



Time-on-task as a Measure of Cognitive Load in TBLT

Jiyong Lee

Seoul National University, South Korea

Recent studies in task-based research have increasingly implemented ways to measure cognitive load in order to ensure that the tasks that were intended to be more complex placed greater cognitive load onto the learner, which in turn would lead to systematic changes in outcome measures. This study aims to introduce a more objective, indirect measure of cognitive load: time-on-task. In this study, 42 native speakers of English carried out three types of oral tasks that each had three levels of task complexity, operationalized as the number of elements. Cognitive load was measured by learner self-ratings, prospective time estimations, the dual task methodology, and time-on-task. Results of a series of correlational analyses and repeated-measures ANOVA showed that time-on-task, which can further be divided into time-on-planning and time-on-speech, proved to be a valid measure of cognitive load.

Keywords: time-on-task, time-on-planning, time-on-speech, cognitive load, task complexity, self-ratings, prospective time estimations, dual task methodology, L1 speech, Task-Based Language Teaching

Introduction

Defined as a multidimensional construct that represents the load that performing a particular task imposes on a learner's cognitive system (Paas & van Merriënboer, 1993), cognitive load in the field of cognitive psychology is not only considered to be a by-product of a learning process, but a major factor that determines the success of an instructional method (Paas, Tuovinen, Tabbers, & Van Gerven, 2003). Assuming that people's capacity for cognitive processing (i.e., working memory) is limited, the Cognitive Load Theory (CLT) concerns how cognitive load plays a crucial role in the learning of complex cognitive tasks. Since the theory's initial development in the 1980s, researchers have been actively searching for more valid and reliable methods to measure cognitive load (Brünken, Plass, & Leutner, 2003, van Merriënboer & Sweller, 2005).

In the field of Task-Based Language Teaching (TBLT), an increasing number of studies are employing various methods to measure cognitive load in order to obtain independent evidence that the task(s) that were intended by the researcher to be more complex actually place greater cognitive load on the learner, which in turn is expected to result in systematic changes in language performance. Measures such as self-ratings, time estimations, the dual task methodology, expert judgments, stimulate recall, and eye-tracking data have been used. In this study, a new method to measure cognitive load is suggested, namely time-on-task.

Measures of Cognitive Load in TBLT

According to Brünken, Plass, and Leutner (2003), cognitive load measures can be classified along two dimensions: objectivity (subjective/objective) and causal relation (direct/indirect). Objectivity concerns the distinction between subjective, self-reported data and objective observations of behavior, physiological conditions, or performance. Causal relation involves whether there is a direct link between the phenomenon observed by the measure and the attribute of interest. Table 1 shows a list of measures that can be used in task-based research, classified by these two dimensions.

TABLE 1
Methods for Measuring Cognitive Load in TBLT (adapted from Brünken, Plass, & Leutner, 2003)

Causal relation	Objectivity	
	Subjective	Objective
Indirect	Learner self-ratings of invested mental effort Expert judgments of invested mental effort Prospective time estimations	Physiological measures (e.g. heart rate, pupil dilation) Behavioral measures (e.g., eye-tracking, time-on-task) Linguistic outcome measures (e.g., complexity, accuracy, fluency)
Direct	Learner self-ratings of stress Learner self-ratings of difficulty of materials Expert judgments of difficulty of materials Stimulated recall protocols	Brain activity measures (e.g., fMRI) Dual-task performance

Due the growing demand for validation of task complexity manipulations, the following measures have been used in task-based research to measure cognitive load: 1) learner self-ratings (Baralt, 2013; Gilabert, Barón, & Llanes, 2009; Lee, 2018a, b, 2019; Malicka & Levkina, 2012; Robinson, 2001; Sasayama, 2016), 2) expert judgments (Lee, 2018a; Révész, Michel, & Gilabert, 2015; Révész, Sachs, & Hama, 2014), 3) time-on-task (Lee, 2018b), 4) time estimations (Baralt, 2013; Lee, 2018b, 2019; Malicka & Levkina, 2012; Sasayama, 2016), 5) dual task methodology (Lee, 2019; Révész, Michel, & Gilabert, 2015; Révész, Sachs, & Hama, 2014; Sasayama, 2016), 6) eye-tracking data (Révész, Sachs, & Hama, 2014), 7) stimulated recall protocols (Kim, Payant, & Pearson, 2015; Révész, Kourtali, & Mazgutova, 2017), and 8) writing measures such as fluency, pausing, and revision behaviors (Révész, Kourtali, & Mazgutova, 2017).

Using the data on learner self-ratings, time estimations, and the dual-task methodology from Lee (2019), the present study investigated whether time-on-task could be a valid measure of cognitive load. In the original study, 42 native speakers of English carried out three types of oral tasks that each had three levels of task complexity. A combination of statistical analyses revealed that increased task complexity led to greater ratings of perceived difficulty, mental effort, and stress, shorter prospective time estimations, and slower reaction times on the secondary task of the dual task method. Therefore, the present study attempted to find significant associations between these cognitive load measures and time-on-task, as well as significant task complexity and task-type effects and their interaction on the new measure.

Time-on-task in TBLT

Time-on-task has mainly been used in the fields of education and educational psychology, mostly used to predict or account for differences in student achievement. In cognitive psychology, time-on-task is claimed to be another indirect measure of cognitive load, and the time that learners spend learning with different forms of instructional interventions could be a result of different amounts of cognitive load induced by these variants (Brünken, Plass, & Leutner, 2003).

Only in very recent years has time-on-task been introduced to task-based research. As Lee (2018a) mentions, “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” (p. 28). In two studies of L2 writing, Lee (2018a, b) further divided time-on-task into three sub-measures: time-on-planning, time-on-writing, and time-on-whole task. In Lee (2018a), 83 Korean learners of English were assigned to either one of two conditions: the Closed condition, in which they carried out closed task versions (i.e., tasks that only had one acceptable solution), and the Open condition, in which they carried out open task versions (i.e., tasks that did not have a limited number of acceptable solutions). Two writing tasks were incorporated, whose task complexity was manipulated in terms of number of elements. Learner self-ratings, expert judgments, and time-on-task were used to assess how much cognitive load was placed on the learner. Regarding the time-on-task measure, results obtained from a series of mixed-effects models revealed not only task complexity effects on all three sub-measures of time-on-task, but also a significant interaction between task complexity and task closure on time-on-planning. Due to the significant interaction effect, a simple effects analysis was conducted, and it was found that task complexity effects on planning times were greater for those who carried out closed tasks. Put simply, participants spent a significantly longer time planning their writing when they performed complex tasks that had a limited number of acceptable solutions. This was attributed to the greater need to compare and contrast the information provided and meet the task requirements before being able to write.

In Lee’s (2018b) investigation on the combined effects of task complexity and L2 proficiency on L2 writing, 41 Korean learners of English were divided into three groups based on their English proficiency. They performed a writing task that had a simple and complex version, and the number of elements determined task complexity. Cognitive load was assessed by two measures: learner self-ratings and time-on-task. Results of a series of two-way repeated-measures ANOVA showed that regardless of proficiency group, task complexity had a significant main effect on all three sub-measures of time-on-task. In other words, participants spent a significantly longer time on the planning and writing phases of task performance when they carried out the complex task version.

These two task-based studies were able to show that time-on-task, especially if it is further divided into time-on-planning and time-on-writing, can be used as another measure of cognitive load. In the present study, borrowing native speaker data from Lee (2019) to get a more reliable window on task complexity effects, additional analyses were conducted in order to obtain independent evidence that time-on-task could be a valid, more objective indirect measure of cognitive load in TBLT.

Research Questions

In order to examine the validity of time-on-task as a measure of cognitive load in task-based research, the following research questions are addressed in the present study:

- RQ1. Can time-on-planning be employed as an indirect measure of cognitive load?
- RQ2. Can time-on-speech be employed as an indirect measure of cognitive load?

Methodology

Participants

The study recruited 42 adult native speakers of English (18 males, 24 females) enrolled at a university in the USA participated in the study, with their ages ranging from 19 to 41 years at the time of study ($M = 26.14$, $SD = 4.646$). Nine were undergraduate students, 25 held a Bachelor’s degree, seven held a

Master's degree, and one person held a doctorate degree. The rationale behind collecting only native speaker data was to ensure that the changes in cognitive load were a result of task complexity effects, and not by other confounding variables that are more likely to affect non-native speakers (see Lee, 2019).

Oral Tasks

In order to maximize generalizability of findings, three types of oral tasks were employed in the study. Task complexity was manipulated in terms of the number of elements, and each task had three levels of task complexity: least complex, mid-complex, and most complex. In a Map Task (MT), participants had to find the quickest route to a certain destination and tell an imaginary friend how to drive there. No-turn signs, one-way streets, closed roads, and construction sites were used to manipulate task complexity. In a Seating Arrangement Task (SAT), they had to recommend the best seating arrangement for a group of imaginary people with certain specific preferences. The number of elements in this task involved the number of people and their preferences. In a Car Accident Task (CAT), after watching a video clip of an actual car accident three times, participants gave a report on it while pretending to be a news reporter. Here, task complexity was manipulated in terms of the number of people and cars involved in the accident. Based on the results of a pilot test, a decision was made to give as much as two minutes for the planning phase of the MT and CAT, and as much as five minutes for that of the SAT.

Measures of Cognitive Load

In order to find out whether the tasks that were designed to be more complex actually placed greater a cognitive load on the learner, triangulation of various methods was employed in the original study: learner self-ratings, prospective time estimations, and the dual task method. Because the use of prospective time estimations made it necessary to record the actual time it took for participants to prepare and make their speeches for each task version, the present study examined how time spent on planning and speaking during task performance correlated with the other measures of cognitive load, and how increasing task complexity had an effect on both time-on-planning and time-on-speech.

Regarding learner self-ratings, a nine-point Likert scale was used to ask participants about their perceptions of (i) overall task difficulty, (ii) level of mental effort they thought was required for task performance, and (iii) level of stress they felt during task performance. On the same questionnaire, they were also asked to estimate the time it took for them to prepare and make their speech, separately. Seeing that participants were fully aware that they had a time limit for preparing their speeches, with some using the entirety of that time during the planning phase, separating these two times was inevitable. These data were later used to calculate prospective time estimations, by dividing the estimations by the actual time participants spent on the planning or speech phases of the task. For the dual-task methodology, while performing the primary oral tasks on a laptop in front of them, participants were required to perform a secondary simple choice reaction task simultaneously: reacting to color changes on the laptop screen that switched from either white to green or white to red every 2.5 seconds. Every time the screen color changed, participants had to react as quickly and accurately as possible by pressing either the left or right shift key. This dual-task methodology was run through DMDX, and participants' error rates and reaction times were used to measure cognitive load. Accuracy was calculated by the ratio of correct responses to the total number of screen color changes, and only correct responses were considered for reaction time.

Procedure

Participants had a one-to-one meeting with the researcher for approximately one hour. Upon completion of a language background questionnaire (adapted from Ellis, 2011), they carried out a series of practice items to familiarize themselves with the actual tasks. Participants were required to carry out nine task versions in total, and the order of the tasks was pseudo-randomized in order to avoid sequencing

effects, as shown in Table 2. After completing each task version, participants filled out a questionnaire that included questions related to the cognitive load of the task.

TABLE 2
Sample of Task Randomization

Participant	Task sequence								
	1 st	2 nd	3 rd	4 th	5 th	6 th	7 th	8 th	9 th
1	CAT 2	SAT 2	MT 2	SAT 1	MT 1	CAT 3	MT 3	SAT 3	CAT 1
2	SAT 2	MT 2	CAT 1	SAT 1	MT 3	CAT 3	MT 1	CAT 2	SAT 3
3	MT 3	SAT 1	CAT 2	MT 1	CAT 3	SAT 2	MT 2	CAT 1	SAT 3
4	MT 1	SAT 2	CAT 1	SAT 3	CAT 2	MT 2	SAT 1	CAT 3	MT 3
5	MT 3	SAT 3	CAT 3	MT 2	CAT 2	SAT 1	MT 1	CAT 1	SAT 2
6	CAT 1	MT 3	SAT 2	CAT 3	SAT 1	MT 1	CAT 2	MT 2	SAT 3

Note. 1=least complex; 2=mid-complex; 3=most complex

Analyses

In order to find out whether there was a relationship between the cognitive load measures used in the original study and two time-on-task measures, i.e., time-on-planning and time-on-speech, two types of correlation analyses were conducted. First, a series of Spearman rank correlations were computed to see how time-on-task was related to learner self-ratings, time judgments, and the accuracy and reaction time of the secondary task. Next, a partial correlation analysis was computed on the data by controlling for task complexity and task-type. Finally, A 3 × 3 repeated-measures ANOVA was then further computed to see if increasing task complexity led to the desired changes in cognitive load. For this analysis, the two time-on-task measures served as the dependent variables, and task complexity and task-type served as the independent variables with three levels each (task complexity: least complex, mid-complex, most complex; MT, SAT, and CAT).

Results

Table 3 presents the descriptive statistics for the four measures of cognitive load employed in the study: time-on-task, learner self-ratings, prospective time estimations, and the dual-task methodology, respectively. As shown in the table, the actual times spent on both planning and speech increase with greater task complexity (with the exception of time-on-planning on the most complex version of the CAT), regardless of task-type. Learner self-ratings also appear to have a linear relationship with task complexity, with ratings rising as task complexity is increased. On the other hand, the other measures do not appear to show a consistent pattern.

TABLE 3
Descriptive Statistics for Cognitive Load Measures

	Map			Seating Arrangement			Car Accident		
	1	2	3	1	2	3	1	2	3
Time-on-planning(s)	67.83 (34.69)	90.83 (33.37)	92.17 (31.84)	116.62 (47.88)	182.17 (59.31)	279.33 (37.53)	24.62 (27.73)	27.79 (25.15)	23.95 (25.99)
Time-on-speech (s)	30.43 (8.96)	44.98 (15.60)	60.50 (21.15)	43.55 (15.11)	59.80 (25.14)	91.55 (40.53)	28.23 (12.74)	38.78 (17.77)	51.40 (21.60)
Perceived difficulty	3.71 (1.38)	5.02 (1.73)	6.24 (1.49)	4.57 (1.88)	5.88 (1.38)	7.95 (1.09)	4.38 (1.55)	5.17 (1.58)	6.86 (1.37)
Mental effort	3.67 (1.52)	5.05 (1.90)	5.86 (1.62)	4.50 (1.81)	5.79 (1.52)	7.75 (1.03)	4.33 (1.68)	5.31 (1.81)	6.62 (1.45)
Stress	3.43 (1.38)	4.57 (1.82)	4.88 (1.95)	4.02 (1.98)	4.67 (1.72)	6.55 (1.89)	3.71 (1.40)	4.48 (1.70)	5.40 (1.80)
Planning estimation	1.25 (0.86)	1.29 (0.70)	1.18 (0.57)	1.06 (0.49)	1.09 (0.44)	0.98 (0.21)	7.49 (8.79)	8.51 (9.44)	10.31 (10.73)
Speech estimation	1.59 (0.89)	1.72 (0.95)	1.37 (0.82)	1.52 (0.79)	1.53 (0.79)	1.26 (0.63)	1.86 (1.23)	1.64 (0.94)	1.58 (0.96)
Accuracy	0.97 (0.07)	0.96 (0.06)	0.96 (0.06)	0.97 (0.08)	0.96 (0.06)	0.95 (0.06)	0.97 (0.07)	0.98 (0.04)	0.95 (0.07)
Reaction time (ms)	1590.27 (921.10)	1725.23 (839.97)	1756.67 (761.97)	1590.92 (941.28)	1742.63 (720.68)	2428.02 (1514.83)	1385.98 (652.15)	1185.54 (358.04)	1448.66 (586.23)

Note. Complexity 1 indicates least complex, 2 mid-complex, and 3 most complex.

As shown in Tables 4 and 5, results of a series of Spearman rank correlations revealed that time-on-planning and time-on-speech had significant correlations with other measures of cognitive load. More specifically, time-on-planning had significant positive associations with learner self-ratings, and significant negative associations with prospective time estimations. On the other hand, time-on-speech had significant negative correlations with self-ratings, and negative associations with prospective time estimations. When looking at the correlation coefficients, moderate to relatively strong correlations were found between time-on-task and learner self-ratings. Concerning time estimations, because they were calculated by the ratio of subjective to objective duration, it was expected that time estimations would decrease as time-on-task increased. On the other hand, no significant relationships were found between time-on-task and the results from the dual-task methodology.

TABLE 4
Spearman's rho Correlation between Time-on-planning and Other Cognitive Load Measures

Task-type Complexity	Map			Seating Arrangement			Car Accident		
	1	2	3	1	2	3	1	2	3
Perceived difficulty	.415** (0.006)	.462** (0.002)	.495** (0.001)	0.197 (0.210)	.494** (0.001)	0.119 (0.460)	0.197 (0.212)	0.238 (0.130)	0.153 (0.332)
Mental effort	0.296 (0.057)	.558** (0.00)	.305* (0.050)	0.215 (0.172)	.394** (0.010)	0.058 (0.723)	.305* (0.049)	0.288 (0.064)	0.085 (0.592)
Stress	0.106 (0.503)	.367* (0.017)	.350* (0.023)	0.236 (0.132)	0.263 (0.093)	0.033 (0.835)	0.026 (0.871)	0.298 (0.055)	0.087 (0.583)
Planning estimation	-.565** (0.000)	-0.225 (0.151)	0.223 (0.156)	0.257 (0.971)	-.491** (0.001)	0.002 (0.989)	-.600** (0.000)	-.770** (0.000)	-.659** (0.000)
Speech estimation	.400* (0.012)	0.258 (0.113)	0.224 (0.170)	0.257 (0.115)	0.033 (0.843)	0.286 (0.073)	-0.073 (0.659)	0.110 (0.499)	0.072 (0.659)
Accuracy	0.052 (0.750)	0.044 (0.789)	-0.076 (0.651)	0.213 (0.194)	-0.025 (0.878)	-0.042 (0.801)	-0.113 (0.492)	-0.247 (0.130)	-0.152 (0.348)
Reaction time (ms)	0.028 (0.865)	0.284 (0.076)	0.034 (0.838)	0.112 (0.497)	0.169 (0.311)	0.083 (0.614)	0.126 (0.438)	0.315 (0.054)	0.050 (0.763)

Note. ** Correlation is significant at the 0.01 level (2-tailed).

* Correlation is significant at the 0.05 level (2-tailed).

TABLE 5
Spearman's rho Correlation between Time-on-speech and Other Cognitive Load Measures

Task-type Complexity	Map			Seating Arrangement			Car Accident		
	1	2	3	1	2	3	1	2	3
Perceived difficulty	0.246 (0.172)	0.174 (0.282)	-.419** (0.007)	0.192 (0.236)	0.246 (0.126)	-.331* (0.040)	0.173 (0.285)	0.251 (0.119)	0.217 (0.179)
Mental effort	0.261 (0.104)	0.123 (0.450)	-0.237 (0.141)	0.084 (0.607)	0.136 (0.403)	-.351* (0.031)	0.213 (0.188)	0.266 (0.097)	0.156 (0.337)
Stress	0.138 (0.396)	0.098 (0.547)	-0.249 (0.121)	0.104 (0.521)	0.246 (0.125)	-0.073 (0.653)	0.200 (0.215)	.362* (0.022)	0.114 (0.485)
Planning estimation	0.298 (0.066)	0.099 (0.542)	-0.133 (0.486)	-0.057 (0.727)	-0.144 (0.376)	-0.256 (0.111)	-.362* (0.024)	-.449** (0.004)	-0.139 (0.399)
Speech estimation	-.320* (0.047)	-0.226 (0.167)	-.319* (0.048)	0.009 (0.957)	-0.200 (0.222)	-0.177 (0.276)	-.404* (0.011)	-0.292 (0.067)	-.441* (0.004)
Accuracy	0.061 (0.708)	0.129 (0.433)	-0.230 (0.165)	0.196 (0.231)	-0.046 (0.782)	0.041 (0.803)	0.017 (0.917)	-0.239 (0.142)	-0.095 (0.559)
Reaction time (ms)	0.205 (0.206)	.491** (0.001)	0.021 (0.902)	-0.041 (0.806)	-0.118 (0.479)	0.005 (0.977)	-0.108 (0.505)	-0.045 (0.787)	-0.077 (0.641)

Note. ** Correlation is significant at the 0.01 level (2-tailed).
* Correlation is significant at the 0.05 level (2-tailed).

A partial correlation was also computed in order to test the associations between time-on-task and other measures of cognitive load while controlling for task complexity and task-type. Results showed that even when we control for task complexity and task-type, time-on-planning and time-on-speech are significantly correlated with the majority of the other cognitive load measures. Results are presented in Table 6.

Table 6
Partial Correlation Results

Measures	Learner self-ratings			Time estimations		Dual-task method	
	Difficulty	Mental effort	Stress	Planning	Speech	Accuracy	Reaction time
Time-on-planning	0.432** (0.000)	0.385** (0.000)	0.340** (0.000)	-0.379** (0.000)	0.026 (0.640)	-0.055 (0.318)	0.278** (0.000)
Time-on-speech	0.204** (0.000)	0.193** (0.000)	0.199** (0.000)	-0.260** (0.000)	-0.245** (0.000)	-0.023 (0.684)	0.135 (0.014*)

Note. ** Correlation is significant at the 0.01 level.
* Correlation is significant at the 0.05 level.

Finally, a 3 × 3 repeated-measures ANOVA was conducted on time-on-planning and time-on-speech separately, with task complexity and task-type as the independent variables. Concerning time-on-planning, Mauchly's Test of Sphericity indicated that the assumption of sphericity had been violated for the interaction between task complexity and task-type, $\chi^2(9) = 39.455, p < .0001$. As a result, a Greenhouse-Geisser correction was used. Results revealed that the main effects for task complexity and task-type were significant, $F(2, 82) = 98.559, p < .001, partial \eta^2 = 0.706, F(2, 82) = 630.505, p < .001, partial \eta^2 = 0.939$, respectively. A significant interaction between the two variables was also found, $F(2.761, 113.188) = 76.717, p < .001, partial \eta^2 = 0.652$. Simple main effects analysis showed that task complexity effects were significantly greater for certain tasks, especially the Seating Arrangement Task (SAT). In the SAT, time-on-planning for all levels of task complexity were significantly different from each other ($p < .001$), and in the CAT, time-on-planning for the least complex task version was significantly lower than those for the mid- and most complex versions ($p < .001$). Figure 1 illustrates such patterns.

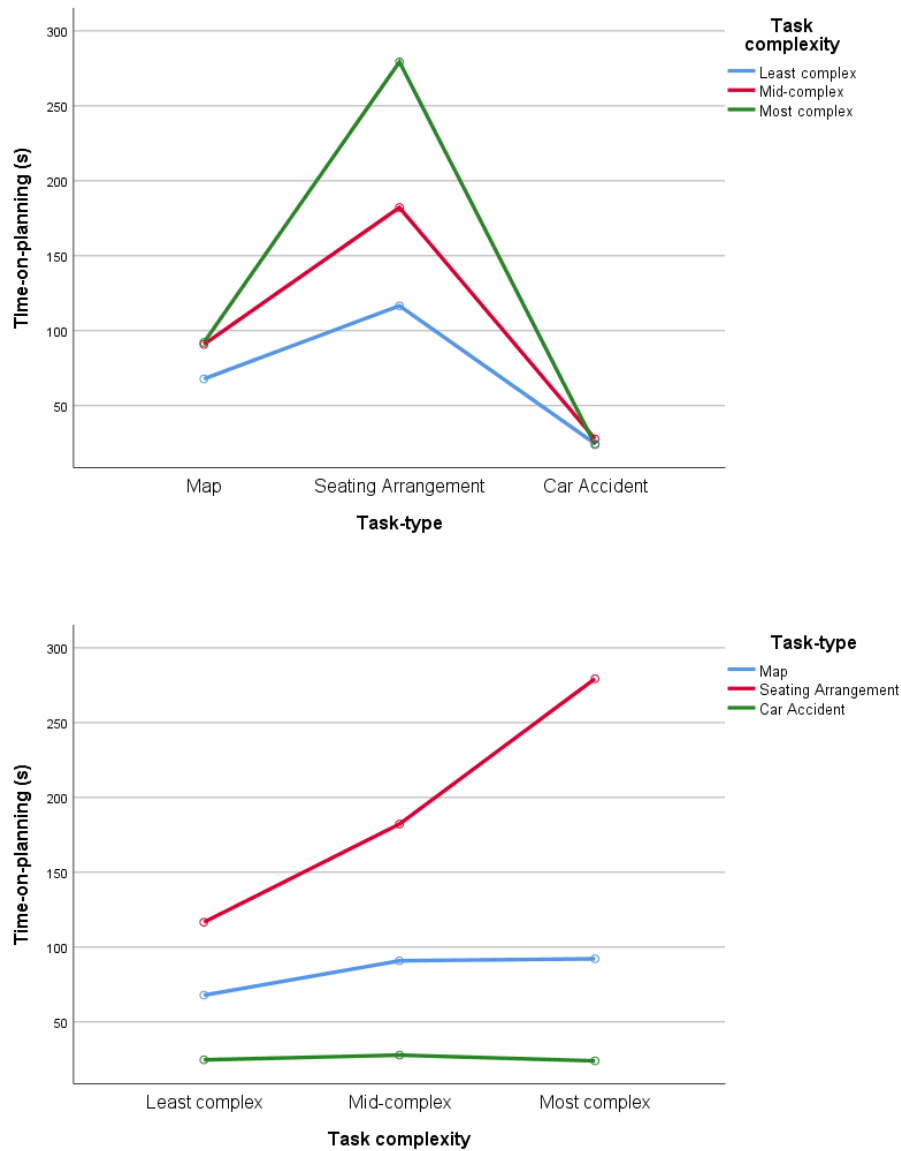


Figure 1. Task complexity and task-type effects on time-on-planning.

Regarding time-on-speech, Mauchly's Test of Sphericity indicated that the assumption of sphericity had been violated for task complexity and its interaction with task-type, $\chi^2(2) = 22.260, p < .0001$, and $\chi^2(2) = 39.827, p < .0001$, respectively. Therefore, a Greenhouse-Geisser correction was used. The inferential statistics for time-on-speech are very similar to those for time-on-planning. The main effects for task complexity and task-type were found to be significant, $F(1.386, 54.041) = 112.774, p < .001, partial \eta^2 = 0.743, F(2, 78) = 63.149, p < .001, partial \eta^2 = 0.618$, respectively. More importantly, the interaction between the task complexity and task-type were found to have a significant effect on time-on-speech, $F(2.483, 96.854) = 10.602, p < .001, partial \eta^2 = 0.214$. Simple main effects analysis revealed that for each level of task complexity, time-on-speech was significantly greater when participants carried out the Seating Arrangement Task, in comparison to the other two tasks, ($p < .001$). Such patterns can be seen in Figure 2.

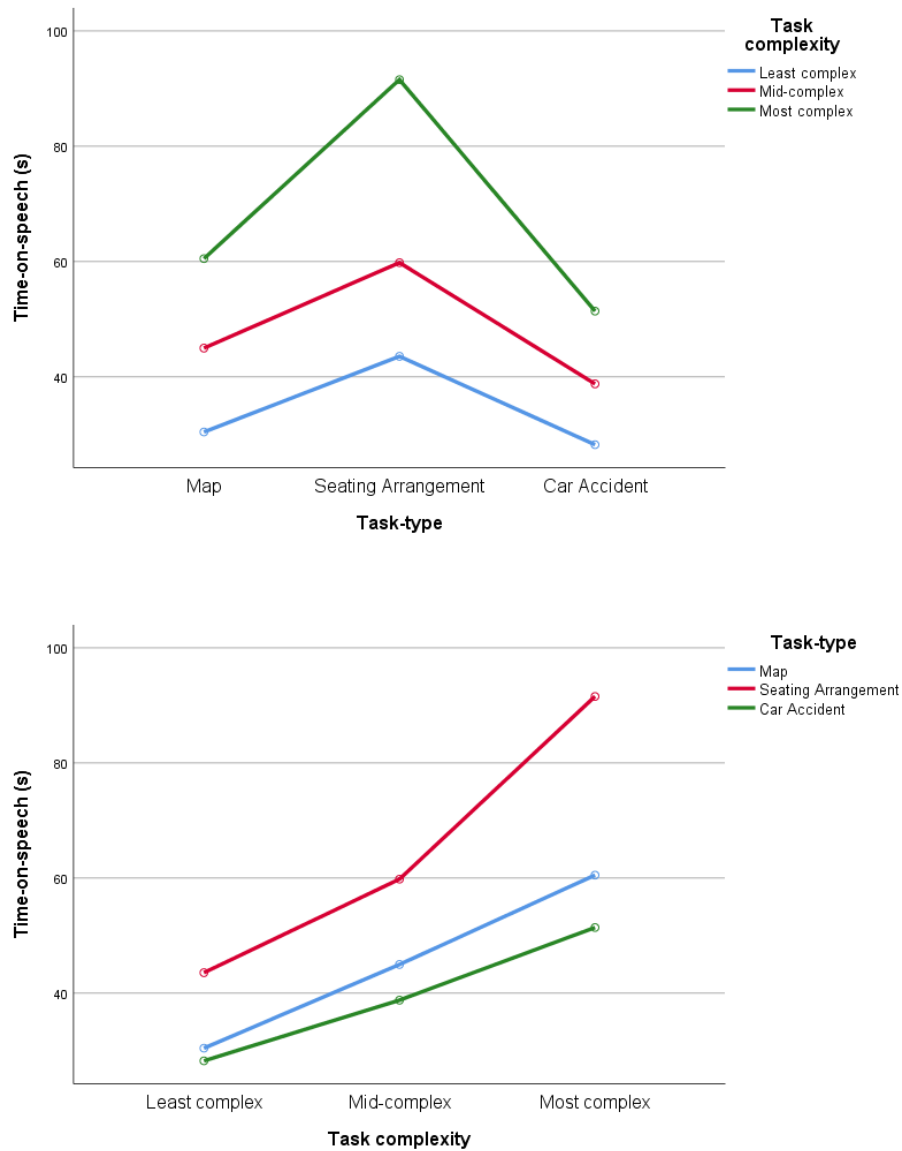


Figure 2. Task complexity and task-type effects on time-on-speech.

Discussion and Conclusion

The purpose of the present study was to find out whether time-on-task could be used as a measure of cognitive load in task-based research. To date, recent studies of task complexity effects have employed various methods in order to see whether the tasks that were intended to be more complex placed greater cognitive load on the learner, which in turn was expected to lead to the desired changes in language performance, mostly in terms of linguistic complexity, accuracy, and fluency. Using data from Lee (2019), which employed learner self-ratings, prospective time estimations, and the dual task methodology as cognitive load measures, the present study examined the validity of the time-on-task measure, by looking into time-on-planning and time-on-speech separately.

RQ 1. Can time-on-planning be employed as an indirect measure of cognitive load?

As participants were aware that there was a time limit for the planning phase of each task (up to two minutes for the MT and CAT, and up to 5 minutes for the SAT), it was considered necessary to separate time-on-planning from time-on-speech in order to avoid any confounding variables. This extra step proved to be essential, for statistical analyses on this measure obtained slightly different results from those on time-on-speech. In fact, time-on-planning showed a greater number of significant associations with the other cognitive load measures than time-on-speech did. In general, for the Map Task (MT) and the Seating Arrangement Task (SAT) in particular, but less so for the Car Accident Task (CAT), high self-ratings of perceived difficulty, mental effort, and stress were associated with longer time spent on planning. As for prospective time estimations, due to how time estimations were calculated, a negative association between time estimations and time-on-planning was expected. This proved to be the case, especially for the CAT that showed significant correlations were found for all levels of task complexity. On the other hand, the accuracy and reaction time results obtained from the dual task methodology did not show any significant correlation with time-on-planning.

Concerning the results of the repeated-measures ANOVA, time-on-planning was found to be significantly affected by the interaction between task complexity and task-type. Put simply, task complexity effects on the time it took for participants to plan their speech differed significantly depending on the type of task they performed. The SAT was found to take much longer to plan for than the MT or CAT, which most likely indicates that the SAT is intrinsically more challenging than the other two tasks. This is not surprising, as the nature of the SAT required participants to come up with a plan to seat imaginary guests with certain requirements around a circular table. Unlike the MT or the CAT, in which participants could find a route relatively quickly or draw upon their memory of watching the video clip of a car accident, the SAT required more active storing and processing information, which showed particularly in their self-ratings of the most complex version of the SAT.

Such results are in line with those from Lee (2018a, b). As Lee (2018b) claimed, time-on-task has been used in the field of TBLT as a measure of cognitive load only in recent years, and the capability to make a distinction between time-on-planning and time-on-speech/writing in a more object method is a major advantage. With the appropriate software tools, it is also much easier or less complicated than prospective time estimations or the dual task methodology, the latter of which is quite difficult to use when the language mode of interest is writing.

RQ 2. Can time-on-speech be employed as an indirect measure of cognitive load?

Similar to the results of time-on-planning, but to a lesser extent, time-on-speech was also found to correlate significantly with some of the other cognitive load measures. However, unlike time-on-planning, it had significant *negative* relationships with the self-ratings, which indicated that the higher participants gave their ratings on perceived difficulty, mental effort, and stress, especially for the most complex task version, the shorter time they spent during the speech part of task performance. On the other hand, the significant negative associations between time-on-speech and prospective time estimations were to be expected, because time estimations are calculated in such a way that longer time spent on speech inevitably results in a greater denominator, and thus a lower time estimation.

In an attempt to account for the reversed V-shape pattern shown in a syntactic complexity outcome measure, Lee (2019), from which the data of learner self-ratings in the present study borrowed, claimed that participants may have “perceived the most complex task versions to be so complex that they short-circuited the task and simplified it, either intentionally ignoring the added elements or unintentionally not being able to notice them” (pp. 533-534). For the MT, whose perceived difficulty rating for the most complex version showed a negative correlation with time-on-speech, many participants did not fully explain why they chose one street/turn over others. In the case of the SAT, whose perceived difficulty and required mental effort ratings negatively correlated with time-on-speech, many did not provide reasons

for why they placed one person in a certain seat. Nonetheless, these participants could still complete the task, albeit in a minimally satisfactory way. As a result, most likely due to a lack of sufficient explanations, many participants spent less time on speech than was expected, leading to a negative association between time-on-speech and learner self-ratings of the complex task versions (see Lee, 2019, Table 8 for a comparison between a task that has been simplified compared to one that has not).

Results obtained from the repeated-measures ANOVA on time-on-speech nearly mirrored those on time-on-planning data. The significant interaction between task complexity and task-type impacted time-on-speech such that task complexity had a significantly greater effect on the time it took for participants to complete the speech phase of the SAT than the MT or the CAT. In other words, the magnitude of the effects that task complexity had on the time spent on both planning and speech differed depending on the type of the task that was carried out.

Conclusion

The present study sought to investigate whether time-on-task, more specifically, time-on-planning and time-on-speech, as a valid, more objective measure of cognitive load in task-based research. Independent evidence was obtained that the measure was similar to other measures of cognitive load, and that it was also affected by task complexity and task-type effects and their interaction. Not only is it less subjective than learner self-ratings and prospective time estimations, which rely on personal judgments that are subject to inconsistencies and other reliability issues, it is easier to administer than other measures such as the dual-task methodology, which is only possible to use with the right equipment and language mode. However, this study does not claim that time-on-task should be the *only* method to measure cognitive load. It is merely trying to suggest an alternative method that can be used in combination with others in order to increase the validity of the research.

For future research on the time-on-task measure, it would be interesting to compare differences between speech and writing, as the two processes vary greatly. Furthermore, the present study analyzed data collected from native English speakers, and further investigations could also examine the time that L2 learners spend on the planning and production phases of tasks with varying degrees of task complexity. Because the main effect for task-type and its interaction with task complexity were found to be significant, tasks should be designed and implemented with caution, bearing in mind that some participants, whether they are native speakers or not, show a tendency to simplify tasks if they perceive it to be too complex.

Acknowledgement

I would like to express my gratitude to Dr. Michael Long for his guidance. I would also like to thank Drs. Steve Ross and Nan Jiang for their support. My gratitude also extends to the anonymous reviewers.

The Author

Jiyong Lee completed her Ph.D. in Second Language Acquisition at the University of Maryland, USA in December 2018. She is currently a lecturer at Seoul National University and Inha University, South Korea.

College English Program
Seoul National University
Building 3, #210
1 Gwanak-ro, Gwanak-gu, Seoul, 151-742

Email: jlee0123@terpmail.umd.edu

References

- Baralt, M. L. (2013). The impact of cognitive complexity on feedback efficacy during online versus face-to-face interactive tasks. *Studies of Second Language Acquisition*, 35, 689–725.
- Brünken, R., & Leutner, D. (2001). Aufmerksamkeitsverteilung oder aufmerksamsamkeitsfokussierung? Empirische Ergebnisse zur “split-attention-hypothese” beim lernen mit multimedia [Split of attention or focusing of attention? Empirical results on the split-attention-hypothesis in multimedia learning]. *Unterrichtswissenschaft*, 29, 357–366.
- Brünken, R., Plass, J. L., & Leutner, D. (2003). Direct measurement of cognitive load in multimedia learning. *Educational Psychologist*, 38(1), 53–61.
- 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.
- Kim, Y., Payant, C., & Pearson, P. (2015). The intersection of task-based interaction, task complexity, and working memory. *Studies in Second Language Acquisition*, 37, 549–581.
- Lee, J. (2018a). *The interactive effects of task complexity, task condition, and cognitive individual differences on L2 writing* (Doctoral dissertation). Retrieved from <http://hdl.handle.net/1903/21755>.
- Lee, J. (2018b). The effects of task complexity and L2 proficiency on L2 written performance. *The Journal of Asia TEFL*, 15(4), 945-958.
- Lee, J. (2019). Task complexity, cognitive load, and L1 speech. *Applied Linguistics*, 40(3), 506-539.
- Malicka, A., & Levkina, M. (2012). Measuring task complexity: Does L2 proficiency matter. In A. Shehadeh & C. A. Coombe (Eds.), *Task-based language teaching in foreign language contexts: Research and implementation* (pp. 43-66). Philadelphia, PA: John Benjamins Publishing.
- Paas, F., & van Merriënboer, J. (1993). The efficiency of instructional conditions: An approach to combine mental-effort and performance measures. *Human Factors*, 35, 737–743.
- Paas, F., Tuovinen, J., Tabbers, H., & van Gerven, P. (2003). Cognitive load measurement as a means to advance cognitive load theory. *Educational Psychologist*, 38, 63–71.
- Révész, A., Kourtali, N., & Mazgutova, D. (2017). Effects of task complexity on L2 writing behaviors and linguistic complexity. *Language Learning*, 67, 208–241.
- 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, 1–35.
- 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, 615–650.
- Robinson, P. (2001). Task complexity, cognitive resources, and syllabus design: A triadic framework for examining task influences on SLA. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 287-318). New York, NY: Cambridge University Press.
- Sasayama, S. (2016). Is a ‘complex’ task really complex? Validating the assumption of cognitive task complexity. *The Modern Language Journal*, 100(1), 231-254.
- van Merriënboer, J., & Sweller, J. (2005). Cognitive load theory and complex learning: Recent developments and future directions. *Educational Psychology Review*, 17(2), 147-177.