

Effects of Task Complexity on L2 Reading and L2 Learning

Jookyoung Jung

(UCL Institute of Education)

Jung, Jookyoung. (2016). Effects of task complexity on L2 reading and L2 learning. *English Teaching*, 71(4), 141-166.

Task-based language teaching (TBLT) has propelled much research into how task type, condition, or demand affects L2 learners' linguistic performance and language learning. To date, however, TBLT has mainly been researched in connection with learners' production, while its applicability to L2 reading has largely been unattended to. To fill this gap, the present study explored whether and how cognitive complexity of L2 reading tasks would affect L2 English reading comprehension and learning of target L2 constructions contained in the texts. The study employed a pretest, posttest, delayed-posttest design with two treatment sessions. The target features were 17 English unaccusative verbs and ten pseudo-words. Participants included 52 Korean college students learning L2 English who were randomly assigned to either – or + complex condition. Reading comprehension was measured with 14 multiple-choice items for each text, and learning of the target constructions was assessed with a grammaticality judgment test and word form and meaning recognition tests. The results of mixed-effects modeling indicated that increased task complexity had limited effects on reading comprehension scores as well as learning of the target unaccusative verbs. Also, task complexity had significant negative effects on vocabulary form recognition scores in the delayed posttest. The results are discussed in relation to models of task-based learning and L2 reading.

Key words: TBLT, task complexity, L2 reading, L2 learning, mixed-effects modeling

1. INTRODUCTION

Since the 1980s, TBLT, in which tasks are deemed to serve as a platform where learners

can enjoy natural opportunities for meaning-oriented communications as well as a medium for infusing focus on form (Ellis, 2003, 2009; Robinson, 2011; Skehan, 1998, 2009), was proposed as a potential approach to L2 instruction, and it has attracted growing attention ever since. A task can be defined as a meaning-oriented activity that requires learners to use the target language (TL) in order to achieve a specified objective (Bygate, Skehan, & Swain, 2001), and it serves not as a mere vehicle for delivering isolated linguistic features, but as a platform where learners can enjoy opportunities for meaning-oriented TL use. The role of task complexity, i.e., the cognitive task demands imposed on learners, has drawn particular attention from researchers, especially with respect to how it influences learners' linguistic performance and language learning that accrues from performing the task. To date, however, the task-based approach has mainly been researched in connection with learners' oral and written production, while its applicability to receptive skills, such as L2 reading, has largely been neglected. In order to fill this gap in the literature, the present study explored whether task complexity affected L2 reading comprehension and learning of L2 constructions contained in the reading texts.

2. LITERATURE REVIEW

2.1. Task Complexity in TBLT

Within the TBLT framework, individual tasks must be designed, selected, and sequenced so as to match the learner's developmental stage and thereby facilitate optimal L2 learning. Among various taxonomies on how to analyze and sequence pedagogic tasks, two rival approaches, i.e., Skehan's *Limited Capacity Model* (Skehan, 1998, 2009; Skehan & Foster, 2001) and Robinson's *Cognition Hypothesis* (Robinson, 2001, 2011), have exerted a substantial impact on recent empirical research on task sequencing and grading. From an attentional capacity perspective, both models attempt to explain and predict how various features of tasks will affect task-generated cognitive demands and in so doing the allocation of learners' attention during task completion. In the Limited Capacity Model, Skehan proposes that the level of task demands depends on three task-related factors: (a) *code complexity*, which pertains to linguistic complexity and variety involved in the task, (b) *cognitive complexity*, which entails processing and computational requirements, and (c) *communicative stress*, which includes time pressure, number of participants, opportunity to control, and so on. Based on a single resource view, all of these factors are considered to have an important bearing on how learners' attention during a task will be shared out and how task performance will be affected in terms of linguistic complexity, accuracy, and fluency (henceforth, CAF). He also suggests competition between linguistic complexity

and accuracy triggered by the natural limitation of attentional resources.

Robinson, on the other hand, defines task complexity as “the result of the attentional, memory, reasoning, and other information-processing demands imposed by the structure of the task on the language learner” (Robinson, 2001, p. 28). As reflected in his definition, Robinson claims that only task-inherent features, not linguistic elements involved in the task, should be considered when determining task complexity. Robinson also proposes the *Multiple Attentional Resources Model* based on Wickens’s (1992) work in cognitive psychology, and claims that increasing task complexity, due to the existence of multiple resource pools, can result in increased linguistic complexity without an expense of accuracy in learner production. The Multiple Attentional Resources Model also provides motivation for the Cognition Hypothesis and its associated *Triadic Componential Framework*. Within this framework, Robinson classifies task features into two dimensions, i.e., *resource-directing* and *resource-dispersing*. Along the resource-directing dimension, a task can become more demanding by increasing the number of elements involved, the amount of the reasoning required, or making reference to a displaced past time event. The cognitive and conceptual need to formulate complex content has the effect of channelling learners’ attention towards lexical and grammatical encoding, which results in greater complexity and accuracy while negatively affecting fluency. By contrast, a task can also become more demanding along the resource-dispersing dimension by reducing the planning time allowed to learners or using unfamiliar task type, content or structures. In this case, learners’ attention is steered towards the consolidation of, and faster access to, the existing interlanguage (IL) system, resulting in a trade-off between linguistic complexity and accuracy. The Cognition Hypothesis further claims that increased task complexity facilitates L2 development. According to Robinson, more complex tasks encourage learners to seek more help from the provided input, which results in greater depth of processing (Craik & Tulving, 1975) and long-term memory of input.

Motivated by these two models, many studies have delved into examining whether and how manipulating task complexity affects learners’ task performance, interactional patterns, and L2 development. The majority of studies have explored how varying levels of task complexity moderated learners’ language production, mostly in terms of CAF (e.g., Foster & Tavakoli, 2009; Michel, 2011, 2013; Révész, 2011). Studies on monologic speech production have produced moderately converging findings: increasing task complexity has small positive effects for accuracy and small negative effects on fluency and linguistic complexity (see Jackson & Suethanapornkul, 2013). In the written mode, availability of planning time seemed to have an effect, as learners were inevitably allowed with more time to prepare and adjust their production (e.g., Kormos & Trebits, 2012). Recently, task complexity has also been shown to have differential effects on face-to-face versus computer-mediated tasks (Baralt, 2013; Yilmaz, 2011).

Researchers have recently started to investigate whether and how task complexity affected the incidence of various interactional features, such as negotiation of meaning (e.g., confirmation check, comprehension check, and clarification request), self-repair or modified output, language-related episodes (LREs), or learners' uptake, all of which deemed as conducive to L2 learning (e.g., Gilabert, Barón, & Llanes, 2009; Nuevo, 2006; Révész, 2011; Révész, Sachs, & Mackey, 2011; Robinson, 2007). The results of the studies have shown that, in general, cognitively complex tasks are likely to increase the amount of negotiation of meaning and the number of LREs (see, however, Nuevo, 2006). There are also studies that measured learning of a specific target form through engaging in interactive tasks, mostly employing pretest-posttest-delayed posttest designs (e.g., Baralt, 2013; Kim & Tracy-Ventura, 2011; Nuevo, 2006; Révész, 2009). The results of these are mixed: increasing task complexity had either no significant (e.g., Nuevo, 2006), marginal (e.g., Baralt, 2013), or positive effects on L2 development (e.g., Kim & Tracy-Ventura, 2011; Révész, 2009).

A critical methodological problem in the existing research on task complexity is that studies so far have more or less eschewed to verify the construct of task complexity, even though it is fundamental to assure the internal validity of task complexity manipulations (Révész, 2014). The few researchers who attempted to check the validity of task complexity manipulations have typically asked learners to respond to post-hoc questionnaires designed to elicit learners' perceptions about how challenging and difficult a task was. The information obtained was used as a basis for inferring the level of cognitive demands induced by the task (e.g., Baralt, 2013; Gilabert et al., 2009; Révész, 2009; Robinson, 2001, 2005, 2007). As expected, however, internal inconsistency and response bias are still open to debate. In order to examine whether the supposedly more complex tasks are indeed more demanding to the learners, new research methods seem highly recommendable, such as subjective time estimation, dual-task performance, and physiological measures (Brünken, Plass, & Leutner, 2003; Norris & Ortega, 2009; Révész, 2014).

2.2. Task Effects on L2 Reading

Reading is a complex psycholinguistic process where a variety of associated component skills come into play (Grabe, 2009; Khalifa & Weir, 2009; Urquhart & Weir, 1998). The entire reading process is affected by the purpose of reading through making metacognitive decisions regarding the scope, depth, and speed of reading required (Khalifa & Weir, 2009). In the case of L2 reading, the incomplete linguistic proficiency of learners may render the importance of regulating reading process across different tasks even more prominent (Horiba, 2000, 2013; Jung, 2012; Taillefer, 1996). For instance, in Taillefer's

(1996) and Jung's (2012) studies, the amount of variance in L2 reading comprehension accounted for by L2 proficiency decreased considerably in a less complex reading task (i.e., scanning) compared to a more complex task (i.e., reading for comprehension). In other words, L2 linguistic processing may make a more marked contribution to a more complex L2 reading task.

While there are not many studies on L2 reading for different purposes, Horiba's (2000, 2013) studies that employed think-aloud protocols provide some useful insights. In her studies that investigated how task instructions affected L1 and L2 readers' text processing and comprehension products, L1 readers were competent and flexible in controlling their reading processes and allocating their cognitive resources strategically according to the type of the text and the task, whereas L2 readers were not able to do so mainly due to linguistic demands. Also, the task effects materialized more clearly at the level of text processing than in the products of comprehension. In other words, different patterns of text processing for different reading goals may not necessarily materialize in the outcome of the reading, which underscores the need for employing concurrent process measures to evince how task affects reading processes more clearly.

In the field of SLA, Yoshimura (2006) investigated whether manipulating foreknowledge of the expected task outcomes leads to different reading behavior, text comprehension, and noticing of L2 form. The reading tasks in this study included reading to memorize the text, reading to retell the text, and reading to visualize the imagery. After reading the text, however, the post-reading tasks were not administered. Rather, the participants were asked to (a) report their reading behaviour by completing a retrospective questionnaire, (b) answer true-or-false comprehension check questions, and (c) fill in the blanks in the text with appropriate verbs. The results showed that the output groups (i.e., reading for memorization and reading for retelling) used more diverse reading strategies. It was further revealed that scores on the verb production test were higher for the memorization group, lower for the retelling group and the lowest for the visualization group. By contrast, comprehension scores were not significantly different across groups. From the findings, Yoshimura suggested that learners' reading processes could be affected by foreknowledge of the required task output, and that different task instructions might have a differential impact on language processing. More importantly, Yoshimura's (2006) study demonstrates that it is viable to use tasks to promote learners' reading for acquisition without interrupting reading for comprehension. That is, if the target construction is regarded as essential for task completion, learners may pay more attention to TL features in the text during reading.

The review of previous studies on the effects of task complexity suggests that the scope of TBLT research has been confined to productive skills and thus needs to be expanded into other language skills, such as L2 reading, for a more nuanced understanding of task

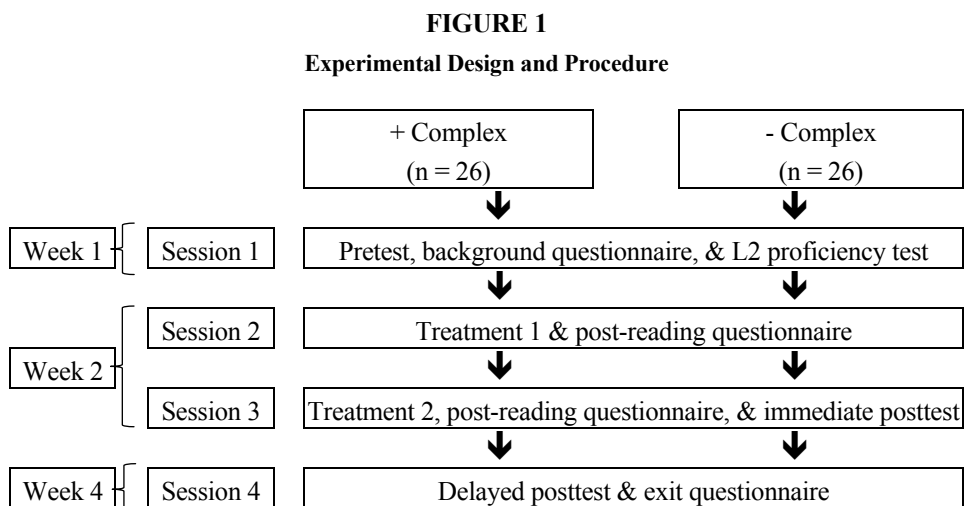
effects of L2 performance and learning. In addition, given the small number of studies on how different task features affect L2 reading, more empirical investigations appear imperative. To fill these gaps, the following research questions will be addressed in the present study:

1. To what extent do the cognitive demands of second language reading tasks affect reading comprehension?
2. To what extent do the cognitive demands of second language reading tasks affect development in the knowledge of target language constructions?

3. METHODOLOGY

3.1. Design

This study examined the impact of task complexity on Korean speakers' L2 English reading and learning. As illustrated in Figure 1, the study employed a pretest, posttest and delayed posttest design, with two treatment sessions. The participants were randomly assigned to either – or + complex condition, and completed two treatment sessions. In each session, participants read a passage taken from a TOEFL exam, while simultaneously answering reading comprehension items. Development in the knowledge of target constructions was measured with a grammaticality judgement test and vocabulary form and meaning recognition tests. More detailed explanations of the research instruments and procedures are provided in the following sections.



3.2. Participants

The participants comprised 14 male and 38 female undergraduate students enrolled in a university in Korea. Their L1 was Korean and their average age was 22.84 years ($SD = 1.94$). They had no explicit instruction on the target construction (i.e., English unaccusative verbs) prior to this study. To ensure the homogeneity of participants' English ability, their English proficiency level was measured with the *Reading and Use of English* section of a practice *Cambridge Proficiency: English* (CPE) test, developed and provided by *University of Cambridge ESOL Examinations*. Based on their scores, stratified random sampling was applied in order to reduce sampling error and ensure equivalence among the groups in terms of English proficiency.

3.3. Texts

For the treatment of this study, two expository texts were selected from passages used for real past TOEFL tests developed by the ETS. The texts were chosen based on two criteria: (a) sufficient number of occurrences of the target constructions and (b) difficult and/or unfamiliar topic to participants. As shown in Table 1, the two texts were comparable in terms of length and readability¹. Texts were presented to the participants in a counter-balanced order.

TABLE 1

Text Characteristics

	Text 1	Text 2
Title	Petroleum Resources	The Cambrian Explosion
Number of words	682	699
Average Readability	11.6	13.4

3.4. Targeted L2 Features

One target L2 feature of the present study was the English unaccusative construction, which Korean learners have been reported to have persistent difficulty in acquiring (e.g., Chung, 2014; Hwang, 1999, 2001; Lee, Miyata, & Ortega, 2008; No & Chung, 2006). Ten pseudo-words were additionally included in order to examine the effects of task complexity on incidental learning of lexical items.

¹ Average readability was calculated from various readability indices including *Flesch-Kincaid grade level*, *Gunning-Fog score*, *Coleman-Liau index*, *SMOG index*, and *Automated Readability index*.

3.4.1. English unaccusative verbs

Perlmutter (1978) first introduced the *Unaccusativity Hypothesis* in which intransitive verbs are classified into either unergatives or unaccusatives. Whereas an unergative verb assigns an agent-role of a volitional act to its subject, the subject of an unaccusative verb lacks volitional control, performing a patient-role (e.g., Unergative: *Mary danced.*; Unaccusative: *The snow melted.*). SLA researchers have consistently found that L2 learners of English tend to overuse passive structures with unaccusative verbs (e.g., *The car was disappeared.*). Multiple factors may be partly responsible for this learnability problem. For example, some researchers (Hwang, 1999, 2001) suggest that if an unaccusative verb has its transitive counterpart (e.g. *ship, change, close, break*), then L2 English learners may have a stronger tendency to make overpassivization errors than with non-alternating unaccusative verbs (e.g. *happen, result, arrive, disappear*). Lack of a conceptualizable agent (Ju, 2000), L1 transfer (No & Chung, 2006), and low input frequency (Lee, Miyata, & Ortega, 2008) have also been explored as potential factors increasing the difficulty in acquiring English unaccusativity.

In the present study, 17 English unaccusative verbs were identified from the two treatment texts and selected as target constructions. As illustrated in Table 2, all target verbs included in the texts were low frequency, and hence the participants were expected to have a limited knowledge about those verbs. Out of the 17 verbs, six verbs were non-alternating unaccusative verbs, whereas the rest were alternating. Each of the unaccusative verbs appeared in the texts once.

TABLE 2
Target English Unaccusative Verbs

	Text 1			Text2			
	Unaccusative verbs	Alternating		Frequency (per 450 million)	Unaccusative verbs	Alternating	Frequency (per 450 million)
1	decompose	A	312	1	fossilize	A	11
2	subside	NA	568	2	date to	A	743
3	ascend	A	759	3	originate	A	1022
4	accumulate	A	1814	4	consist of	NA	2140
5	cease	A	2554	5	persist	NA	2684
6	diminish	A	2701	6	evolve	A	3184
7	drift	NA	4477	7	disappear	NA	7581
8	collect	A	10525	8	emerge	NA	9116
9	settle	A	10873				

3.4.2. Pseudo-words

In addition to the English unaccusative verbs, ten lexical items were also included as

target constructions in this study (see Table 3). The lexical targets were carefully selected from the two texts based on the following conditions: (a) the word is a noun (to control the part of speech) and (b) the word appears once (to control for frequency). Five words were selected from each text and replaced with pseudo-words that followed English orthographic and morphological rules (Pulido, 2007). When the original word was in a plural form, plurality was also marked in the corresponding pseudo-word by attaching *-s*. Each of the pseudo-words consisted of two syllables, containing seven letters, in order to control the length. The target unaccusative verbs were not substituted with pseudo-words as the focus was learning the unaccusative usage of the verbs rather than their forms or meanings.

TABLE 3
Target Pseudo-words

	Text 1			Text2	
	Pseudo-words	Original words		Pseudo-words	Original words
1	stragon	bottom	1	cabrons	changes
2	golands	spouts	2	fration	absence
3	phosens	discoveries	3	zenters	clues
4	klaners	parks	4	morbits	descendants
5	stovons	beaches	5	tralion	predator

3.5. Treatment Tasks

The treatment task in this study is fundamentally what learners have to do when taking a reading section of a TOEFL test, i.e., reading the provided text and answering to multiple-choice comprehension questions. In other words, reading comprehension measure was embedded in the learning task so that the level of participants' text understanding could be simultaneously measured in tandem with task completion. The multiple-choice reading comprehension items were also taken from the previous TOEFL tests developed and tested by ETS. The reading comprehension items asked participants to identify factual/negative factual information, make inferences, understand rhetorical purpose, recognize vocabulary meaning, determine reference, simplify/paraphrase a sentence, insert a sentence into a paragraph, and select main ideas of the text (Educational Testing Service, 2012). The texts were divided into five segments, comprised of either one or two paragraphs, and followed by reading comprehension questions relevant to each segment.

In this paper, task complexity was defined as task-induced demands made on learners' cognitive resources while performing a task. In the – complex condition, participants were asked to read and answer the comprehension questions as they normally would when working on a reading section of a TOEFL test. In the + complex condition, the segments were jumbled and presented to participants in a mixed order. Thus, in addition to reading

the paragraphs and answering to comprehension questions, participants in the complex group also had to reorder the segments in a correct order to make a coherent text. The latter task was judged as cognitively more demanding in that readers' comprehension is substantially influenced by the degree of clarity and coherence of text structure (Meyer & Ray, 2011). There was no time limit required for task completion. The total score for reading comprehension was 15 for each text.

3.6. Assessment Tasks

L2 learning in this study was operationalized as (a) the ability to recognize grammaticality of English unaccusative verbs and (b) the ability to recognize the form and meaning of the target lexical items.

3.6.1. Grammaticality judgment test

The GJT used for this study contained 80 sentences in total, including (a) 34 sentences for target unaccusative verbs, (b) 16 for novel unaccusative verbs, and (c) 30 distracters. First, for each of the 17 target unaccusative verb, one grammatical and one ungrammatical passive sentence were created, resulting in 34 sentences (e.g. *The sun was soon disappeared* vs. *The tension soon disappeared*).

Also, in order to explore if participants' learning from the treatment transferred to other unaccusative verbs, eight additional verbs (i.e., *occur*, *remain*, *appear*, *fall*, *burn*, *stop*, *break*, and *change*) were selected from the list of 2000 most frequently used English words, consulting the *Compleat Lexical Tutor* version 6.2. For these eight verbs, 8 grammatical and 8 ungrammatical sentences were produced, totaling in 16 sentences. Caution was paid to control the number of syllables, syntactic complexity, semantic plausibility, vocabulary familiarity, and the position of the unaccusative verbs for each of the 25 pairs of the unaccusative sentences.

Lastly, thirty sentences, 15 grammatical and 15 ungrammatical, were included as distracters. The grammatical rules the distracters draw on included gerunds, to-infinitives, subjective moods, comparatives, participial adjectives, reflexives, relative pronouns, inversion, and prepositions, which covered the topics generally dealt in English grammar lessons in Korea (No & Chung, 2006). Across the pretest, posttest, and delayed posttest, the same 80 sentences were randomly presented to participants.

The GJT for this study was constructed using E-Prime 2.0. The 80 sentences were presented on a computer screen, and participants were asked to press the "z" key if the sentence seemed grammatical and the "m" key if the sentence seemed ungrammatical. These particular keys were chosen considering their placement in the QWERTY keyboard,

which is the norm layout in Korea. Participants were instructed to make their decision as fast as they could. Each sentence remained on the screen until the decision on the sentence well-formedness was made. Between each stimulus, a fixation cross appeared at center of the screen for 500 milliseconds to signal an upcoming sentence. The total score was 50 (34 for the target verbs and 16 for the novel verbs), and the test took approximately 10 to 12 minutes to complete.

3.6.2. Vocabulary form recognition test

In order to measure if task complexity affected form recognition of the target pseudo-words, 20 items were constructed using E-Prime 2.0. Participants were asked to press either the “*z* (yes)” or the “*m* (no),” depending on whether they remembered seeing the word from the texts. Ten items were the target pseudo-words, whereas the other ten were distracters that were constructed drawing upon the pseudo-words in Godfroid, Housen, and Boers (2013). Each of the distracters contained two syllables and seven letters as the target pseudo-words. In addition to the 20 form recognition items, 20 additional multiple-choice items, modeled after Martinez-Fernández’s (2010) meaning recognition test, asked participants to select a correct Korean translation of the given target word out of three choices. Among these, ten items were the target words while the other ten were the distracters used in the form recognition test. As in the GJT, each set of the form recognition and the meaning recognition items were randomized and presented on a computer screen, and participants were asked to choose the answer for each item. Again, between each stimulus, a fixation cross appeared on the screen for 500 milliseconds to signal the next item. Total score for each of the vocabulary form and meaning recognition test was 10 (1 for each correct and 0 for each incorrect item), and each test took approximately 2 to 3 minutes.

3.7. Questionnaires

Participants were asked to answer a background questionnaire, a post-reading questionnaire, and an exit questionnaire. The aim of the background questionnaire was to collect information about participants’ demographics and English learning experiences. The post-reading questionnaires asked participants to provide their retrospective subjective time estimation taken to complete the given reading task and familiarity with the topic of the reading text. As the time estimation task was conducted under a retrospective paradigm (Fink & Neubauer, 2001), participants were unaware of the upcoming duration judgement task until it had to be done. As such, subjective time estimations were only collected after the first treatment session. It was expected that the duration ratio would increase after

performing cognitively more demanding tasks. Finally, the exit questionnaire asked participants to give retrospective comments about the treatment sessions. All questionnaires were administered in Korean.

4. PROCEDURE

The data were collected over four weeks. All participants took the pretest, background questionnaire, and the L2 proficiency test in the first session. In sessions 2 and 3, they took part in the treatment sessions, each followed by a post-reading questionnaire. In session 3, they also completed an immediate posttest. In the fourth week, participants in the experimental conditions completed a delayed posttest. Each session took approximately 45 minutes to an hour. The participants in the experimental conditions carried out the sessions in a computer-laboratory at a university as a group.

5. ANALYSIS

SPSS 22.0 for Mac was used for examining reliability of the tests as well as computing descriptive and correlational statistics of the data. More specifically, the reliability of the different tests was determined using Cronbach's alpha, and interrelationships between the various test scores were computed using Pearson's coefficients. The level of significance for this study was set at alpha level of $p < .05$. The relationships between the variables were analyzed by constructing various mixed-effects models constructed with the package *lme4* (Bates, Maechler & Bolker, 2012), using the statistical program R version 3.3.0 (R Development Core Team, 2016). The fixed effect was Complexity and the random effects were Subjects and Items. For GJT scores and vocabulary recognition scores, Time (pretest – posttest – delayed posttest) was put into the models as an additional fixed effect in order to explore changes in the data over the repeated measurements. For *t* statistics, absolute *t*-values above 2.0 was set for testing significance of the models (Gelman & Hill, 2007). Effect sizes for the linear mixed-effects models were computed using *r.squaredGLMM* function in the package *MuMIn* (Barton, 2015), whereas that of the logit mixed-effects models was calculated with *C* index of the concordance using *somer2* function in *Hmisc* package (Harrell & Dupont, 2015). Following Plonsky and Oswald (2014), R^2 's value of .06, .16, and .36 were interpreted as small, medium, and large respectively. A *C*-index of .70 was considered as a moderate, .80 as good, and .90 and above as excellent fit for the data (Baayen, 2008; Rogers, 2016). For *t*-tests, Cohen's *d* was calculated to examine effect sizes. As suggested by Plonsky and Oswald, the benchmarks were .40 for small, .70 for

medium and 1.00 for large effect sizes for independent-sample *t*-tests, and .60 for small, 1.00 for medium and 1.50 for large effect sizes for paired-sample *t*-tests.

6. RESULTS

6.1. Preliminary Analysis

Prior to answering the research questions, some preliminary steps were taken to ensure the reliability of the instruments and validity of the results. The following methodological concerns were taken into consideration: reliability of the tests, participants' prior knowledge of the target items, potential effects of topic familiarity on reading comprehension scores, and validation of task complexity.

6.1.1. Test reliability

In order to check consistency and stability of the instruments, reliability coefficients for the proficiency test, reading comprehension tests, grammaticality judgment tests, and vocabulary recognition tests were computed using Cronbach's alpha. As summarized in Table 4, values of Cronbach's alpha were found to be high for the proficiency test and grammaticality judgment tests, but low for the reading comprehension tests and vocabulary recognition tests. Also, the variances in the scores of reading comprehension tests and vocabulary recognition tests were relatively small, presumably contributing to the low reliability coefficients. In addition, the mean reading comprehension scores appeared to imply a ceiling effect, which could have further contributed to the low internal consistency reliability of the reading comprehension tests.

TABLE 4
Descriptive Statistics for Test Scores

Test	<i>N</i>	<i>M</i>	<i>SD</i>	<i>Cronbach's alpha</i>
CPE test	52	14.29	4.94	.70
Reading comprehension (Text 1)	52	11.04	2.22	.47
Reading comprehension (Text 2)	52	12.85	1.51	.37
Grammaticality judgment (Target items)	52	58.52	11.77	.80
Grammaticality judgment (Novel items)	52	31.92	6.07	.67
Vocabulary recognition (Form)	52	11.64	3.65	.59
Vocabulary recognition (Meaning)	52	5.64	2.96	.45

Note. Maximum score for: CPE test = 45, reading comprehension = 15, grammaticality judgment (target) = 102, grammaticality judgment (novel) = 48, vocabulary form recognition = 20, vocabulary meaning recognition = 20.

6.1.2. Equivalence among groups

To check the equivalence of English proficiency level among the groups, a mixed-effects model was constructed, with the CPE scores as the dependent variable, Group as a fixed effect, and Subject and Item as random effects. When compared with a null model that contained only random effects, the results showed that the inclusion of Group as a fixed effect did not make a significant difference to the null model, $\chi^2(1) = .29, p = .59$. In other words, the groups did not significantly differ from each other in terms of their English proficiency (for descriptive statistics, see Table 5).

TABLE 5
Descriptive Statistics for Proficiency Test by Group

Group	Proficiency test			
	<i>N</i>	<i>M</i>	<i>SD</i>	<i>SE</i>
- Complex	26	14.62	4.97	.98
+ Complex	26	13.96	4.96	.98

Note. Maximum score = 45.

Next, in order to test whether the four groups started out at a developmentally parallel stage, another set of likelihood ratio tests were conducted on pretest GJT scores comparing null models with random effects only and models additionally containing Group as a fixed effect (for descriptive statistics, see Table 11). The results indicated that Group did not improve the null models to a significant degree, Target verbs: $\chi^2(1) = .25, p = .62$; Novel verbs: $\chi^2(1) = 1.14, p = .29$. In sum, the results showed that, at the time of the pretest, there was no significant difference among the groups in their ability to judge the grammaticality of the English unaccusative sentences.

6.1.3. Effects of topic familiarity

To assess the potential impact of topic knowledge on comprehension of the treatment texts, participants' familiarity with the two topics was measured using post-reading questionnaire items (i.e., Item 1: *I thought this topic of the reading was familiar*, Item 2: *I had some background knowledge about the reading topic*).

The descriptive statistics are presented in Table 6. The responses to the two items were significantly correlated with each other, Text 1: $r(52) = .68, p < .01$, Text 2: $r(52) = .56, p < .01$, suggesting that the items assessed overlapping constructs. In order to examine the effects of topic familiarity on reading comprehension scores, likelihood ratio tests were conducted comparing a null model with the random effects and models additionally including topic familiarity as a fixed effect. The dependent variable was reading

TABLE 6
Descriptive Statistics for Topic Familiarity by Item

Item	N	Topic familiarity					
		Text 1			Text 2		
		M	SD	SE	M	SD	SE
# 1	52	3.60	.24	1.75	3.10	.24	1.76
# 2	52	3.48	.23	1.69	2.46	.19	1.34
Total	52	7.08	.44	3.16	5.55	.38	2.75

Note. Maximum value for each item = 7.

comprehension scores for Text 1 and Text 2. The results showed that adding topic familiarity did not make a significant improvement to the null models, Text 1: $\chi^2(1) = .01$, $p = .91$, Text 2: $\chi^2(1) = 2.25$, $p = .13$. In short, the participants' topic familiarity with the texts did not affect their scores on the reading comprehension items.

6.1.4. Validation of task complexity

To validate the operationalization of task complexity, all participants were asked to judge the perceived time duration taken to complete each task immediately after the task completion. As mentioned earlier, only the time estimations made after completing the first task were analyzed. In order to examine whether the subjective time estimations differed as a function of task manipulation, the estimated-to-target duration ratios were calculated by dividing estimated time by real time taken to complete the given task. In the retrospective time estimation paradigm, duration judgment ratio is expected to increase with greater cognitive load.

As shown in Table 7, for both Text 1 and Text 2, duration judgment ratios in the + complex conditions were on average larger than those in the – complex conditions. The results from independent samples *t*-tests on the duration judgment ratios across + and – complex conditions also revealed significant effects of task complexity for both Text 1 and Text 2, Text 1: $t(50) = 2.86$, $p = .01$, 95% CI [.04, .22]; Text 2: $t(50) = 3.85$, $p < .01$, 95% CI [.11, .36]. Cohen's *d*s were .79 and 1.09 respectively, which were considered as medium and large effect sizes. In other words, duration judgment ratios in the + complex conditions were significantly greater than those in the – complex conditions, implying that the + complex tasks induced heavier cognitive loads on the participants compared to the – complex tasks.

To infer the effects of task complexity on the amount of mental effort posed on the participants, three questionnaire items were included in the post-task questionnaires (Item 3: *I thought this task was difficult*, Item 4: *I invested a large amount of mental effort to complete this task*, Item 5: *I thought this task was demanding*). The Cronbach's alpha for

the three items was .63 for Text 1 and .75 for Text 2. Descriptive statistics for the responses to the three items are presented in Table 8.

TABLE 7
Descriptive Statistics for Duration Judgment Ratio

Condition	N	Text 1		Text 2	
		M	SD	M	SD
- Complex	13	1.03	(.16)	.95	(.11)
+ Complex	13	1.16	(.17)	1.19	(.29)
Total	26	1.09	(.18)	1.08	(.25)

TABLE 8
Descriptive Statistics for Reported Mental Effort

Item	Condition	N	Reported mental effort					
			Text 1			Text 2		
			M	SD	SE	M	SD	SE
# 3	- Complex	26	4.23	1.03	.20	3.77	1.03	.20
	+ Complex	26	4.65	.80	.16	4.46	1.07	.21
# 4	- Complex	26	5.12	1.11	.22	4.69	1.12	.22
	+ Complex	26	4.85	1.19	.23	4.42	1.42	.28
# 5	- Complex	26	4.00	1.17	.23	3.62	1.10	.22
	+ Complex	26	4.46	1.42	.28	4.19	1.27	.25
Total	- Complex	26	13.35	2.50	.49	12.08	2.79	.55
	+ Complex	26	13.96	2.71	.53	13.08	3.02	.59

Note. Maximum value for each item = 7.

In order to see if there was significant differences between the + and the - complex conditions in participants' ratings of perceived task difficulty, independent-sample *t*-tests were conducted. The results revealed that there was no significant difference between the conditions, $t(50) = .852, p = .40, 95\% \text{ CI } [-.83, 2.07]$; Text 2: $t(50) = 1.242, p = .22, 95\% \text{ CI } [-.62, 2.62]$. Cohen's *d*s were .23 and .34 respectively, indicating small effect sizes. In short, participants' perceived level of task difficulty appeared comparable regardless of task manipulation.

6.2. Effects of Task Complexity on L2 Reading Comprehension

The descriptive statistics for the reading comprehension scores of each group are displayed in Table 9. Reading comprehension scores on Text 2 were on average higher than those on Text 1.

In order to examine whether task complexity had a significant impact on L2 reading comprehension scores, linear mixed-effects models were constructed with R. Null models contained random effects (i.e., Subject and Item) only, and Complexity was entered and

compared against the null models with likelihood ratio tests using χ^2 statistics. As summarized in Table 10, task complexity was shown to have no significant effects on reading comprehension scores.

TABLE 9
Descriptive Statistics for Reading Comprehension Scores

Group	N	Text 1			Text 2		
		M	SD	SE	M	SD	SE
- Complex	26	11.08	2.26	.44	13.08	1.13	.22
+ Complex	26	11.00	2.23	.44	12.62	1.81	.36
Total	52	11.04	2.22	.31	12.85	1.51	.21

Note. Maximum score = 15.

TABLE 10
Summary of Likelihood Ratio Tests for Complexity on Reading Comprehension Scores

	χ^2	df	p	R ²
Text 1	.02	1	.90	.23
Text 2	.40	1	.53	.14

6.3. Effects of Task Complexity on Learning of Unaccusative Verbs

Table 11 presents the descriptive statistics for the GJT scores by group. It can be noticed that mean gain scores were overall higher in the delayed posttest than in the immediate posttest.

TABLE 11
Descriptive Statistics for Gains in GJT

Group	Test	N	Target items			Novel items		
			Mean	Mean gain	SD	Mean	Mean gain	SD
- Complex	Pretest	26	17.92		4.34	10.50		2.57
	Immediate posttest	26	19.58	1.58	5.05	10.38	-.04	2.22
	Delayed posttest	26	20.54	2.54	4.89	11.15	.73	2.41
+ Complex	Pretest	26	17.35		3.36	10.04		2.91
	Immediate posttest	26	19.77	2.62	4.05	10.77	.58	2.27
	Delayed posttest	26	21.73	4.35	4.06	11.23	1.19	2.58

Note. Maximum score for: target GJT items = 34, novel GJT items = 16.

In order to examine whether task complexity had a significant impact on target GJT scores, logit mixed-effects models were constructed with R. The dependent variable was GJT gain scores, i.e., the changes in the GJT scores compared to the pretest scores. The fixed effect was Complexity. The null models contained only the random effects, i.e., Subject, Item, and Time (changes in the GJT scores over the repeated measures). Then,

Complexity was added and tested against the null models to see whether the inclusion of the fixed effects significantly improved the model fit. As shown in Table 12, Complexity was shown to have no significant influence on the GJT gain scores for the target verbs.

TABLE 12
Summary of Likelihood Ratio Tests for Complexity on GJT Gain Scores for Target Verbs

	χ^2	df	<i>p</i>	<i>R</i> ²
Immediate gain	.34	1	.85	.81
Delayed gain	2.40	1	.30	.81

Next, another series of likelihood ratio tests were conducted to identify whether Complexity made significant difference to the null models on the GJT gain scores for the novel verbs. As summarized in Table 13, significance was not found, indicating transfer of learning did not occur.

TABLE 13
Summary of Likelihood Ratio Tests for Complexity on GJT Gain Scores for Novel Verbs

	χ^2	df	<i>p</i>	<i>R</i> ²
Immediate gain	1.92	1	.38	.81
Delayed gain	2.94	1	.23	.82

6.4. Effects of Task Complexity on Learning of Pseudo-Words

Table 14 presents the descriptive statistics for the vocabulary recognition scores by group. The mean scores on the form recognition test were overall higher than those on the meaning recognition test. Also, the mean and form recognition scores from the delayed posttest were higher than those from the immediate posttest, whereas the mean meaning recognition scores on the delayed posttest were lower than those on the immediate posttest.

TABLE 14
Descriptive Statistics for Vocabulary Recognition Score

Group	Test	<i>N</i>	Form		Meaning	
			<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
– Complex	Immediate posttest	26	5.38	2.28	3.04	1.66
	Delayed posttest	26	6.77	2.05	2.35	1.41
+ Complex	Immediate posttest	26	5.50	1.98	3.15	1.85
	Delayed posttest	26	5.62	1.85	2.96	1.78

Note. Maximum score for: form recognition = 10, meaning recognition = 10.

In order to examine whether task complexity improved null models to a significant degree, repeated likelihood ratio tests were conducted using χ^2 statistics. The dependent

variables were scores in the immediate posttest and those in the delayed posttest. The null models included random effects only (i.e., Subject and Item) and each of Complexity was added and tested against the null models. As shown in Table 15, Complexity improved the null model in the delayed posttest.

TABLE 15**Summary of Likelihood Ratio Tests for Complexity on Vocabulary Form Recognition Scores**

	χ^2	df	<i>p</i>	<i>R</i> ²
Immediate gain	.26	1	.61	.77
Delayed gain	4.98	1	.03	.79

Then, logit mixed-effects models were constructed with Complexity on the delayed posttest scores. As Table 16 presents, Complexity had significant negative effects in the delayed posttest. The *C* index of concordance was .79, which indicated good model fit for the data. In short, participants in the + complex conditions scored significantly less in the delayed vocabulary form recognition test than those in the – complex conditions.

TABLE 16**Results for the Best-Fit Logit Mixed-Effects Models on Immediate Vocabulary Form Recognition Scores**

		Fixed effects				Random effects	
		Estimate	<i>SE</i>	<i>z</i>	<i>p</i>	by-Subject <i>SD</i>	by-Item <i>SD</i>
Delayed	Intercept	.59	.24	2.48	.01	.45	.62
	Complexity	-.63	.27	-2.30	.02	.93	.06

Note. Formula: VF ~ Complexity + (Complexity | Subject) + (Complexity | Item); *C* = .79.

Finally, effects of task complexity on vocabulary meaning recognition were explored, beginning with another series of likelihood ratio tests using χ^2 statistics. Again, the null models included random effects only, and the fixed effects, i.e., Complexity, was added to the null model one by one and examined if this improved the null models to a significant extent. As summarized in Table 17, Complexity had no significant effects on vocabulary meaning recognition scores.

TABLE 17**Summary of Likelihood Ratio Tests for Complexity on Vocabulary Meaning Recognition Scores**

	χ^2	df	<i>p</i>	<i>R</i> ²
Immediate gain	.31	1	.58	.80
Delayed gain	1.64	1	.20	.83

7. DISCUSSION AND CONCLUSION

In this study, it was investigated whether task complexity affected Korean undergraduate students' English reading comprehension and their learning of target lexical constructions contained in the texts. Task complexity was manipulated by disarranging paragraphs of each text, based on the understanding that coherent and clear text structure considerably facilitates reading comprehension (Meyer & Ray, 2011). In this study, reading comprehension scores were not affected by task complexity of the texts. It should be noted, however, that as shown in the relatively high mean scores and small SDs, participants overall performed well in the reading comprehension tests, and thus a ceiling effect might have masked between-group differences. More difficult reading comprehension tests may result in higher SDs and reliability of the reading comprehension tests. In addition, there was no time limit and participants were allowed to stay on the task as long as they felt necessary, which might have contributed to the non-significant effects of task manipulation on reading comprehension scores. Another possibility is that participants' reading processes could have been affected by task complexity, although the effects did not surface in the reading comprehension scores. In order to explore this issue, participants' verbal reports on their internal processes while performing tasks or eye-movement data will be highly informative.

It was also found that task complexity failed to affect the learning of the English unaccusative verbs. It seems possible to assume that paragraph-ordering task in the + complex conditions did not necessarily encourage participants to process the target unaccusative verbs to a significantly greater extent. More specifically, the ordering task might have led participants to depend on the initial or the final part of each paragraph selectively, rather than paying attention to each paragraph thoroughly. Also, arranging paragraphs might have promoted higher-level conceptual reasoning rather than lower-level text-based processing, thereby not influencing learning of the target verbs. In other words, when re-arranging the paragraphs, participants might have focused more on the main idea of each paragraph and tried to figure out the logical order among the global ideas. In addition, the task in the + complex conditions in this study could have not been complex enough than the – complex task. Indeed, the task manipulation was shown successful only by the subjective time estimations, but not by self-reports on the perceived level of task difficulty.

Task complexity, though, had significant negative effects on form recognition in the delayed posttest. That is, participants assigned in the + complex conditions were less successful in recognizing target word forms than those in the – complex conditions. It seems possible to assume that the increased level of task complexity could have driven participants' attention to the paragraph-ordering task, and accordingly away from attending

the pseudo-word forms. When participants were allowed to read the text in a coherent order under the – complex conditions, they might have enjoyed extra mental resources to be shared out for processing the target word forms. This is open to empirical investigation, preferably using on-line methodologies such as intro- or retrospective verbal reports or eye-movement data.

The results from this study cast valuable insight for future research. Firstly, it was speculated that a ceiling effect could have masked the effects of task complexity on reading comprehension scores. Indeed, mean scores were low while variances were small, suggesting an inherent limitation in detecting significant effects of task complexity on reading comprehension scores. Therefore, it was assumed that the difficulty of reading comprehension tests might need to be increased in the future studies so that the variances among the participants could be inflated. It was also concluded that, in order to better detect the effects of task complexity, task manipulation had to be conducted on a more localized level in such a way that text-bound linguistic processing could be facilitated in a more complex condition. In this study, the tasks were manipulated on a global-level (re-arranging paragraphs into a coherent order) and thus led the participants to rely on top-down conceptual processing, which in turn failed to affect linguistic processing of the target constructions. It was speculated that, local-level task manipulation might encourage learners to read the given text more thoroughly so that their processing of target features could be more likely to differ across the + and – complex task conditions. It was also considered that a time limit might also play a role in magnifying the effects of cognitive complexity of each task by placing additional cognitive demands on participants.

Last but not least, the reading tasks in this study involved reading the passages provided while answering multiple-choice reading comprehension questions, which is fundamentally what learners would normally do when taking a language aptitude or proficiency test. In this regard, there may be concerns regarding the ecological validity of the task in terms of its resemblance to real-world reading tasks. Yet, the applicability of the reading tasks used in this study has implications in many cases in academic settings where learners take exams. Also, in the present study, the extent to which the reading task invoked the kind of cognitive processes that are essential in performing a real-world task was considered more important than how much the task approximated to a target task in its appearance. For instance, if a learner can read a given text and identify the author's intention, as one of the reading comprehension questions required, we can make a valid assumption that the learner is likely to perform other similar tasks, such as reading an editorial and understanding the author's opinion. It should be acknowledged, however, that ecologically more valid tasks that closely resemble real-world reading practices would also generate insightful findings as to the task-based approach to L2 reading instruction.

REFERENCES

- Baayen, R. H. (2008). *Analyzing linguistic data: A practical introduction to statistics*. Cambridge, UK: Cambridge University Press.
- 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, 689-725.
- Barton, K. (2015). *MuMIn: Multi-Model Inference*. R package version 1.13.4. <http://cran.r-project.org/package=MuMIn>.
- Bates, D. M., Maechler, M., & Bolker, B. (2012). *lme4: Linear mixed-effects models using Eigen and syntax*. R package version 0.999999-0.
- Brünken, R., Plass, J. L., & Leutner, D. (2003). Direct measurement of cognitive load in multimedia learning. *Educational Psychologist*, 38(1), 53-61.
- Bygate, M., Skehan, P., & Swain, M. (2001). *Researching pedagogic tasks, second language learning, teaching and testing*. Harlow: Longman.
- Chung, T. (2014). Multiple factors in the L2 acquisition of English unaccusative verbs. *IRAL*, 52(1), 59-87.
- Craik, F. I. M., & Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology: General*, 104, 268-294.
- Educational Testing Services (2012). *Official TOEFL iBT Tests*. McGraw-Hill: New York.
- Ellis, R. (2003). *Task-based language learning and teaching*. Oxford, UK: Oxford University Press.
- Ellis, R. (2009). Task-based language teaching: Sorting out the misunderstandings. *International Journal of Applied Linguistics*, 19(3), 221-246.
- Fink, A., & Neubauer, A. C. (2001). Speed of information processing, psychometric intelligence: And time estimation as an index of cognitive load. *Personality and Individual Differences*, 30, 1009-1021.
- Foster, P., & Tavakoli, P. (2009). Native speakers and task performance: Comparing effects on complexity, fluency, and lexical diversity. *Language Learning*, 59(4), 866-896.
- Gelman, A., & Hill, J. (2007). *Data analysis using regression and multilevel/hierarchical models*. New York, NY: Cambridge University Press.
- Gilabert, R., Barón, J., & Llanes, À. (2009). Manipulating cognitive complexity across task types and its impact on learners' interaction during oral performance. *IRAL*, 47, 367-395.
- Godfroid, A., Housen, A., & Boers, F. (2013). An eye for words: Gauging the role of attention in incidental L2 vocabulary acquisition by means of eye-tracking. *Studies in Second Language Acquisition*, 35, 483-517.

- Grabe, W. (2009). *Reading in a second language: Moving from theory to practice*. New York: Cambridge University Press.
- Harrell, F. E., & Dupont, C. (2015). *Hmisc: Harrell Miscellaneous*. URL <https://CRAN.R-project.org/package=Hmisc>. R package version 3.17-0.
- Horiba, Y. (2000). Reader control in reading: Effects of language competence, text type, and task. *Discourse Processes, 29*, 223-267.
- Horiba, Y. (2013). Task-induced strategic processing in L2 text comprehension. *Reading in a Foreign Language, 25*(2), 98-125.
- Hwang, J. B. (1999). L2 acquisition of English unaccusative verbs under implicit and explicit learning conditions. *English Teaching, 54*(4), 145-176.
- Hwang, J. B. (2001). Focus on form and the L2 learning of English unaccusative verbs. *English Teaching, 56*(3), 111-133.
- 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.
- Ju, M. K. (2000). Overpassivization errors by second language learners: The effect of conceptualizable agents in discourse. *Studies in Second Language Acquisition, 22*, 85-111.
- Jung, J. (2012). Relative roles of grammar and vocabulary in different L2 reading tasks. *English Teaching, 67*(1), 57-77.
- Khalifa, H., & Weir, C. J. (2009). *Examining reading: Research and practice in assessing second language learning*. Cambridge, UK: Cambridge University Press.
- Kim, Y.-J., & Tracy-Ventura, N. (2011). Task complexity, language anxiety, and the development of the simple past. In P. Robinson (Ed.), *Researching second language task complexity: Task demands, language learning and language performance* (pp. 287-306). Amsterdam: John Benjamins.
- Kormos, J., & Trebits, A. (2012). The role of task complexity, modality, and aptitude in narrative task performance. *Language Learning, 62*(2), 439-472.
- Lee, S.-K., Miyata, M., & Ortega, L. (2008). A usage-based approach to overpassivization: The role of input and conceptualization biases. *Paper presented at the 26th Second Language Research Forum*, Honolulu, HI, October 17-19.
- Martinez-Fernández, A. M. (2010). *Experiences of remembering and knowing in SLA, L2 development, and text comprehension: A study of levels of awareness, type of glossing, and type of linguistic item*. Unpublished dissertation, Georgetown University, Washington, D.C.
- Meyer, B. J. F., & Ray, M. N. (2011). Structure strategy interventions: Increasing reading comprehension of expository text. *International Electronic Journal of Elementary Education, 4*(1), 127-152.

- Michel, M. C. (2011). Effects of task complexity and interaction on L2 performance. In P. Robinson (Ed.), *Researching second language task complexity: Task demands, language learning and language performance* (pp. 141-173). Amsterdam: John Benjamins.
- Michel, M. C. (2013). The use of conjunctions in cognitively simple versus complex oral L2 tasks. *The Modern Language Journal*, 97(1), 178-195.
- No, G., & Chung, T. (2006). Multiple effects and the learnability of English unaccusatives. *English Teaching*, 61(1), 19-39.
- 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.
- Nuevo, A. (2006). *Task complexity and interaction: L2 learning opportunities and interaction*. Unpublished doctoral dissertation. Georgetown University, Washington D.C.
- Perlmutter, D. M. (1978). Impersonal and the unaccusative hypothesis. *Proceedings of the 4th Annual Meeting of the Berkeley Linguistics Society*, 157-190.
- Plonsky, L., & Oswald, F. L. (2014). How big is "big"?: Interpreting effect sizes in L2 research. *Language Learning*, 64(4), 878-912.
- Pulido, D. (2007). The effects of topic familiarity and passage sight vocabulary on L2 lexical inferencing and retention through reading. *Applied Linguistics*, 28(1), 66-86.
- R Development Core Team (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>.
- Révész, A. (2009). Task complexity, focus on form, and second language development. *Studies in Second Language Acquisition*, 31, 437-470.
- Révész, A. (2011). Task complexity, focus on L2 constructions, and individual differences: A classroom-based study. *The Modern Language Journal*, 95, 162-181.
- Révész, A. (2014). Towards a fuller assessment of cognitive models of task-based learning: Investigating task-generated cognitive demands and processes. *Applied Linguistics*, 35(1), 87-92.
- Révész, A., Sachs, R., & Mackey, A. (2011). Task complexity, uptake of recasts, and second language development. In P. Robinson (Ed.), *Researching second language task complexity: Task demands, language learning and language performance* (pp. 203-236). Amsterdam: John Benjamins.
- Robinson, P. (2001). Task complexity, cognitive resources and syllabus design: A triadic framework for examining task influences on SLA. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 193-226). Cambridge: Cambridge University Press.
- Robinson, P. (2005). Cognitive complexity and task sequencing: Studies in a componential

- framework for second language task design. *International Review of Applied Linguistics*, 43, 1-32.
- Robinson, P. (2007). Task complexity, theory of mind, and intentional reasoning: Effects on L2 speech production, interaction, uptake and perceptions of task difficulty. *IRAL*, 45, 193-213.
- Robinson, P. (2011). *Researching second language task complexity: Task demands, language learning and language performance*. Amsterdam: John Benjamins.
- Rogers, J. R. (2016). *Developing implicit and explicit knowledge of L2 case marking under incidental learning conditions*. Unpublished dissertation, University College London Institute of Education, London, UK.
- Skehan, P. (1998). *A cognitive approach to language learning*. Oxford: Oxford University Press.
- Skehan, P. (2009). Modelling second language performance: Integrating complexity, accuracy, fluency, and lexis. *Applied Linguistics*, 30(4), 510-532.
- Skehan, P., & Foster, P. (2001). Cognition and tasks. In P. Robinson (Ed.), *Cognition and second language learning* (pp. 183-205). New York: Cambridge University Press.
- Taillefer, G. E. (1996). L2 reading ability: Further insight into the short-circuit hypothesis. *The Modern Language Journal*, 80, 461-477.
- Urquhart, A. H., & Weir, C. J. (1998). *Reading in a second language: Process, product, and practice*. New York: Longman.
- Wickens, C. D. (1992). *Engineering psychology and human performance*. New York, NY: Harper Collins.
- Yilmaz, Y. (2011). Task effects on focus on form in synchronous computer-mediated communication. *The Modern Language Journal*, 95(1), 115-132.
- Yoshimura, F. (2006). Does manipulating foreknowledge of output tasks lead to differences in reading behavior, text comprehension and noticing of language form? *Language Teaching Research*, 10(4), 419-434.

Application levels: Secondary

Jookyong Jung
Department of Culture, Communication and Media
University College London
Gower Street, London
WC1E 6BT
Phone: 44 (0)20-7679-2000
Email: jookyongjung14@ucl.ac.uk

Received on September 1, 2016

Reviewed on October 15, 2016

Revised version received on November 15, 2016