English Teaching, Vol. 77, No. 4, Winter 2022, pp. 49-69
DOI: https://doi.org/10.15858/engtea.77.4.202212.49
http://journal.kate.or.kr

Effects of Task Complexity on EFL Learners' Writing Behaviors and Performance

Yoonseo Kim *

Kim, Yoonseo. (2022). Effects of task complexity on EFL learners' writing behaviors and performance. *English Teaching, 77*(4), 49-69.

This study explored how task complexity, writing behaviors (i.e., pausing and revision behaviors), and writing performance (i.e., task completion, coherence and cohesion, language use, and expression and tone) influence and relate to each other. Thirty advanced-level Korean EFL undergraduates completed writing tasks differing in complexity. A combination of keystroke logging and stimulated recall interview was employed. It was found that the simple task group showed a greater number of pauses and revisions related to lower-order writing processes, whereas the complex task group showed longer pauses related to higher-order writing processes. While task complexity had no influence on writing performance, writing behaviors revealed significant relationships with text quality. In the simple task group, pause length and revision were negatively related to writing scores, whereas pause frequency revealed mixed results. In the complex task group, consistent negative relations were found between pausing behaviors and text quality, and fewer revisions were related to better scores in expression and tone.

Key words: second language writing, writing process, task complexity, pausing, revision, writing performance, keystroke logging

*Author: Yoonseo Kim, Graduate Student, Department of English Education, Yonsei University; 50, Yonsei-ro, Seodaemun-gu, Seoul, Republic of Korea; Email: yoonseokim@yonsei.ac.kr

Received 12 September 2022; Reviewed 1 November 2022; Accepted 1 December 2022

1. INTRODUCTION

In the field of second language acquisition, Task-based language teaching (TBLT) has attracted considerable attention from both instructors and researchers. This is because tasks not only engage learners in practicing authentic communication (Nunan, 2004) but also provide a suitable platform where meaning-based and form-based instruction can be integrated (Révész, Kourtali, & Mazgutova, 2017). To maximize the benefits of TBLT, understanding the effects of various task types (e.g., task complexity, task condition, and task difficulty) is crucial. Among them, task complexity is especially significant in that it provides insights into curricular decisions such as task sequencing and grading (Robinson, 2001a). Despite its importance, compared to studies on speaking skills, relatively scarce attention has been paid to the effects of task complexity on writing skills.

The effects of task complexity on L2 writing can be investigated in terms of writers' behaviors and performance. First, investigating writing behaviors using a keystroke logging program can be useful in understanding underlying cognitive processes of L2 writing, predicting difficulties in learning, and testing writing models. Specifically, pausing behaviors provide measurable data indicating where, when, and how long writers planned or revised texts (Schilperoord, 1996). Also, revision behaviors reveal complex cognitive activities which involve reading, editing, and monitoring (Révész, Michel, & Lee, 2019). Therefore, investigating pausing and revision behaviors will provide insights into cognitive processes which L2 writers undergo (Barkaoui, 2019).

Second, writing performance is also a good indicator of the effects of task manipulation on L2 writing, especially when a combination of research methods is employed. Nevertheless, few studies have explored the effects of task complexity on L2 writers' performance, and fewer studies have looked into the subcomponents of writing performance (i.e., task completion, coherence and cohesion, language use, and expression and tone). Thus, using a mixed-method approach of keystroke logging and stimulated recall, the current study explores how task complexity, pausing and revision behaviors, and subcomponents of writing performance influence and relate to each other.

2. LITERATURE REVIEW

2.1. Theoretical Background

Information processing model, which explains the human mind in comparison with a computer, was set as a theoretical basis of the present study (Baddeley, 1986). According to this model, people receive input via sensory register and transfer the information to

working memory (i.e., a mental storage where information is held in an accessible form) through perception and attention. The encoding process allows information in working memory to be retained in long-term memory, an infinite but unconscious memory. To utilize information, knowledge must be available in working memory, however, due to the limitation of its capacity, excessive information can impose cognitive burden on the mind.

Indeed, L2 writing is a complex process which requires thinking, language production, and translation (Kellogg, Whiteford, Turner, Cahill, & Merlens, 2013). If writers try to manage information beyond their working memory capacity, they will likely experience cognitive overload. Cognitive overload is influenced by how tasks are designed, and task design can be manipulated based on Robinson's (2003) Triadic Framework. According to this framework, there are three dimensions of a task (i.e., task condition, task difficulty, and task complexity), and all the dimensions exert different impacts on task performance. Task condition refers to interactional factors that can be identified by needs analysis, and task difficulty is how learners perceive a task, which is affected by task complexity, individual abilities, and affective variables. Task complexity, which is the focus of the current study, is defined as information processing demands of a task on attention, memory, and reasoning.

Robinson (2003) explains that task complexity is composed of two dimensions (i.e., resource-directing and resource-dispersing dimension), and these dimensions belong to separate resource pools affecting linguistic output differently. The dimensions are expressed using "±" features, and task complexity can be raised by selecting "-" features in a certain task. On the one hand, resource-directing factors impose conceptual demands on learners and are divided into ±here-and-now, ±few elements, and ±no reasoning demands. When task complexity is raised in this respect, accuracy and complexity of learners' linguistic production during tasks increase while fluency decreases. This is because the manipulation can influence how cognitive resources are assigned to language form (Ishikawa, 2008). On the other hand, resource-dispersing factors impose procedural demands on learners and are classified into ±planning time, ±prior knowledge, and ±single task. Raising task complexity along these factors can deteriorate accuracy, complexity, and fluency of learners' linguistic production. In the current study, task complexity was manipulated using one of the resource-directing factors, ±few elements. It was expected that this manipulation will lead to higher complexity and accuracy in writing performance.

Along with backgrounds on linguistic performance, backgrounds concerning writing behaviors have to be addressed. Specifically, previous studies indicate where writers pause and revise manifest reasons for the behaviors. First, pauses at higher-level textual units (e.g., between clauses and sentences) were related to higher-order writing processes (e.g., planning organization and contents), whereas pauses at lower-level textual units (e.g., within words and between words) were associated with lower-level writing processes (e.g.,

retrieving lexis) (Schilperoord, 1996). Second, regardless of location, revision behaviors were more concerned with translation rather than planning processes. However, while lower-level revisions were primarily related to linguistic processes, higher-level revisions were associated with not only linguistic but also planning processes (Révész et al., 2019).

2.2. Previous Empirical Research

2.2.1. Task complexity and L2 writers' behaviors

A study that initiated the investigation of the link between task complexity and L2 writers' behaviors was Spelman Miller (2000). Using keystroke logging data of two essays differing in task complexity, this study revealed that writers paused longer at higher-level textual units and paused most frequently between intermediate constituents. The result indicated that planning is associated with pauses between clauses and sentences. However, task complexity was found to have no influence on pausing and fluency.

In the same vein, Révész et al. (2017) explored the effects of task complexity on L2 writers' fluency, pausing, and revision. Regardless of task complexity, pauses at higher textual units were associated with higher-level writing processes. Also, the simple task writers showed more pauses between sentences, which were related to improved linguistic complexity, whereas the complex task writers showed less revisions below word due to the greater cognitive demands of the complex task.

Most recently, Conijn, Roeser, and van Zaanen (2019) explored how cognitive demands of writing tasks affect pausing and revision behaviors. It was found that writing behaviors differed across writing tasks, which confirmed that an academic task had higher task complexity than an email writing task and that keystroke features were susceptible to different cognitive demands of tasks. In conclusion, although keystroke data have been used to detect the relationship between task complexity and writing processes, the relation still remains vague, and the number of studies is scarce. To supplement the research gap, it is worth exploring how keystroke features are influenced by task complexity.

2.2.2. Task complexity and L2 writers' performance

Previous studies which investigated the effects of task complexity on L2 writers' performance are twofold: resource-dispersing and resource-directing dimensions. To begin with, despite few contradicting outcomes (e.g., Ong & Zhang, 2010), findings concerning resource-dispersing factors are mostly consistently in line with Robinson (2001b), who suggested that raising task complexity along resource-dispersing factors would have a negative impact on accuracy, complexity, and fluency (e.g., Sadeghi & Mosalli, 2012).

However, studies that explored the effects of resource-directing factors show inconsistent results. Specifically, findings of Rahimi (2019) contradict Robinson's (2001b) assumption that increasing task complexity would improve accuracy and complexity. Also, Robinson's (2001b) hypothesis that raising task complexity would harm fluency was debunked by Salimi, Dadaspour, and Asadollahfam (2011). In addition, previous studies reported that task complexity has no influence on fluency (Ong & Zhang, 2010), lexical complexity (Ong & Zhang, 2010), and accuracy (Rahimpour & Hosseini, 2010).

Nevertheless, several studies are supportive of Robinson's (2001b) assumption on resource-directing factors. Specifically, Kuiken and Vedder (2007) found that increased task complexity results in more accurate texts. Similarly, task complexity led to improved accuracy in Kuiken and Vedder (2008) although it had no influence on syntactic complexity. Most recently, Abdi Tabari (2022) found that writers showed higher complexity and fluency under a less complex condition. Thus, considering such varied results, a further study is necessary to explore the effects of resource-directing factors on writing performance.

2.2.3. L2 writers' behaviors and performance

The majority of studies which explored the relationship between L2 writers' behaviors and performance has focused on overall quality of text production. In terms of pause duration, Xu and Ding (2014) and Xu and Qi (2017) reported that pause duration at the prewriting stage positively correlated with final text quality, whereas Alves and Limpo (2015) found that shorter pauses were related to better text quality. However, according to Spelman Miller, Lindgren, and Sullivan (2008), only a weak association was found between pause duration and writing performance.

Regarding p-bursts (bursts which were terminated by pauses), Alves and Limpo (2015) found that longer bursts are related to higher fluency. In the same vein, Choi and Deane (2021) reported that writers who showed consistent within-burst pauses and shorter within-word pauses gained higher essay scores. In contrast, Deane (2014) concluded that the relationship between burst span and text quality (i.e., fluency, accuracy, and content) is unclear. Additionally, Deane (2014) reported that revision behaviors are not significantly related to text quality.

Concerning fluency, Spelman Miller et al. (2008) reported that a strong association between fluency and writing scores was detected. Similarly, according to Alves and Limpo (2015), shorter pauses were related to higher fluency. Zhang, Hao, Li, and Deane (2016) also found that when pauses were evenly distributed, writers showed better fluency and gained higher essay scores. In conclusion, previous studies mainly discussed how writing behaviors are associated with total scores rather than subcomponents of scores. To fill this

research gap, this study delves into the following research questions:

1) To what extent does task complexity influence EFL learners' pausing and revision behaviors?
2) To what extent does task complexity influence EFL learners' writing performance (i.e., task completion, coherence and cohesion, language use, and expression and tone)?
3) Are there links between EFL learners' pausing and revision behaviors and writing performance (i.e., task completion, coherence and cohesion, language use, and expression and tone)?

3. METHODOLOGY

3.1. Participants

Participants of this study were 30 advanced level Korean EFL undergraduates (13 male and 16 female) with the mean age of 19.63 years (*SD* = .81). They were randomly assigned to two groups (simple group: *n* = 13, complex group: *n* = 17), and the compatibility of the groups was determined by a reading comprehension test ($t(28) = .30$, $p = .798$). The students began learning English at the mean age of 6.07 years (*SD* = 1.87), which ranged from 2 to 10 years old. 10 of them had lived in English-speaking countries (*M* = 3.75, *SD* = 3.92), which ranged from 1 to 14 years of residence. 5 of them had taken TOEIC (*M* = 943.00, *SD* = 47.38) and 13 of them had taken iBT TOEFL (*M* = 102.24, *SD* = 11.40). Their majors included social sciences (26.67%), administration (20%), liberal arts (16.67%), engineering (13.33%), etc. They had employed various academic resources to learn English: formal instruction (28%), private instruction (25%), communication with native speakers (20%), media (16%), online class (11%), and private tutoring (8%).

3.2. Design and Materials

At the outset of the study, participants filled out a background information questionnaire and completed a reading comprehension test adapted from iBT TOEFL. Then they were divided into two groups differing in task complexity controlled by ±few elements. The task was to write a letter to a friend recommending accommodation for their trip. To complete the writing task, they had to choose one accommodation out of five alternatives considering given criteria (see Appendix A). In the simple task group, writers had to consider two criteria, whereas in the complex task group, six criteria had to be considered.

Their writing behaviors were recorded using a keystroke logging software, Inputlog 8.0. After the composition, all students answered a perception questionnaire, which was designed in 7 Likert-scale to measure their perception of task complexity. Subsequently, 9 randomly selected students (simple group: n = 5, complex group: n = 4) participated in a stimulated recall session, which was conducted individually for approximately 60 minutes. During the interview, each participant watched the keystroke logging video and answered reasons for pausing and revision in specific locations. The entire stimulated recall session was video-recorded and transcribed to capture the students' verbal comments.

3.3. Analysis and Coding

The data were analyzed and coded as follows. First, writing behaviors were analyzed in terms of pausing and revision keystrokes. Pause threshold was set as 200 milliseconds following the convention in writing research (e.g., Michel et al., 2020). Revision behaviors detected from keystroke log files were manually coded according to location. Second, writing scores were used as a performance measure. The essays were rated by two raters, and the inter-rater reliability was high (Cohen's Kappa = .94). For rating, an analytic scoring rubric of four categories (i.e., task completion, language use, coherence and cohesion, expression and tone) was employed. Third, stimulated recall comments were transcribed and classified into four subcategories (i.e., planning, formulation, monitoring, and unspecified) based on Kellogg et al. (2013).

3.4. Statistical Analysis

To explore the effects of task complexity on writing behaviors, independent-samples t tests were computed. For pausing behaviors, the dependent variables were set as the number of pauses, pause length, the number of pauses/100 words, and the number of p-bursts. The rationale for analyzing both the number of pauses and pauses/100 words was to investigate pause patterns in detail. Though the number of words in students' essays were similar (200 to 250 words), statistical analysis captured minute differences between the number of pauses and pauses/100 words. All the dependent variables except p-burst were subdivided into four locations: total, within word, between words, and between sentences. For revision behaviors, the dependent variables were set as the number of revisions, revision length, the number of revisions/100 words, and the number of r-bursts (bursts which were terminated by revisions). The number of revision behaviors were divided into five locations: below the word level, at the word level, below the clause level, at the clause level and above, and the sentence level and above. The revision length was divided into normal production, deletion, and insertion. To explore the effects of task complexity on

writing performance, essay scores were set as the dependent variable and it was subdivided into task completion, coherence and cohesion, language use, and expression and tone. Pearson's correlation coefficient was performed to explore the relation between writing behaviors and performance. Using independent-samples *t* tests, the compatibility of the groups was determined, and the perception questionnaire responses were analyzed.

4. RESULTS

4.1. Task Complexity and Writing Behaviors

Table 1 provides the descriptive and inferential statistics for pausing behaviors of the two groups. The simple task group paused significantly more often in total ($t(28) = 2.11$, $p = .044$), between words ($t(28) = 2.22$, $p = .035$), and between words/100 words ($t(28) = 2.47$, $p = .020$). The complex group showed significantly longer pause length in total ($t(24) = 2.52$, $p = .019$). Provided in Table 2 is descriptive and inferential statistics for revision behaviors of the two groups. The simple task group made more revisions in total ($t(27) = 2.12$, $p = .043$), and revised more frequently at clause and above ($t(15) = 3.00$, $p = .009$).

4.2. Task Complexity and Writing Performance

According to the independent-samples *t* tests, there was no significant difference between the two groups in terms of total scores ($t(28) = 0.28$, $p = .785$). They also showed no differences in terms of task completion ($t(28) = 0.22$, $p = .829$), coherence and cohesion ($t(28) = 0.91$, $p = .369$), language use ($t(28) = 0.93$, $p = .362$), and expression and tone ($t(28) = 0.90$, $p = .378$). Thus, task complexity did not affect the writing performance of the participants significantly.

4.3. Writing Behaviors and Writing Performance

Table 3 shows the relationship between essay scores and pausing behaviors of the simple task group. Higher scores in task completion were related to shorter pauses between words ($r = -.81$, $p = .001$), and more pauses between sentences meant improved coherence and cohesion ($r = .58$, $p = .038$). Those who paused more often within words ($r = .61$, $p = .028$) and within words/100 words ($r = .63$, $p = .022$) showed better performance in language use. Writers who paused less within words ($r = -.65$, $p = .016$) and within words/100 words ($r = -.67$, $p = .012$) gained better scores in expression and tone.

TABLE 1

Descriptive and Inferential Statistics for Pausing Behaviors

Measures	Simple (n = 13)		Complex (n = 17)		Comparison (t test)			
	M	*SD*	*M*	*SD*	*t*	*df*	*p*	*d*
Number of pauses								
Total	1083.46	336.17	772.71	440.13	2.11	28.00	.044*	.79
Within word	306.69	111.11	214.94	149.19	1.86	28.00	.074	.70
Between words	298.85	93.56	209.53	119.49	2.22	28.00	.035*	.83
Between sentences	20.23	4.83	18.35	12.67	.56	21.63	.581	.20
Pause length								
Total	1.12	.24	1.47	.49	2.52	24.07	.019*	.91
Within word	.57	.09	.55	.12	.42	28.00	.675	.19
Between words	1.58	.71	1.76	.79	.63	28.00	.530	.24
Between sentences	2.76	1.59	2.62	2.26	.19	27.00	.852	.07
Pauses/100 words								
Total	507.32	196.77	348.38	225.43	2.02	28.00	.053	.75
Within word	143.76	59.09	98.01	77.80	1.76	28.00	.089	.66
Between words	139.35	54.47	90.81	52.47	2.47	28.00	.020*	.91
Between sentences	9.30	1.89	8.05	5.74	.84	20.32	.413	.29
Number of p-bursts	100.54	27.08	91.59	50.15	.63	25.57	.537	.22

*p < .05

TABLE 2

Descriptive and Inferential Statistics for Revision Behaviors

Measures	Simple (n = 13)		Complex (n = 17)		Comparison (t test)			
	M	*SD*	*M*	*SD*	*t*	*df*	*p*	*d*
Number of revisions								
Total	269.17	173.72	158.47	107.31	2.12	27.00	.043*	.77
Below word	94.15	83.40	60.29	44.69	1.43	28.00	.163	.51
At word	65.85	45.28	41.29	30.36	1.78	28.00	.086	.64
Below clause	44.08	30.49	29.53	23.33	1.48	28.00	.149	.54
At clause	40.54	25.77	17.65	10.97	3.00	15.33	.009**	1.16
Sentence & above	12.31	6.14	9.71	7.26	1.04	28.00	.308	.39
Revision length								
Normal production	638.85	505.59	902.29	685.87	1.16	28.00	.255	.44
Deletion	1190.62	1436.18	421.47	440.77	2.09	28.00	.046	.72
Insertion	1565.54	1732.44	501.00	133.34	2.14	13.86	.051	.87
Revisions/100 words								
Total	122.34	100.33	70.98	53.00	1.81	28.00	.081	.64
Below word	45.97	49.60	27.11	21.35	1.41	28.00	.169	.49
At word	31.23	25.67	18.76	15.78	1.64	28.00	.112	.59
Below clause	20.94	17.24	13.29	11.34	.49	28.00	.154	.52
At clause	18.72	12.43	7.63	4.75	3.05	14.69	.008	1.18
Sentence & above	5.49	2.32	4.20	3.40	1.17	28.00	.251	.44
Number of r-bursts	53.92	69.34	46.59	39.14	.37	28.00	.716	.13

*p < .05, **p < .01

TABLE 3

Correlations Between Essay Scores and Pausing Behaviors of Simple Task Group

Measures (n = 13)	Total		Task completion		Coherence and cohesion		Language use		Expression and tone	
	r	*p*	*r*	*p*	*r*	*p*	*r*	*p*	*r*	*p*
Number of pauses										
Total	.12	.697	-.09	.768	.20	.504	.31	.305	-.21	.490
Within word	-.43	.140	-.06	.838	-.17	.581	-.17	.578	-.65	.016*
Between words	.43	.145	-.33	.264	.36	.225	.61	.028*	.19	.537
Between sentences	.50	.080	.14	.651	.58	.038*	.08	.794	.49	.092
Pause length										
Total	.01	.963	.94	.025	-.04	.909	-.17	.572	.26	.397
Within word	.05	.884	.02	.961	-.12	.709	.11	.727	.08	.788
Between words	-.44	.133	-.81	.001**	-.05	.872	-.26	.387	-.20	.515
Between sentences	-.25	.403	-.14	.641	.12	.708	-.36	.226	-.21	.489
Pauses/100 words										
Total	.11	.727	.05	.862	.08	.801	.37	.209	-.29	.341
Within word	-.37	.209	.06	.839	-.22	.474	-.06	.850	-.67	.012*
Between words	.36	.226	-.12	.703	.19	.525	.63	.022*	.03	.934
Between sentences	.46	.113	.34	.259	.45	.120	.14	.642	.31	.308
Number of p-bursts	.03	.932	-.13	.679	.09	.777	.03	.936	.03	.916

*p < .05, **p < .01

TABLE 4

Correlations Between Essay Scores and Pausing Behaviors of Complex Task Group

Measures (n = 17)	Total		Task completion		Coherence and cohesion		Language use		Expression and tone	
	r	*p*	*r*	*p*	*r*	*p*	*r*	*p*	*r*	*p*
Number of pauses										
Total	-.46	.061	-.45	.071	-.25	.339	-.20	.453	-.56	.019*
Within word	-.62	.008**	-.39	.123	-.41	.107	-.42	.097	-.61	.009**
Between words	-.02	.939	-.23	.377	.11	.671	.04	.876	-.13	.612
Between sentences	-.04	.778	-.15	.569	.16	.538	-.11	.665	-.23	.365
Pause length										
Total	-.05	.85	.21	.424	-.10	.713	.00	.997	.16	.533
Within word	-.29	.267	.08	.765	-.09	.732	-.18	.486	-.45	.068
Between words	-.33	.197	.08	.753	-.17	.504	-.38	.135	-.36	.158
Between sentences	-.55	.026*	-.05	.853	-.52	.041*	-.36	.177	-.50	.049*
Pauses/100 words										
Total	-.62	.008*	-.33	.192	-.38	.134	-.39	.126	-.69	.002**
Within word	-.73	.001**	-.03	.236	-.50	.040*	-.54	.026*	-.71	.002**
Between words	-.27	.302	-.19	.467	-.07	.779	-.19	.474	-.37	.141
Between sentences	-.31	.226	-.14	.604	-.03	.919	-.30	.243	-.47	.057
Number of p-bursts	-.51	.038*	-.27	.305	-.33	.199	-.35	.169	-.52	.032*

*p < .05, **p < .01

TABLE 5

Correlations Between Essay Scores and Revision Behaviors of Simple Task Group

Measures (n = 13)	Total		Task completion		Coherence and cohesion		Language use		Expression and tone	
	r	*p*	*r*	*p*	*r*	*p*	*r*	*p*	*r*	*p*
Number of revisions										
Total	.28	.383	-.13	.682	.34	.254	.43	.144	.08	.788
Below word	.29	.345	.03	.924	.28	.358	.38	.195	-.04	.889
At word	.30	.314	-.09	.777	.32	.285	.32	.292	.11	.714
Below clause	.39	.185	-.16	.609	.35	.237	.50	.084	.13	.672
At clause	.28	.350	-.58	.039*	.30	.319	.45	.126	.24	.434
Sentence and above	.38	.197	-.08	.788	.44	.137	.10	.748	.45	.122
Revision length										
Normal production	.25	.406	.34	.257	.31	.301	.25	.413	-.19	.532
Deletion	.36	.229	.05	.867	.30	.325	.03	.932	.53	.065
Insertion	.18	.566	-.13	.663	.04	.898	.03	.914	.44	.137
Revisions/100 words										
Total	.32	.293	-.03	.932	.24	.431	.47	.106	.00	.995
Below word	.27	.370	.07	.818	.20	.512	.43	.144	-.08	.791
At word	.29	.335	-.00	.992	.23	.451	.39	.183	.03	.934
Below clause	.35	.235	-.05	.867	.25	.402	.52	.070	.04	.895
At clause	.31	.309	-.41	.159	.23	.448	.53	.060	.16	.608
Sentence and above	.36	.231	-.01	.970	.41	.164	.07	.818	.40	.174
Number of r-bursts	.25	.410	.20	.524	.33	.265	.26	.388	-.14	.643

*p < .05

TABLE 6

Correlations Between Essay Scores and Revision Behaviors of Complex Task Group

Measures (n = 17)	Total		Task completion		Coherence and cohesion		Language use		Expression and tone	
	r	*p*	*r*	*p*	*r*	*p*	*r*	*p*	*r*	*p*
Number of revisions										
Total	-.27	.294	.30	.249	.03	.914	-.16	.542	-.49	.046*
Below word	-.29	.252	-.29	.261	.04	.874	-.21	.417	-.53	.030*
At word	-.25	.327	-.18	.490	.02	.933	-.20	.437	-.45	.072
Below clause	-.20	.450	-.32	.216	.09	.745	-.04	.890	-.46	.060
At clause	-.18	.502	-.30	.238	-.09	.736	-.05	.844	-.18	.482
Sentence and above	-.24	.364	-.37	.141	-.07	.788	-.01	.962	-.37	.143
Revision length										
Normal production	-.08	.760	-.23	.371	-.03	.898	-.08	.749	.00	.999
Deletion	-.44	.074	-.15	.565	-.17	.506	-.34	.178	-.60	.011*
Insertion	.12	.656	.12	.657	.21	.430	.16	.552	-.13	.613
Revisions/100 words										
Total	-.42	.090	-.23	.382	-.11	.680	-.32	.214	-.62	.008**
Below word	-.44	.081	-.24	.357	-.08	.754	-.35	.168	-.64	.005**
At word	-.37	.141	-.11	.665	-.09	.724	-.33	.193	-.53	.028*
Below clause	-.33	.201	-.23	.381	-.03	.918	-.19	.467	-.57	.017*
At clause	-.43	.084	-.28	.269	-.29	.267	-.28	.284	-.43	.087
Sentence and above	-.46	.064	-.35	.163	-.25	.338	-.19	.462	-.59	.012*
Number of r-bursts	-.35	.166	-.46	.063	-.09	.734	-.24	.357	-.42	.091

*p < .05, **p < .01

In Table 4, the complex task group participants who gained higher total scores paused less within word (r = -.62, p = .008), in total/100 words (r = -.62, p = .008), and within words/100 words (r = -.73, p = .001). Those with better total scores also paused shorter between sentences (r = -.55, p = .026) and showed less p-bursts (r = -.51, p = .038). In terms of the subcategories of writing scores, participants who paused shorter between sentences (r = -.52, p = .041) or less frequently within words/100 words (r = -.50, p = .040) showed better coherence and cohesion. Also, those who paused less within words/100 words scored better in language use (r = -.54, p = .026), and those who paused shorter between sentences scored higher in expression and tone (r = -.50, p = .049). In addition, writers who paused less frequently in total (r = -.56, p = .019), within words (r = -.61, p = .009), in total/100 words (r = -.69, p = .002), within words/100 words (r = -.71, p = .002) gained better scores in expression and tone. Less p-burst was related to better expression and tone (r = -.52, p = .032).

In Table 5, the simple task group's higher scores in task completion were associated with less revision in clauses (r = -.58, p = .039). In Table 6, the complex task group's better scores in expression and tone were associated with less revision in total (r = -.49, p = .046), less revision below word (r = -.53, p = .030), and shorter revision for deletion (r = -.60, p = .011). For revisions/100 words, higher expression and tone scores were related to less revision in total (r = -.62, p = .008), below word (r = -.64, p = .005), at word (r = -.53, p = .028), below clause (r = -.57, p = .017), and sentence and above (r = -.59, p = .012).

4.4. Task Complexity and Writers' Underlying Cognitive Processes

Perception questionnaire responses were analyzed to determine whether the task complexity manipulation of the present study is valid. It was confirmed that the participants perceived the task complexity in line with the deliberated task complexity manipulation. That is, they rated the complex task as significantly more difficult (t(28) = 0.21, p = .044) and requiring more mental effort (t(28) = 2.76, p = .010).

Table 7 summarizes the number of stimulated recall comments for pausing. Under the simple task condition, the largest percentage of comments referred to formulation (42.21%), followed by planning (31.17%) and revision (11.04%). In the complex task condition, the largest percentage of comments referred to formulation (36.44%), followed by planning (32.20%) and monitoring (12.71%).

TABLE 7

Reasons for Pausing (number of comments) from Stimulated Recalls

	Simple (*n* = 5)	Complex (*n* = 4)	Total (*n* = 9)
Planning			
Content	34 (22.08%)	27 (22.88%)	61 (22.43%)
Organization	14 (9.09%)	11 (9.32%)	25 (9.19%)
Total	48 (31.17%)	38 (32.20%)	76 (27.94%)
Formulation			
Lexical retrieval	24 (15.58%)	21 (17.80%)	45 (16.54%)
Syntactic encoding	39 (25.32%)	19 (16.10%)	58 (21.32%)
Coherence & cohesion	2 (1.30%)	3 (2.54%)	5 (1.84%)
Total	65 (42.21%)	43 (36.44%)	108 (39.71%)
Revision	17 (11.04%)	5 (4.24%)	22 (8.09%)
Monitoring	10 (6.49%)	15 (12.71%)	25 (9.19%)
Resource	10 (6.49%)	13 (11.02%)	23 (8.46%)
Unspecified	4 (2.60%)	4 (3.39%)	8 (2.94%)
Total	154 (100.00%)	118 (100.00%)	272 (100.00%)

Concerning syntactic encoding, the simple task writers focused on accuracy while their counterparts focused on naturalness. That is, the simple task writers' comments included 82.5% regarding accuracy and only 2.56% regarding naturalness. Moreover, they frequently used metalinguistic terms such as a relative pronoun, subject-verb agreement, and tense. To be specific, comments of the simple task group included '*I was checking the agreement between subject and verb.' and 'Which tense is appropriate for this sentence?'* On the other hand, 42.11% of the complex group writers' comments were related to naturalness, whereas only 26.32% were associated with accuracy. For instance, one of the complex task group writers responded that '*I didn't think this sentence reads well.*'

5. DISCUSSIONS

5.1. Task Complexity and L2 Writers' Behaviors

The first research question of this study concerned the extent to which task complexity influences EFL learners' writing behaviors. Task complexity was found to have a significant impact on pause length, pause frequency, and revision frequency. First, participants showed longer pauses in total under high complexity conditions. Based on Kellogg et al. (2013), it can be interpreted in terms of cognitive overload on writers' working memory. Due to the cognitive burden caused by a greater number of elements, the complex group writers might have needed more time to process the information, and it might have been revealed as longer physical inactivity during writing. Previous studies also support that longer pauses are associated with greater cognitive overload (e.g., Alves,

Castro, de Sousa, & Stromqvist, 2007). This result is also in line with Matsuhashi (1981), who found that the mean pause length of the complex task group was longer, and that during longer pauses, they engaged in revisions related to conceptual planning. The current study's stimulated recall comments support Matsuhashi (1981) in that complex group writers paused longer due to conceptual planning concerns (e.g., *I paused here to plan a way to persuade the recipient of the letter to agree with my decision*.).

Second, the simple task was found to result in more pauses in total, between words, and between words per 100 words. According to the information processing model, this might be because the simple task group had more spare working memory to assign to factors other than task completion. Considering the location of their pauses (between words), they might have used their remaining working memory to deal with language such as retrieving words (Medimorec & Risko, 2017). The stimulated recall comments of the simple task group support this (e.g., *I could not think of an English word for pricey*.). However, these findings contradict Révész et al. (2017), who reported that participants paused less frequently under the simple task condition. A possible explanation for the inconsistency might be that discrete operation was implemented to manipulate task complexity: the availability of content support in Révész et al. (2017) and ±few elements in the current study. Employing consistent manipulation might reveal congruous results.

Third, the simple task group demonstrated more revisions in total and at the clause level and above, which corresponds to Révész et al. (2017). The stimulated recall comments were supportive of the quantitative results. Specifically, writers showed more revisions in the simple task group (11.04%) than in the complex task group (4.24%). The results indicate that the smaller number of elements decreased cognitive burden on writers' working memory and allowed them to focus more on redundant processes such as revision.

5.2. Task Complexity and L2 Writers' Performance

The second research question of the present study asked the extent to which task complexity affects EFL learners' writing performance. Based on Kellogg et al. (2013), it was initially assumed that increased task complexity will have a significant impact on writing performance. That is, since complex tasks put more cognitive demands on writers' working memory, writers were expected to be less successful at engaging in a writing task, which will be reflected in text quality. Another rationale of the assumption derived from Robinson (2001b). According to the hypothesis, increasing task complexity along resource-directing factors was expected to cause more accurate and complex performance.

Contrary to expectations, it was revealed that task complexity does not have a significant impact on neither total scores nor scores of subcategories. Such findings are in line with Ong and Zhang (2010), who reported that task complexity has no significant influence on

lexical complexity and fluency. A possible explanation might be that the effects exerted by the manipulation of different task factors might have led to varied outcomes. Moreover, the effects of task manipulation on learners' attentional resources and how they allocated their working memory capacity to what perspectives of task completion might have differed.

Despite its statistical insignificance, one thing to note is that no one in the complex task group attained more than 75% of the total score in the language use category. Along with the complex task group's less frequent pausing and revision behaviors, it could be explained based on the information processing model. Since the complex task group writers had more cognitive demands imposed on their working memory, they might not have been able to allocate their working memory to linguistic encoding processes, and it might have been reflected in their scores. This supports Ishikawa (2008), who reported that the manipulation of resource-directing dimensions could affect how cognitive resources are allocated to language form.

5.3. L2 Writers' Behaviors and Performance

The last research question of this study is to examine the relationship between EFL learners' writing behaviors and performance. First, regardless of task complexity, higher total essay scores were related to less frequent pausing (within words, within words/100 words) and shorter pausing (between words, between sentences). The result corresponds to Alves and Limpo (2015) and Choi and Deane (2021), who argued that shorter pauses are related to improved text quality. On the other hand, regardless of task complexity, revision behaviors were not significantly related to total essay scores. It indicates that more writing behaviors do not directly lead to better writing performance.

Second, in the simple task group, writing behaviors and total writing scores were not significantly related. This result confirms Spelman Miller et al. (2008) and Deane (2014), who concluded that only a minor association was detected between pausing behavior and writing scores. Therefore, writing patterns may not be a useful indicator of total writing score in the simple task group. Nevertheless, significant relations were found between writing behaviors and subcategories of writing scores. To be specific, writers who paused more frequently between sentences gained higher scores in coherence and cohesion, which supports Révész et al.'s (2017) finding that pauses between sentences are associated with higher order writing processes such as coherence. Also, writers who paused more frequently between words in total and per 100 words gained higher scores in the language use category. Considering the stimulated recall comments of the simple task writers, this might be due to the fact that simple task writers had more spare working memory to deal with language, which led to improved performance in the language use category. These findings are also in line with Révész et al.'s (2017) finding that pauses between words are

related to lower order writing processes such as lexical retrieval. Furthermore, writers who revised more often at clause gained higher scores in coherence and cohesion. The result confirmed Révész et al. (2019) in that higher-level revisions were related to higher-order writing processes such as planning organization.

In the complex task group, consistent negative relations were found between pausing behaviors and total scores. In other words, more frequent and longer pauses did not lead to higher writing scores. The result confirms the findings from Révész et al. (2019) where negative relation between complex task groups' performance and pausing behaviors was reported. However, the result contradicts Xu and Qi (2017), who argued that pausing patterns were positively related to text quality. The discrepancy between the two studies might be explained by the different English proficiency levels of the participants: intermediate to upper-intermediate level in Xu and Qi (2017) and advanced level in the current study. Thus, it can be concluded that the link between pausing behaviors and performance are inconsistent across the proficiency level. With regard to revision behaviors of the complex task group, more frequent revision and longer revision length were consistently correlated with lower scores in expression and tone. The result indicates that higher quantity in revision does not necessarily lead to better quality in expression and tone.

6. CONCLUSION

The present study investigated the effects of task complexity on EFL learners' writing behaviors and performance. The findings are as follows. First, the simple task group writers showed more frequent pauses and revisions, which may be due to less cognitive demands imposed on their working memory. The pause location, stimulated recall comments, and essay scores indicate that they allotted their remaining working memory to language. On the other hand, due to cognitive overload, writers of the complex task group showed longer pauses, which were associated with higher-order writing processes. Second, contrary to assumptions derived from Robinson (2001b) and Kellogg et al. (2013), task complexity was found to have no influence on writing performance. Third, regardless of task complexity, higher total essay scores were related to less frequent and shorter pauses, which corresponds to previous studies. However, revision behaviors were not significantly related to total essay scores, which shows that more revisions do not directly result in better text quality. In the simple task group, more pauses between sentences led to better scores in coherence and cohesion, and more pauses between words led to improved language use. The results support Révész et al. (2017) in that pauses between sentences were associated with higher order writing processes and pauses between words were related to lower order

writing processes. Also, more revisions at clause were associated with better scores in coherence and cohesion, which confirmed the results of Révész et al. (2019). In the complex task group, writers showed consistent negative relation between subcategories of writing scores and writing behaviors.

There are several implications for this study. In theoretical perspective, the results of this study contradict Robinson's (2001b) hypothesis that increasing task complexity along resource-directing factors would have a positive impact on accuracy and complexity. Possible explanation may be that the manipulation of task complexity might have caused different effects on individual learners' working memory allocation. However, this study supports Kellogg et al. (2013) in that increased task complexity imposed more cognitive burden on learners, which led to longer pauses in higher level discourse units.

Pedagogically, the findings of the current study can provide insights into cognitive processes of L2 writing, such as planning, execution, and revising. Understanding L2 learners' cognitive processes, especially pausing and revision behaviors, will benefit various stakeholders of L2 learning. Specifically, data of writers' behaviors can be effective means of predicting difficulties in L2 writing which cannot be detected in the final text (Révész et al., 2019). Therefore, instructors could rely on online writing data to provide learners with more precise and appropriate feedback on specific areas that require practice. Moreover, keystroke logging data and stimulated recall comments can provide language testers with the information regarding the links between writing tasks, intended construct, and their cognitive processes. It can assure language testers that the construct functions as intended in a specific writing task and that performance-based score inferences properly demonstrate learners' linguistics knowledge (Purpura, 2013). Furthermore, task complexity and its effects on writing behaviors and performance can be used by instructors and syllabus designers to evaluate different cognitive demands imposed by their selected task designs. The data can provide insights into curricular decisions such as task grading and sequencing (Robinson, 2001a). In classrooms, language instructors could control the complexity of writing tasks in order to regulate learners' writing behaviors, which could influence their performance. Also, various levels of task complexity would aid learners' adjustment to different writing processes. Thus, manipulating task complexity appropriately according to learners' needs and proficiency levels will facilitate L2 learning.

Despite its implications, this study bears a number of limitations to be acknowledged. One limitation relates to the relatively small number of participants. The present study analyzed 17 essays from the complex task group and 13 essays from the simple task group. Further research could analyze larger numbers of essays to enhance the reliability of the result. Also, the participants' English proficiency levels as well as nationality were congruous, which might affect the generalizability of other populations. That is, all the participants were advanced-level students, and their nationality was Korean. Later studies

could investigate whether results of the current study can be generalized to learners of different proficiency levels and nationality. Additional limitation pertains to the implementation of stimulated recall methodology. Inherent limitations of this method include the fact that it only provides data concerning writers' conscious behaviors, and that it is impossible for participants to report all conscious processes due to the loss of memory. Such problems might be alleviated by using other methods such as eye tracking in future research. Lastly, another methodological limitation lies in the fact that the relatively high pause threshold was utilized in this study (200 milliseconds). Although the adoption of the 200 milliseconds threshold enabled the comparability of the present study to previous studies, engaging a lower pause threshold may reveal more significant relations between task complexity and writing behaviors.

Applicable level: Tertiary

REFERENCES

Abdi Tabari, M. (2022). Investigating the interactions between L2 writing processes and products under different task planning time conditions. *Journal of Second Language Writing, 55*, 100871

Alves, R. A., Castro, S. L., de Sousa, L., & Stromqvist, S. (2007). Influence of keyboarding skill on pause-execution cycles in written composition. In M. Torrance, L. van Waes & D. Galbraith (Eds.), *Writing and cognition: Research and applications* (pp. 55-65). Amsterdam, The Netherlands: Elseveir.

Alves, R. A., & Limpo, T. (2015). Progress in written language bursts, pauses, transcription, and written composition across schooling. *Scientific Studies of Reading, 19*(5), 374-391.

Baddeley, A. (1986). *Working memory.* Oxford, England: Oxford University Press.

Barkaoui, K. (2019). What can L2 writers' pausing behavior tell us about their L2 writing processes? *Studies in Second Language Acquisition*, *41*(3), 529-554.

Choi, I., & Deane, P. (2021). Evaluating writing process features in an adult EFL writing assessment context: A keystroke logging study. *Language Assessment Quarterly, 18*(2), 107-132.

Conijn, R., Roeser, J., & van Zaanen, M. (2019). Understanding the keystroke log: The effect of writing task on keystroke features. *Reading and Writing, 32*(9), 2353-2374.

Deane, P. (2014). *Using writing process and product features to assess writing quality and explore how those features relate to other literacy tasks* (Research Report No. RR-14-03). Princeton, NJ: Educational Testing Service.

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

Kellogg, R., Whiteford, A., Turner, C., Cahill, M., & Merlens, A. (2013). Working memory in written composition: An evaluation of the 1996 model. *Journal of Writing Research, 5*(2), 159-190.

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

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

Matsuhashi, A. (1981). Pausing and planning: The tempo of written discourse production. *Research in the Teaching of English*, *15*(2), 113-134.

Medimorec, S., & Risko, E. F. (2017). Pauses in written composition: On the importance of where writers pause. *Reading and Writing, 30*, 1267–1285.

Michel, M., Révész, A., Lu, X., Kourtali, N. E., Lee, M., & Borges, L. (2020). Investigating L2 writing processes across independent and integrated tasks: A mixed-methods study. *Second Language Research*, *36*(3), 307-334.

Nunan, D. (2004). *Task-based language teaching*. Cambridge, England: Cambridge University Press.

Ong, J., & Zhang, L. J. (2010). Effects of task complexity on the fluency and lexical complexity in EFL students' argumentative writing. *Journal of Second Language Writing*, *19*, 218-233.

Purpura, J. E. (2013). Cognition and language assessment. In A. J. Kunnan (Ed.), *The companion to language assessment* (pp. 1452–1476). Oxford, England: John Wiley & Sons.

Rahimi, M. (2019). Effects of increasing the degree of reasoning and the number of elements on L2 argumentative writing. *Language Teaching Research, 23*(5), 633-654.

Rahimpour, M., & Hosseini, P. (2010). The impact of task complexity on L2 learners' written narratives. *English Language Teaching, 3*(3), 198-205.

Révész, A., Kourtali, N., & Mazgutova, D. (2017). Effects of task complexity on L2 writing behaviors and linguistic complexity. *Language Learning, 67*(1), 208-241.

Révész, A., Michel, M., & Lee, M. (2019). Exploring second language writers' pausing and revision behaviors: A mixed-methods study. *Studies in Second Language Acquisition, 41*, 605-631.

Robinson, P. (2001a). Task complexity, cognitive resources, and syllabus design: A triadic framework for investigating task influences on SLA. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 287–318). New York: Cambridge University Press.

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

Robinson, P. (2003). The Cognition Hypothesis, task design, and adult task-based language learning. *Second Language Studies, 21*(2), 45-105.

Sadeghi, K., & Mosalli, Z. (2012). The effect of task complexity on fluency and lexical complexity of EFL learners' argumentative writing. *International Journal of Applied Linguistics and English Literature*, *1*(4), 53-66.

Salimi, A., Dadaspour, S., & Asadollahfam, H. (2011). The effect of task complexity on EFL learners' written performance. *Social and Behavioral Sciences, 29*, 1390-1399.

Schilperoord, J. (1996). *It's about time: Temporal aspects of cognitive processes in text production*. Amsterdam, The Netherlands: Rodopi.

Spelman Miller, K. (2000). Academic writers on-line: Investigating pausing in the production of text. *Language Teaching Research*, *4*(2), 123-148.

Spelman Miller, K., Lindgren, E., & Sullivan, K. P. H. (2008). The psycholinguistic dimension in second language writing: Opportunities for research and pedagogy using computer keystroke logging. *TESOL Quarterly, 42*(3), 433-454.

Xu, C., & Ding, Y. (2014). An exploratory study of pauses in computer-assisted EFL writing. *Language Learning and Technology, 18*(3), 80–96.

Xu, C., & Qi, Y. (2017). Analyzing pauses in computer-assisted EFL writing: A computer-keystroke-log perspective. *Educational Technology and Society*, *20*(4)*,* 24-34.

Zhang, M., Hao, J., Li, C., & Deane, P. (2016). Classification of Writing Patterns Using Keystroke Logs. In L. A. van der Ark, D. M. Bolt, W. C. Wang, J. A. Douglas & M. Wiberg (Eds.), *Quantitative psychology research: The 80th annual meeting of the psychometric society, Beijing, 2015* (pp. 299-314). New York: Springer.

APPENDIX A

Writing task for simple condition (200-250 words, 30 minutes)

You are planning a vacation to Jeju Island with your friend. **You have to**: 1. choose one accommodation 2. write a letter to your friend explaining why you think the accommodation is the most suitable. You and your friend want the accommodation to satisfy **two criteria**: 1. located near airport 2. includes breakfast.

Lists of accommodations

1. Yoon's Guest house
Location*: Jeju airport - 30 min by bus. Downtown Jeju city - 5 min walk.*
Description*: Breakfast included in the price; omelets and pies. Bedroom in local style, large terrace view with garden. Pets are welcome.*
Facilities*: Free Wi-Fi access in rooms. Free coffee/tea in a shared area.*

2. Starry Hotel
Location*: Free pick-up from Jeju airport - 25 min by bus. At a considerable distance from the city center.*
Description*: American style buffet, between 8:30 AM and 11:00 AM ($20 per person). Pets are not allowed.*
Facilities*: Fitness-room. Air conditioning. Free parking. Pool access available from 10:00 AM to 11:00 PM. Bathtub. Daily housekeeping.*

3. Harubang Inn
Location*: Jeju airport - 5 min by car. 30 km from Downtown.*
Description*: No breakfasts served. Guests can prepare their own breakfast. Attractively priced.*
Facilities*: Wheelchair-accessible. Free parking. Elevator. Free Wi-Fi. Cable TV channels.*

4. Orange Resort
Location*: Situated on the coast. 25 min by bus from Jeju airport and city center, respectively. Free shuttle bus from the airport.*
Description*: Ideally located for those seeking to spend a quiet holiday on the beach, or to go hiking in the mountains. Breakfast service (special discounts for guests).*
Facilities*: Free off-street parking area. Swimming pool 1km away. Barbecue grills.*

5. Marine Guest house
Location*: 40 km from Jeju airport. Situated in the center of the city.*
Description*: Multilingual staff (Korean, English, Chinese, Japanese). The bathroom is shared between two bedrooms.*
Facilities*: Breakfast available. Wi-Fi in rooms and public areas. Espresso maker and iron on request.*