The effects of task complexity on FL learners' online writing behaviors and the correlated underlying cognitive processes

Dr. Elham Sweilam Ahmad Desouky
Lecturer of Curriculum & Instruction (TEFL),
Faculty of Education, October 6 University

Abstract

The main target of the current study was to discover whether task complexity has any effect on the online behaviors of FL writers and the correlated cognitive processes underlying them. The participants were chosen after finishing their simple or complex essay tasks; they reached 146 participants, and four of them were chosen to be involved in spur callback. The keystroke logging software was used to record the writing behaviors of the participants. The presence of the content support was compared to the absence of task complexity. The results revealed that the presence of content support increased linguistic complexity, but at the same time there was less pausing and more revision. The absence of content support led to more repetition in pauses and the language was less developed in revisions.
Key words: Foreign language learners; task complexity; writing

The importance of tasks in language came from its role of providing a base to merge learning and evaluation. The evaluation here is based on the form and the meaning determined by language, as well as the participation of the learners in communication tasks, which enjoys high credibility (Révész et al., 2017). Lately researchers began to direct their attention to investigating the effect of task complexity on writing skills, while their interest before was on the impact of task complexity on speaking skills only. The current study tried to explore this area to examine the effect of the underlying cognitive demands of tasks on foreign language performance and improvement. The purpose was to formulate criteria for classifying and succession of tasks to identify the elements that cause difficulty of assessment in a foreign language as suggested by Brown et al., (2002).

Byrnes (2014) claimed that most studies of writing skills were concentrated on the relation between task complexity and the quality of writing production. Until now very little experimental studies have studied the effect of cognitive complexity of tasks on participants' online behaviors in writing (i.e., pausing and revision) and the underlying cognitive process that appear during writing (i.e., planning) (Macaro, 2014). Révész (2014) affirmed that no study can investigate fully writing models in relation to task complexity in the absence of

studying the casual process that appears as a result of task manipulations. In addition, it is important for any study to obtain evidence concerning behaviors that are generated by tasks and the underlying cognitive processes behind those behaviors.

The current study seeks to explore if there was any efficacy of task complexity on the behaviors of foreign language writers, particularly in fluency, pausing, and revision, as well as in linguistic encoding and planning as associated cognitive processes. Additionally, this study tried to find out whether task complexity manages the relationship between writing quality and writers' behaviors concerning fluency, pausing, and revision. The methodological method used in this study was online keystroke capturing software to measure writing behaviors of the participants and information gained by spur callback. This study added another investigation about the impact of task complexity on the linguistic complexity of the foreign language writers' production.

Kormos (2011) expected that the content support will make it harder for the foreign language writers to keep away from complex morphosyntactic and lexical constructions, and this makes the conditions more appropriate to develop and assess linguistic production. Therefore, in this study the presence of the content support was compared to the absence of task complexity for two reasons: to explore the effects of content support on the performance of writing, and to test predictions of cognitive patterns of writing, which play an important role in planning.

REVIEW OF PREVIOUS RESEARCH

The researcher divided this part into two sections: section one discussed previous theoretical research, and section two discussed previous experimental research.

SECTION (A)

Previous theoretical research

The current study used Kellogg's (1996) cognitive pattern of writing as a theoretical foundation to explore this area. This cognitive pattern is used to investigate the first language writing, and this makes it beneficial to the foreign language writing process because it foretells the linguistic encoding processes in detail. This foretelling may encourage foreign language writers to create significant cognitive demands. In this cognitive pattern, writing was defined in general as a repeated and reactive process, but at the same time it can be classified into three parts: formulation,

execution, and monitoring as subsidiary processes. Kormos (2011) explained the three subsidiary processes as follows: firstly, the formulation stage encompasses planning the content of the writing text, then transferring this written text into linguistic shape.

The writers tried to form a coherent plan for the written content through combining the instructions for the task and their long-term memory ideas. The subprocesses of lexical retrieval, syntactic encoding, and expression of cohesion come as a result of transferring ideas into linguistic form through translation. Secondly, in the execution phase, both handwritten and typed pieces are produced through motor motions. The last stage was monitoring, which gives the writers a chance to compare what were written with what they intended to say; if they found mismatches, revision may be a good solution. Révész (2017) believed that the interaction between the three stages leads to the emergence of complex models of processes, which are influenced by some variables such as task complexity in the field of writing in a foreign language.

Skehan (2014) argued that Kellogg's pattern did not offer any direct predictions concerning the relationships between writing processes, production, and task manipulations. Both Skehan (1998) and Robinson (2001) have discerned two different patterns: the first one was about limited capacity, while the other was about cognitive hypothesis. These two patterns have been modified to design research on task complexity in writing; however, Manchón (2014) argued that it is difficult to apply these theoretical frameworks to writing because they were designed specifically for speaking. The nature of producing written pieces is different from the psycholinguistic processes engaged in generating speech. Manchón gave some evidence to support his opinion, such as the writers can expend more time to plan the content that showed their message and use some resources like the expression of cohesion, retrieval, and syntactic encoding to help in translation processes. This happens because writing is always less restricted by time than speaking, which means that writers can deal with the limitations of their working memory capacity by dividing resources among different phases of writing. Additionally, the writers have good opportunities to revise their written work, unlike speakers who can only reform their instantly previous speech.

Even so, and depending on Kellogg's pattern, Kormos (2011) believed that the working memory of the writers is limited, so if the task complexity increases, this means the writers will be less successful at dealing with the reinforced demands placed on writing subprocesses. Therefore, it is expected that the influences on the quality of the written

pieces and writing behaviors are due to the cognitive task demands and the limitations of the writer's working memory. According to Kellogg's (1996) pattern, the manipulation of the task complexity used in the current research as provision compared to the absence of content support may possibly affect linguistic complexity as well as writing behaviors. Based on the absence of content support, it is anticipated that the status will be difficult for the writers because they are asked to access more ideas from long-term memory, and then connect these ideas by themselves. Kormos (2011) considered the absence of content support might lead to pressure on planning processes, as well as enhancing the use of more grammatical and lexical structures, which causes more pressure on the translating processes. In contrast to this case, the situation in a simple condition is different because the instructions of the task might include some of the relevant grammar and vocabulary.

Révész (2017) believed that these processes are like a chain in that each element depends on the other; this means the task processing will be slower because of the increased effort used in planning and translating content. The slower way of dealing with a task may lead to a slower pace in writing, while the number of pauses and their lengths are increased. In addition, sentences and clauses are considered larger units, and their pauses are probably associated with planning process; consequently pausing might become more lengthy and frequent. Fewer awareness resources were left for observing language use, so the participants might make fewer language-related revisions; hence, the behaviors of revision and pausing might not appear as predictors of major linguistic complexity.

The participants might produce less linguistically complex written pieces in the absence of the content support, while they direct more attention to grammatical encoding and lexical retrieval as a translation processes because of the decreased pressure on planning operations in the presence of content support. It is anticipated that when the writers involved in larger discourse units with the opportunity of reducing effort are requested to generate content, their ability to write faster may increase and pauses become shorter or less frequent. The participants can concentrate on translating their ideas in effective ways when modifications engaging lower discourse units linked with syntactic encoding and lexical processes are predicted to be more frequent. Therefore, it is expected that writers who pause and revise more at the lower discourse units will perhaps create written pieces with higher lexical and syntactic complexity if the ideas are presented (Kormos, 2011). It is important to put these previous anticipations into

consideration in the light of present results related to the influences of task complexity on the writing behaviors and the production of the foreign language writers, despite the little direct evidence about the validity of these expectations in previous experimental studies.

SECTION B

Previous experimental research

The researcher divided this section into three parts: part A discussed the previous work on task complexity and foreign language writing behaviors, then part B showed the relationship between text quality and foreign language writing behaviors. Finally, part C addressed the relationship between task complexity and linguistic complexity.

Part A

Task complexity and FL writing behaviors

There are few studies that have tried to investigate the relation between task complexity and foreign language writing behaviors. A study by Spelman (2000) was focused on exploring the relation between cognitive task complexity and writing behaviors. The main target of his study was to examine whether writing pausing behavior and fluency differed according to the type of essay writing task, whether evaluative or descriptive. He presumed that evaluative tasks need critical synthesis and assessment from different points of view because it poses greater cognitive demands. Spelman recorded the participants' online keyboard activity, and the results of these records showed that the longest pauses happened between sentences, as well as the writers pausing longer at higher levels of text unit. Moreover, pauses in noun and verb phrases were the most frequent. Spelman found that the different tasks did not influence pausing or fluency in writing.

The studies that explored the effect of task complexity on FL revision behaviors were limited. One of them was Thorson's (2000) research, where he used keystroke-logging software to find how revision models of FL writers differ according to the type of task. The writers in his study used their first language (English) and their foreign language (German) to perform two tasks: the first one was a newspaper article, while the second was a letter to a pen pal. The results were contrary to his anticipations, as there were no task impacts on the type or the quantity of FL revision; however, the writers revised the article more than the letter in their first language.

Notwithstanding the fact that cognitive writing processes have obtained growing attention in foreign language writing study, the field is still destitute, as only a few studies explored the effects of the task. While no studies have explored in what ways task complexity may have effects on the cognitive processes that underlie the behaviors of pausing and revision (Schoonen et al., 2009; Révész, 2017). A study by Ong (2014) tried to clarify that the writers who received support in planning time, topic, macrostructure, availability of content and any other guidelines devoted fewer of their awareness resources to metacognitive processes more than those who had no support. This result was along the same lines as Kellogg's (1990) predictions.

Part B

The relation between text quality and FL writing behaviors

In the field of first language writing, many studies have explored the relationship among the behaviors of pausing, fluency, and revision, and the quality of the text. While only a few studies have investigated these associations in foreign language writing area, a study by Stevenson et al. (2006) looked into whether the type of revision behaviors can help to anticipate the quality of the text. In this study the writers were asked to write four essays, two of them in English as a foreign language and the other two essays in Dutch as their first language. The results of this study went against the researchers' forecast, as there was no relationship between the quality of the text and the revision type. The researchers anticipated finding a negative relationship between the conceptual quality of the text and the revisions in its lower level, which did not happen.

The area of searching on the effect of both pausing and fluency on the quality of the text has been interesting in the study conducted by Spelman et al. (2008), which looked into whether pausing and fluency can help to predict the quality of the text. The speed of writing was evaluated in terms of fluency twice, once when typed characters between pauses and/or revisions and once during writing time between pauses and/or revisions; the numbers of revisions in terms of insertions or cancellations were also counted. The results disclosed that neither pausing nor revision behaviors showed any significant differences between the text quality scores. From the previously mentioned studies, it can be concluded that fluency may have positive effect on the quality of the text, while both pausing and revision have no correlation with the quality of the text. Obviously, these trends may be need to be confirmed via more studies as well as looking into how to moderate the connection

between the quality of the text and writing behaviors through task complexity.

Part C

The relation between Task Complexity and Linguistic Complexity

Many studies looked into task complexity dimensions in relation to linguistic complexity, including different factors such as planning time (Ellis, 2004), storyline complexity (Tavakoli, 2014), reasoning demands (Kuiken, 2008), telling a story with the presence of content support versus the absence of content support (Kormos, 2011), number of task elements (Ruiz-Funes, 2014), and revising conditions and provision of writing support (Ong, 2010). The current study worked along the same lines with those studies of Ong (2010) and Kormos (2011) as their focal point to investigate the influence on the quality of the text after giving content support. The area that examines the impacts on the linguistic complexity of written production because of task complexity is a matter of recent research interests; hence, the present research tried to contribute to previous studies in this domain.

To investigate whether lexical complexity is influenced by writing support, a study by Ong (2010) used three conditions: topic, ideas, and macrostructure given; topic and ideas given; and topic given then asked of Chinese participants who studied English as a foreign language (EFL) to write a polemic essay under one of these conditions. The ratio of word types squared to the total number of words in the final text was used to evaluate lexical complexity. The results showed that when comparing the premier drafts generated by the participants, no differences were observed, while the matter was different when revising the original written pieces of the participants. The writers generated more lexically complex texts when ideas were presented alone or when macrostructure and ideas were given, compared to when they were given the essay topic only.

Another study by Kormos (2011) used a more extensive range of linguistic characteristics to explore the impacts of content support. The researcher gave the participants who were EFL learners' six ordered portraits and six unconnected portraits. At the first stage the participants used the six portraits to construct a comic strip, while in the second stage the participants used the other unconnected portraits to narrate a story. The participants in this study were forced to use their fiction to connect the portraits in order to create a story based on these portraits; this action was considered by Kormos to be a heavy cognitive burden on the

participants because there was no plot. The results of two other studies by Heatley (2002) and McNamara (2005) revealed that when the content was predetermined, the participants would use a wider abstract vocabulary. This means that if the content is available, certain aspects of lexical complexity can be enhanced, while syntactic complexity is not influenced by using the content. However, these results need to be affirmed through more future investigations.

Research Questions

To achieve the purpose of the study, the researcher tried to pose some research questions that may help this study to be more accurate, taking into consideration former studies on tasks, linguistic complexity, FL writing behaviors, and finally the quality of the text.

These questions were formulated as follows:

1-If there are any effects of task complexity on FL writing behaviors and the correlated cognitive processes underlying them, what are these effects?

2- If there are any effects of task complexity on linguistic complexity of texts in FL, what are these effects?

3- If there are any connections between linguistic complexity of texts in FL and writing behaviors of FL, what role does task complexity play in these connections?

To calculate fluency speed, pausing, and revision, the FL writing behaviors were used as online measures. Task complexity was formulated in terms of ideas to be included in the controversial article as the presence or absence of content support. Syntactic complexity and lexical indicators were used to define linguistic complexity. The participants' stimulated recall comments on their inner installing were used to explore their cognitive writing processes.

Method

Participants

The learners who participated in this study were 146 males and females. Their age ranged between 20 and 21 years old, all of them in their fourth year in the Faculty of Education. All of the learners were Arabic speakers who had studied formal English for at least 9 years. The

researcher chose the participants randomly without any plan or bias and divided them into two groups as follows:

Group A: simple, the learners were 70, 60 female, 10 male

Group B: complex, the learners were 76, 58 female, 18 male

Both simple and complex groups had similar writing proficiency levels, as well as similar demographic characteristics. The simple and complex groups achieved comparable scores on a version of the Integrated Skills in English (ISE) III Correspondence Task. The rating criteria utilized by the faculty examiners to evaluate the ISE III controlled written examination and the findings of the independent sample t-tests showed no significant differences between the simple and complex groups (see Appendix A). In addition, the texts generated by both simple and complex groups on the ISE III Correspondence Task showed no differences from the linguistic complexity measures that the current study presented.

Study Design

The current study randomly divided the participants into two groups: simple and complex with different missions. The simple group was asked to write a less complex version of a controversial essay, while the complex group carried out the same version of the task but in a more complex way. The keystroke-logging software Inputlog 5.2 and screen-capturing technology were used to record the online writing processes used by the participants as was suggested by Leijten, (2013). The researcher randomly assigned 16 participants from both groups to accomplish an additional task, in which they used stimulated recall motivated by the playback of the recordings of their keystrokes and mouse clicks to characterize their thought processes over the performance of the task. After the writing task, the researcher asked the non-stimulated recall learners to finish a short perception questionnaire.

Tools

The tools used in this study included: writing tasks, perception questionnaires, and stimulated recall. These parts will be clarified as follows:

a- Writing Tasks

The ISE III Controversial Writing Task was operationalized as the complex controversial writing task, as well as the task prepared for The Common European Framework of Reference for Languages (CEFR) at level C1. This level means an ability to communicate with emphasis on how well it is done in terms of appropriateness, sensitivity, and the

capacity to deal with unfamiliar topics. The text used was: "Scientists and technology experts seem to be more valuable to modern society than musicians and artists. Do you agree?" The learners who were engaged in the simple task were supplied with ideas to be contained in the essay, as they should keep in mind two questions: (a) what can we learn about the importance of scientists and technology experts comparing to musicians and artists in 20th century? (b) What are the advantages of learning about scientists, technology experts, musicians, and artists? The learners were offered subtopics for both questions, from which they were encouraged to select and develop those chosen (see Appendix B). Along with Révész (2017), the current study utilized 200-250 words as a word restriction for both essay models, and the learners were given 45 minutes to finish the missions. This study used Shintani's (2013) study to satisfy the standards for task definition. It was expected that the task would produce essential concentration on meaning, as the learners were forced to produce certain content on their own because they were only given broad ideas in both cases.

b- **Perception Questionnaire**

The main purpose of this questionnaire was to affirm that the planned task version was likely to be more cognitively challenging that asked for more intellectual effort through examining the validity of the manipulation of task complexity (Révész, 2014). This questionnaire consisted of five questions, three of them related to the present research, and the learners were evaluated on a 9-point scale. The aim of the questions was to judge the perception of three points: firstly, the applied total intellectual effort; secondly, the difficulty of the task in general; and finally, the onerousness of designing the content of the essay. Great effort and awkwardness appeared in the higher values of the scale.

c- Stimulated Recall

To identify whether the task version was planned to be characteristically complex, produced qualitatively and/or quantitatively, particularly cognitive processes therefore, the stimulated recall was used to extract the ideas of the learners whilst finishing the writing task. While the participants were performing the task, the researcher asked them to stop the recording of their own writing any time they wanted to exchange ideas with others. At the same time, the participants had a chance to watch their own writing performance through the screen recording. When the participants stopped or checked out their own performance, the researcher paused the screen recording as well to gather ideas from the participants. The researcher conducted stimulated recall sessions in

English; thus, the participants did not suffer from difficulty in expressing their ideas.

Collection of the data

As previously mentioned, the researcher contracted simulated recall sessions for both a stimulated recall group and non-stimulated recall group. Although all the learners attended the sessions, the time that the researcher spent with each group was different, as it took one hour for the non-stimulated recall group and two hours for the stimulated recall group, as well as meeting them separately. Before the stimulated recall, both groups followed exactly the same steps. And all of them carried out the writing task in the faculty computer lab. After the participants had completed the writing task, their online writing behaviors were recorded by the keystroke-logging software Inputlog 5.2 and the screen-capture technology of the (Leijten, 2013). The last step was the perception questionnaire, which was given at once after the learners finished their writing task.

ANALYSIS OF THE DATA

The current research included three factors that need to be analyzed: firstly, behaviors of online writing, then stimulated recall comments, and finally the written texts produced by the learners. These elements will be clarified as follows:

Behaviors of Online Writing

The participants' output of premier drafts were the center of interest in behaviors of online writing analyses, as well as targeting both linear events that fashion forward progression and nonlinear events (i.e., revision). This study benefited from Baaijen et al.'s (2012) research in removing some texts that was generated as a part of titles, as well as some phases like plan on the screen, revision drafts, and final revisions because these phases engage processes unlike those that presuppose the production of premier drafts. In the next step after separating the output of written pieces as a portion of the first outlines, the files of keystroke log were analyzed in relation to pausing, revision, and the speed of fluency. To capture the speed of fluency, four measures were used (number of words/characters occurring between pauses and total writing time divided by total number of words/characters, excluding pauses) to be in line with the conclusions of Abdel-Latif (2013) who considered a valid measurement of the fluency speed involved calculating indicators of the length of the participant's output units. Hence, bursts took place during pauses; therefore, the current study used speed fluency narrowly and fostered a process-oriented perspective (Van Waes, 2015). Spelman et al.

(2008) considered that the operationalization of Van Waes's multidimensional pattern of fluency falls into the considerable class of output fluency; thus, the pausing behaviors in this study were located at 2 seconds.

The current study used the same considerable limitations in writing studies like Wengelin's (2006), where pause length and pause frequency are used to express pausing behavior. Additionally, the mean length of pauses and the number of pauses were counted per 100 words. The pauses were organized according to the place of appearance, whether pauses appeared between sentences or clauses, between words or within words. Predominantly between-word pauses contain just one pause prior to pressing the spacebar and before the beginning of the following word; thus, the present research dealt with between-word pauses as one pause. Cancellations, substitutions, and all the revision behaviors were calculated according to type and quantity. To evaluate the quantity of revision, there was a comparison between the number of letters/words generated over the whole writing process and the number of letters/words in the final version of the written piece. Along the same lines of Stevenson et al. (2006), a site was used to code revisions whether they took pace at the same standard of clause or higher, or beneath the standard of the word or below the clause level. The intercoder convention reached 94% after analyzing ten percent of the data, randomly selected.

Stimulated Recall Comments

In this stage, the researcher first copied the stimulated recall comments, and then reviewed them. Like Kellogg's pattern (1996) the researcher classified these comments into groups of planning, monitoring, and translation (when possible). In the planning phase, comments were further sub-classified into content-organization linked to comments, while comments in the translation stage were furthermore categorized according to whether they utilize cohesive devices, syntactic encoding, or include reference to lexical retrieval. Learners were not able to remember the reason behind pausing during writing 17 percent of the time; after that the researcher reexamined all of the notes. The comments that fell into a certain group were added up to form a recurrence computed for task versions as well as for each learner. Concerning the comments that linked to pausing and revision, the researcher used type of revision and pause site to compute frequency. The intercoder convention reached 95% after randomly selecting two learners and double-coding the data by the researcher's colleague.

Participants' Written Texts

Following Jarvis (2013), the current research planned to use lexical diversity in terms of paucity, variability, and disproportion. As the learners in the present research have to generate texts of the same length, Jarvis believed that, to be able to grasp lexical diversity, six substructures at least were needed, which are size, equivalence, dispersion, paucity, variability, and finally disproportion. Jarvis discovered that there is a high connection among three aspects, which are size, equivalence and dispersion. The researcher analyzed the written pieces generated by the learners according to a group of lexical diversity and syntactic complexity measures. The system of computer-based text analysis tools was utilized to measure the complexity indicators. Punctuation errors and misspellings were found in the written texts, which the researcher corrected before giving the texts to machine coding (Mazgutova, 2015).

Paucity was calculated according to ratio of the first and the second thousand most repeatedly utilized words in the English language, respectively (Cobb's online Vocabprofiler, 2016). In addition, the current study followed Coxhead (2000) and added the ratio of off-list words and ratio of academic words. The formulation of MTLD (McCarthy, 2010) was employed to calculate the textual lexical diversity, as well as the formula of Malven (1997) to evaluate the lexical variability. Building on a probabilistic mathematical pattern that employs a group of randomly sampled codes to produce a form - codes proportion curve versus increasing token volume, the value of D was carried out. Coh-Metrix 3.0 (McNamara et al., 2005) was utilized to gain the rates of MTLD and D after identifying MTLD as the mean length of sequential word strings that preserve a given sill of type–token proportion in a written work. Coh-Metrix 3.0 also created a latent semantic analysis (LSA) indicator used in the present research to evaluate variance, as suggested by Jarvis (2013). The mission of the LSA indicator was to keep in mind the semantic overlap among the words in the sentences to grasp conceptual resemblance of each sentence to other sentences in a written piece.

Following a study done by Bulté (2012), syntactic complexity was evaluated according to four types of indicators: syntactic sophistication, phrasal complexity, overall complexity, and subordination complexity. After former studies on task complexity in FL writing, this study decided to use the t-unit as the main unit of analysis. The general complexity was showed as the proportion of words in relation to t-units, while subordination complexity was employed as the ratio of clauses in terms of their relevance to t-units (Kuiken, 2008). Like Lu (2010), this research

used text analysis software SynLex to measure all these indicators as well as Coh-Metrix 3.0 for two reasons: firstly, to count the mean number of modifiers per noun phrase, and secondly, to calculate phrasal complexity. Another mission was done by Coh-Metrix 3.0 that employed a syntactic construction similarity indicator to evaluate the standard of syntactic sophistication. The advantage of this measure was to calculate consistency of syntactic structures in written work. This means that a lower syntactic construction similarity indicator shows more varied use of constructions.

Statistical Analyses

To examine the validity of task complexity manipulation in this study, the researcher analyzed the data from the perception questionnaire. To compare the responses of the learners in different conditions concerning using or not using content support, independent sample *t*-tests were used. To identify the outliers, the researcher checked the data for all measures of linguistic complexity and writing behaviors. It was clear that the outliers were shorted for values of two standard deviations from the mean per measure each set. In terms of pause length, the same threshold was used to determine and shorten outliers within learners. Additionally, to compare the impacts of task complexity on both linguistic complexity and writing behaviors, the researcher employed a chain of independent sample *t*-tests. Also, to compare the impacts of task complexity on the indicators of linguistic complexity (lexical and syntactic complexity) and writing behaviors (fluency, pausing, and revision), the researcher employed a chain of independent sample *t*-test. The current research used two measures: firstly, alpha standard for all tests set at .05, and secondly, to count the effect size, Cohen's d was used as a measure. As previously mentioned in a study by Plonsky (2014), the values larger than .40, .70, and 1.00 were deemed respectively as small, medium, and large. To investigate the connections between the measures of linguistic complexity and writing behaviors, Pearson correlations were employed. A more conservative alpha standard of .01 was utilized to limit the chance of Type 1 error as a result of the large number of connections. Following Plonsky's (2014) study, the current study used standard diagnostic procedures to ensure the suitability of using parametric connections and *t*-tests. In addition, the correlation coefficients were regarded as small at .25, medium at .40 and large at .60.

Results

The researcher divided this section into five parts as follows:

Part A: Validity Evidence for Task Complexity Manipulation

Part B: Task Complexity and FL Writing Behaviors

Part C: Task Complexity and Cognitive Processes Underlying FL Writing Behaviors

Part D: Task Complexity and Linguistic Complexity

Part E: Task Complexity, Revision Behaviors, and Linguistic Complexity

Part A

Validity Evidence for Task Complexity Manipulation

The current study used descriptive statistics through the nonstimulated recall learners in the perception questionnaire in terms of learners' perceptions about mental effort, awkwardness in planning, and task difficulty. The manipulation of task complexity in this research was along the same lines of this data, as the learners' performance in the complex task version was in contrast to their performance in the simple task version. The participants estimated the simple task version where they can find content support, which means less mental effort and less difficulty in planning the content of the essay, as well as the task being less difficult. There were significant differences in rating for the three scales as the independent sample t-test assured. The impact size was large for mental effort, t (126) = –6.42 , p = .002, 95% confidence interval (CI) = [–3.92, –0.9], d = 1.58; but the impact sizes were medium for both: difficulty of planning content, $t(126) = -4.06$, p = 0.018, 95% CI = [–3.62, –0.02], d = 1.02 and task difficulty in general $t(126) = -4.84$, p = 0.047, 95% CI = [–3.48, –0.34], d = 1.2. The following Table 1 provides descriptive statistics for perceptions of mental effort and task difficulty.

Table1 Descriptive statistics for perceptions of mental effort and task difficulty

Rated item	Simple (n =62)			Complex (n =68)		
	M	SD	95%CI	M	SD	95%CI
Mental effort	9.36	3.24	(8.26,10.48)	11.76	2.82	(10.84,12.64)
Task difficulty	7.74	2.9	(6.74, 8.76)	9.64	3.4	(8.44,10.84)
Difficulty in planning content	8.7	3.2	(7.64, 9.9)	10.52	3.94	(9.24,11.8)

Part B

Task Complexity and FL Writing Behaviors

It was clear that task complexity has a significant, medium-sized impact on only just two of the indicators: revisions below the word standard and the number of pauses between sentences. It was found that the revisions of the learners below the clause level were significant, and that pauses between sentences were significantly less frequent when the learners were furnished with ideas to be included in the essay, while there were no significant results in *t*-tests in terms of revision, fluency, and the residual measures of pausing. The following Table 2 gives the descriptive statistics for the measures of fluency, pausing, and revision behaviors for both the simple and complex task texts. The results of the independent sample t-tests that compared the behaviors of the learners under the terms of the simple and complex tasks are provided in the table as well.

Table 2 Descriptive statistics for the measures of fluency, pausing, and revision behaviors

Fluency	Simple (n =70)			Complex (n =76)			Comparison *t* test			
	M	SD	95%CI	M	SD	95%CI	t	p	95%CI	d
Minutes/ Word	0.08	0.02	(0.08,0.08)	0.1	0.02	(0.08,0.1)	-2.94	.15	(-0.018,0.002)	0.72
Minutes/ Character	0.02	0	(0.02,0.02)	0.02	0	(0.02,0.02)	-1.92	.34	(-0.002,0)	0.46
Words/P-burst	7.38	3.5	(6.26,8.7)	7.12	3.8	(6.1,8.56)	0.62	.76	(-1.38,1.84)	0.14
Characters/P-burst	45.52	21.78	(38.64,53.82)	46.1	29.6	(38.52,57.4)	0.18	.93	(-12.94,9.84)	0.04

Table2 Continued

Pause length in milli-seconds (log)	N	Simple (n =70)			N	Complex(n =76)			Comparison *t* test			
		M	SD	95%CI		M	SD	95%CI	t	p	95%CI	d
Gross	70	16.88	0.28	16.78,16.98	76	16.9	0.34	16.78,17.02	0.08	.97	-0.14,0.16	0.12
Within words	68	16.36	0.66	16.16,16.6	74	16.34	0.48	16.16,16.46	0.42	.84	-0.24,0.3	0.06
Between words	70	16.8	0.32	16.68,16.92	76	16.78	0.36	16.66,16.88	0.74	.71	-0.12,0.18	0.32
Between clauses	62	16.86	0.66	16.58,17.04	68	16.94	0.42	16.76,17.04	1.12	.58	-0.36,0.2	0.28
Between sentences	66	17.34	0.56	17.14,17.56	74	17.34	0.64	17.12,17.56	0.4	.84	-0.02,0.14	< 0.02
Number of pauses per 100 words (log)												

Pause length in milli-seconds (log)	Simple (n =70)				Complex(n =76)				Comparison *t* test			
	N	M	SD	95%CI	N	M	SD	95%CI	t	p	95%CI	d
Gross	70	6.58	0.78	6.26,6.78	76	6.68	1.14	(6.22,7.02)	-0.82	.69	-0.56,0.36	0.2
Within words	69	1.34	2	0.68,2.04	74	1.76	1.74	(1.22,2.3)	-1.9	.34	-1.3,0.46	0.44
Between words	70	5.58	1.42	5.1,6.06	76	5.46	1.82	(4.86,6.04)	0.68	.74	0.64,0.9	0.14
Between clauses	62	2.34	1.1	1.98,2.82	68	2.84	1.52	(2.28,3.34)	-3	.14	-1.16,0.16	0.67
Between sentences	70	2.5	2.5	2.18,3.02	76	3.12	1.38	(2.66,3.56)	-4.38	.03	-1.18,-.06	1.02
Revision overall (words/characters in product out of words/characters during process)												
Words		1.56	0.26	(1.48,1.64)		1.54	0.24	(1.46,1.62)	0.72	.72	-0.1,0.14	0.16
Characters		1.48	0.24	(1.4,1.56)		1.46	0.24	(1.38,1.52)	1.38	.50	-0.08,0.16	0.16
Number of revisions per 100 words (log)												
Below word	70	6.44	1.24	6.02,6.84	76	5.82	1.38	(5.38,6.28)	3.94	.05	0.02,1.22	0.94
Below clause	70	2.52	0.78	2.26,2.78	76	2.24	0.7	(2, 2.46)	3.06	.13	-0.08,0.62	0.76
Clause and above	54	-0.1	0.82	0.38,0.2	50	0.22	0.64	-0.46,0.04	1.24	.54	-0.28,0.54	0.32

Part C

Task Complexity and Cognitive Processes Underlying FL Writing Behaviors

When content support was obtainable in the simple case, the biggest ratio of stimulated recall comments pointed to translation processes, as it reached 59%. After that, comments that describing planning processes where they recorded 27% and came directly after translation operations. The situation was different in the complex case, where the percentage of stimulated recall comments that pointed to translation processes (33) were fewer than planning (48) as a result of the absence of content support. At the same time, monitoring was indicated by the learners with similar frequency in both simple (7%) and complex (6%) written pieces. The larger part of form comments concentrated on the mechanisms of lexical encoding (simple: 51%, complex: 61%), while the majority of comments on planning indicated planning content (simple: 83%, complex: 91%). Regardless of task complexity, the suboperations related to translation and planning were similar in their distributions. Additionally, the pauses between words reflected either translation or planning while more pauses between sentences were connected with planning only (simple: 67%, complex: 81%). In addition, when content support was available, the minor ratio of planning to connect pausing happened between words (13%) in contrast to preventing content support (22%), where the larger percentage of pausing was found. It can be concluded that absence of content support can badly affect planning operations. The following

Table 3 sums up the reasons for pausing (number of comments) from stimulated recall comments in order to examine the cognitive operations underlying learners' pausing behavior for both the complex and simple written work. For participant-level breakdown of the data, see Appendix C.

Table3. Reasons for pausing (number of comments) from stimulated recalls

Pause location in simple tasks	Planning			Translation				Monitoring	No recall	All
	Cont.	Org.	All	Lex. Ret.	Syn. Enc.	Cohesion	All			
Within words	2	0	2 (1%)	4	2	0	8 (2%)	0 (0%)	14 (4%)	24 (7%)
Between words	50	2	52 (13%)	110	8	6	186 (47%)	2 (1%)	14 (4%)	254 (65%)
Between clauses	4	0	4 (1%)	6	6	0	24 (6%)	2 (1%)	2 (1%)	32 (8%)
Between sentences	32	16	48 (13%)	0	2	2	16 (5%)	24 (7%)	2 (1%)	90 (26%)
Total	88	18	106 (27%)	120	18	8	234 (59%)	28 (7%)	32 (8%)	400 (100%)
Pause location in complex tasks										
Within words	4	0	4 (1%)	6	2	0	12 (4%)	0 (0%)	16 (6%)	32 (11%)
Between words	60	2	62 (22%)	48	4	0	64 (22%)	0 (0%)	20 (7%)	146 (51%)
Between clauses	18	0	18 (6%)	0	0	0	8 (3%)	0 (0%)	2 (1%)	28 (10%)
Between sentences	44	10	54 (18%)	4	0	2	10 (3%)	16 (5%)	2 (1%)	82 (27%)
Total	126	12	138 (48%)	58	6	2	94 (33%)	16 (6%)	40 (14%)	288 (100%)

Note. Cont. = content, org. = organization, Lex. Ret. =Lexical retrieval, Syn. Enc. = Syntactic encoding.

The data that appeared in the following Table 4 displays some adopted comments from stimulated recall to characterize the ideas that come to the learners' minds during revision. For learner-level breakdown of the data, see Appendix D. The percentage for both cases in planning was simple: 19%, complex: 26%, while the ratio of the same cases in translation was different (simple: 78, complex: 72%). There was a slightly higher total ratio of translation that connected to comments in the simple task versions. Additionally, some other models of revision were broadly comparable in both simple and the complex task versions, such as below word and below sentence.

Table 4 Reasons for revision (number of comments) from stimulated recalls

Revision location in simple task	Planning			Translation				No recall	Total 3
	Cont.	Org.	All 2	Lex. Ret.	Syn. Enc.	Cohesion	All		
Below word	10	0	10 (2%)	4	2	4	26 (5%)	4 (1%)	40 (8%)
Single word 1	22	0	22 (4%)	86	10	22	212 (43%)	10(2%)	244 (49%)
Below clause	42	0	42 (9%)	30	10	14	102 (21%)	4 (1%)	148 (31%)
Clause and above	16	2	18 (4%)	8	2	4	40 (8%)	0 (0%)	58 (12%)
Total	90	2	92 (19%)	128	24	44	380 (78%)	18 (4%)	490 (100%)
Revision location in complex task									
Below word 1	8	0	8 (2%)	12	10	0	48(10%)	4 (1%)	60 (13%)
Single word	32	0	32 (7%)	40	26	0	144 (29%)	4 (1%)	180 (37%)
Below clause	38	2	40 (8%)	20	14	22	126 (26%)	6 (1%)	172 (35%)
Clause and above	38	8	46 (9%)	4	4	0	34 (7%)	0 (0%)	80 (16%)
Total	116	0	126 (26%)	76	54	22	352 (72%)	14 (3%)	492 (100%)

Note. Cont. = content; org. = organization; Lex. Ret. =Lexical retrieval; Syn. Enc. = Syntactic encoding.
1= Single complete word was inserted, replaced or canceled. 2= some of the participants' comments given were not specific enough to permit for more sub-categorization, so there is no need to add subcategories' values to the total. 3= Do not gather up to 100 because of rounding some scores.

Part (D)

Task Complexity and Linguistic Complexity

The learners generated a higher number of words per t-unit and utilized significantly larger ratios of K2 words, but smaller ratios of K1 words in the simple task where the content was provided. This finding means that the participants used of less frequent words and greater total complexity more widely because of reduced task complexity. Nonetheless, task complexity did not have any significant effect on most of the linguistic complexity measures, but the impact sizes for these lexical and syntactic complexity indicators were of large and medium size, respectively. The following Table 5 showed that the significant results appeared in three tests only. The table presents the descriptive statistics and results of independent sample t-tests for lexical diversity and syntactic complexity across the complex and simple task models.

Table 5 Descriptive statistics for lexical diversity and syntactic complexity

Lexical diversity	Simple (n =70)			Complex(n =76)			Comparison *t* test			
	M	SD	95%CI	M	SD	95%CI	t	p	95%CI	d
K1 words	174.62	8.3	84.93,81.63	181.48	8.18	88.45,92.96	-7.12	. 001	-4.35,-2.51	1.88
K2 words	6.74	2.96	2.90,3.90	4.12	2.42	1.65,2.45	8.32	<.001	.66,1.96	1.98
Academic words	10.28	4.5	4.42,5.90	8.48	4.02	2.60,5.89	3.62	.08	-.07,1.91	0.86
Off-list words	7.84	4.62	3.25,4.67	6.06	4.48	2.33,3.72	3.36	.10	-.16,1.97	0.8
MTLD	134.86	27.66	62.17,72.79	133.62	27.9	60.53,73.30	0.38	.85	-5.85,7.14	0.1
D value	137.48	30.64	63.04,74.71	135.94	32.96	60.66,75.23	0.4	.84	-5.68,9.21	0.1
LSA	0.4	0.1	0.18,0.22	0.42	0.14	.16,.25	-0.84	.68	-0.03,0.02	0.2
Syntactic complexity										
Words/t-unit	44.54	8.38	19.69,24.70	39.12	7.92	17.31,21.86	5.68	.006	0.80,4.62	1.34
Clause/t-unit	4.24	0.84	1.99,2.26	3.96	0.76	1.80,2.4	3.02	.14	-.03,.34	0.72
Modifiers per NP	1.82	0.26	.85,.96	1.7	0.28	.80,.91	3.38	.09	-0.01,0.12	0.8
Structural similarity	0.14	0.04	.06,.09	0.16	0.04	.08,.08	-2.68	.18	-0.02,.003	0.64

Note. Based on Malvern and Richards (**1997**).MTLD = measure of textual lexical diversity; LSA = latent semantic analysis; D value = measure of lexical variability.

Part E

The relation between Linguistic Complexity and Revision Behaviors

Concerning the relation between linguistic complexity and revision behaviors, the current study found that there were three significant interconnections, one of them in the simple essay, while the other two correlations were identified in the complex essay. In the simple condition, diverse syntactic constructions were lower for those learners who paused longer between clauses, while in the complex condition, the learners who paused longer between sentences generated texts with a less developed lexicon. The infrequent words used in the essays as well as the lower number of off-list stand as evidence of that. In addition, at the clause level and above, more revision was linked with less developed lexical options. That was obvious in the smaller ratio of academic words that were used in the texts. The following Table 6 sums up the considerable relations between linguistic complexity measures and writing behavior indicators.

Table 6 Correlations (Person r) between linguistic complexity measures and writing behavior

Writing behavior	Linguistic complexity	R	95%CI	P
Simple				
Pause length between clauses (log)	Structural similarity	0.92	.14,.70	.010
Complex				
Pause length between sentences (log)	Off-list words	-0.94	-.65,-.24	.003
Revisions clause level and above	Academic words	-1	-.72,-.24	.005

Note. R. = Significant correlations (Pearson *r*).

Discussion

The current study posed three research questions; the first question tried to investigate the impact of task complexity, fluency, pausing, revision behaviors and the cognitive processes underlying these behaviors if any when giving content support to the participants. The results revealed that task complexity had a significant impact on one revision indicator and one pausing, while it did not have any significant effect on fluency, and this result was supported by a group of independent sample t- tests. The learners' behavior in terms of revising below the word level and pausing more frequently between sentences was due to the absence of ideas in complex task versions. Building on whether ideas were accessible, there was difference in the processes underlying pausing and revision behaviors, as was clear from the stimulated recall information. The stimulated recall comments that were obtained from the participants' response to pausing behaviors showed that the absence of ideas led to more planning than translation- linked pauses and revisions.

The results of the current study concerning speed fluency are in line with these of Spelman Miller (2000) and Révész (2017). According to Spelman Miller, the relative reluctance of this construct to task differences in writing might stand behind the deficiency of impacts for speed fluency. Following Spelman Miller's suggestion, a study by De Jong et al. (2013) revealed that linguistic skills and knowledge had the most powerful connection with speed fluency between measures of speed, repair fluency, and breakdown in FL speaking. Revesz (2017) applied the previous idea to FL writing, and thus he considered that the effect of proficiency (a variable controlled for in the current study) on speed

fluency was large and incomparable with the magnitude of the impact of task complexity, which was negligible.

The current study relies on Kellogg's (1996) expectations in his model in terms of the pressure that would increase on planning as a result of the absence of ideas. This action would make pausing at higher-level discourse units wider. Schilperoord (1996) considered the incidence of pauses at a higher level overwhelmingly reflects higher-order writing processes like creating content. There was significant evidence that pauses happened more between sentences in the complex condition, as detected by the keystroke logs. In most conditions, pauses at sentence limits were connected with planning processes, as confirmed by the stimulated recall comments. The effort demanded in translation-related stimulated recall comments generated in the complex case was not obvious in the overall proportionately higher planning, in contrast to the planning stage. The current study agreed with the observation of Ong (2014) that learners were more involved in metacognitive operations, like organizing ideas when content support was available. At the same time, the current results are in accord with those of Révész (2017) in denying the findings of Spelman Miller (2000), which revealed that there were no impacts on task complexity in checking the frequency of pauses.

The findings of the current study concerning the length of the pause revealed that task complexity did not have any significant correlation with the period of the pause. Thus, this result agreed with these of Spelman Miller (2000) and Révész, (2017). In the context of pausing, both Spelman Miller (2008) and De Jong (2013) revealed that pause frequency decreased over time while pause period remained stable, and there was no correlation between linguistic skills in FL speaking and duration of silent pauses. De Jong (2013) predicted that length of pause might be identified through personal writing style or personality characteristics because the length of time writers' pause, on average, might be resistant to elements like proficiency and task complexity.

Concerning pausing at word boundaries, the distribution of both planning and translation, which is linked to stimulated recall comments, was different in the simple and complex cases. On the other hand, pausing at sentence boundaries was different as the learners neglected task complexity in favor of planning-linked operations when they were asked to recall their ideas during pauses between sentences. The comments describing both translation and planning processes were produced through pauses between words in the simple case, but the

comments describing planning processes were less than for translation. This indicates that pausing at higher discourse units is more likely to be related to higher-level writing operations, even when providing enough cognitive resources to permit for translation and planning processes. This result is comparable to what was suggested by Schilperoord (1996) in his study of L1 writing. Certainly this point needs more study to verify this result, particularly in terms of lower proficiency for writers.

The results regarding revision behavior showed that the learners would have to have fewer attentional resources to assign to monitoring and translation, because of increased requests made on planning operations in the absence of content support. As a result, the language revisions were expected to be reduced. The stimulated recall comments recorded a larger proportion at the level of word-described translation than planning operations in simple and complex groups. At the same time, significantly more below-word revisions were recorded in the keystroke logs in the simple case only. It can be concluded from the previous findings that the learners in the low-task complexity case made more language-linked revisions below the word level. A larger ratio of the revision comments pointed to translation than planning-related mechanisms, which affirmed the previous conclusion through the stimulated recall comments. The current study agrees with both Thorson's (2000) and Révész's (2017) results, where there were no significant differences between trends for revision below the clause level and clause and above. In addition, there were no task impacts revealed for overall amount of revisions.

To answer the second research question, the researcher used indicators of syntactic complexity and lexical diversity to determine the impact of task complexity on the linguistic complexity of FL written works if any. Based on Kellogg's (1996) model, the current study anticipated that the presence of content support would facilitate increased linguistic complexity, as well as that all the independent sample t-tests would support this expectation. The learners were able to present written works with considerable higher-ranking total syntactic complexity and a more developed vocabulary. The lexical complexity results in this study were in contrast with Kormos's (2011) findings, where she found no impacts for syntactic complexity, while both studies were completely in agreement in terms of their results for lexical complexity. In addition, these results agreed with both studies of Ong (2010) and Révész (2017), where similar trends for revision outlines were found. In spite of the premier, outlines did not show these models as happened in the current study.

The last question posed in this study aimed to investigate the relationships between linguistic complexity and FL writing behaviors if there were such relationships and whether task complexity moderates these relations. The findings showed that when the ideas were not available, as well as decreasing attentional resources, the learners were less able to make language-related revisions, and they would be likely to pause to involve in planning phase. This result supports the anticipation of the current study in that revision behaviors and pausing would possibly not be positively connected to linguistic complexity in the high-task complexity group.

In the absence of content support for the simple case, there were no positive relations between the measures of pausing and revision and linguistic complexity. Therefore, at the clause level and above, the participants used a less developed lexicon to deal with considerable amounts of pausing between sentences and revision. The current study anticipated this fact as a result of the reduced requests on planning processes; thus, linguistic complexity would be positively connected to the amount of pausing and revisions. This discovery correlates with the fact that the learners in the complex group indicated that when they were required to recall their ideas during revisions at the clause level and above and pauses between sentences, they were more often associated with planning than translation-related processes compared to the simple group. The participants in the simple case who paused more between clauses used less varied syntax as a result of allocating more time for planning; leaving fewer resources for lexical encoding operations, and this was contrary to the expectation of the current study. The only potential clarification for this result is that greater length of pausing between sentences was an aspect of less sophisticated syntactic knowing. The results of the current study showed that there was agreement between the studies of Spelman Miller et al. (2008) and Révész (2017) in terms of fluency, while they were at odds with the studies of Stevenson et al. (2006) and Spelman Miller et al. (2008) concerning writing work connected to revision behaviors and pausing.

Implications of the study

The importance of the current study results can be sorted into three levels: pedagogical, methodological, and theoretical. On a theoretical level, the results of the current study had proven the anticipations acquired from Kellogg's cognitive pattern (1996). This means that this pattern can be used as a starting point for conceptualizing research on task-produced FL pausing and revision behaviors, as well as the

underlying related operations. The importance of this point comes from the results of other studies (Tavakoli, 2014; Manchón, 2014) regarding the suitability of utilizing task-based patterns of FL speech outcome as a theoretical backbone for exploring task-based writing. At the pedagogical level, the current study tried to support providing ideas that may help the participants pay more attention to linguistic encoding operations, which might lead to expanding their interlanguage, in spite of the probability of showing genuineness. At the methodological level, the current study showed that the ability to produce more accurate and valid explanations concerning task-created operations than depending on just one data source was because of the connection between keystroke logging and stimulated recall .

Conclusions

The main purpose of the current study was to find out whether the manipulations of the task complexity have any impact on the online behaviors of FL writers and the connected cognitive operations underlying them. This was a trial to address the gap in present task-based research on expository operations, mediating the connection between the linguistic output of the writing operation and task complexity. Additionally, the current study aimed to extend the present research in terms of exploring the impacts of task complexity on linguistic complexity, as well as looking into whether task complexity may moderate the connections between linguistic complexities and writing behaviors. In terms of task-based research, the methodological contribution of the current study was in using the trilogy of information gained from computer-based textual analysis, keystroke logging, and stimulated recall. The findings indicated that providing ideas led to less frequent pausing and greater amount of revision and resulted in increased lexical complexity, and these findings meshed largely with the anticipations of Kellogg's (1996) pattern of writing. The availability of giving ideas leading to more frequent pauses was also linked to the production of more lexically complex language. This means that the writers devoted more awareness to linguistic encoding operations, which in turn led to the noticed impacts on the quality of the text and writing behaviors.

Suggestions for Future Research

The current study used two seconds as a pause threshold to make it easy to compare the previous studies to the current one, as it was considered as an exemplary threshold in writing research until now. But at the same time it was considered as a methodological weakness and was attacked by some other researchers. In a study by Baaijen et al. (2012), it

was deemed that this process restricts the analysis to longer pauses, which are likely to reflect higher-level writing operations. This means that the shorter pauses that are more often linked with lower-level processing will be precluded, which is another shortcoming related to the between-subjects design used in the current study. Despite the fact that the current study found that the two groups were similar, future researchers could ask participants to generate sundry written works to raise the generalizability of the results, which means that the researchers could use within-subjects designs. The use of the stimulated recall methodology is deemed as another restriction. In spite of treating cases with reactivity, an inherent limitation in this proceeding is that it can yield data only about conscious processes. Additionally, during writing the learners will be able to recall only one group of conscious operations, because of memory deterioration.

The researchers can utilize introspective protocols together with information from eye-tracking to alleviate this issue. Brunfaut (2015) argued that this combination would help the researchers to gain a more complete picture of the operations taking place during the output of the FL written work. In addition, this will give insights for involved conscious and some unconscious processes. Analyzing revision drafts, in addition to the output of the first written work, would lead to a more complete picture of the writing operation that can be a starting point for future studies. Another interesting focus for future research is that the researchers could stretch the questions of the research here to other task complexity manipulations, populations, and task types. Future researchers can investigate the findings of the current study to discover whether these results would transfer to other task types, lower-proficiency participants, and different task complexity manipulations.

References

Abdel-Latif, M. M. (2013). What do we mean by writing fluency and how can it be validly measured? *Applied Linguistics*, *34*, 99–105.

Baaijen, V. M., Galbraith. D., & de Glopper, K. (2012). Keystroke analysis: Reflections on procedures and measures. *Written Communication*, *29*, 246–277.

Brown, J. D., Hudson, T. D., Norris, J. M., & Bonk, W. (2002). *An Investigation of Second Language Task-Based Performance Assessment*. Honolulu: University of Hawai'i Press.

Brunfaut, T., & McCray, G. (2015). Looking into test takers' cognitive processes whilst completing reading tasks: A mixed-method eye-tracking and stimulated recall study. ARAGs Research Reports Online, AR/2015/001.

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

Byrnes, H., & Manchón, R. (Eds.). (2014). *Task-based Language Learning: Insights from and for L2 Writing*. Amsterdam: John Benjamins.

Cobb,T. (2016). Vocab Profiler[Computer software]. Retrieved from http://www.lextutor.ca/vp/eng

Coxhead, A. (2000). A new academic word list. *TESOL Quarterly*, *34*, 213–238.

De Jong, N. H., Steinel, M. P., Florijn, A., Schoonen, R., & Hulstijn, J. H. (2013). Linguistic skills and speaking fluency in a second language. *Applied Psycholinguistics*, *34*, 893–916.

Ellis, R., & Yuan, F. (2004). The effects of planning on fluency, complexity and accuracy in second language narrative writing. *Studies in Second Language Acquisition*, *26*, 59–84.

Heatley, A., Nation, I. S. P., & Coxhead, A. (2002). Range and frequency programs [Computer software]. Retrieved from http://www.victoria.ac.nz/lals/resources/range.aspx

Jarvis, S. (2013). Defining and measuring lexical diversity. In S. Jarvis & M. Daller (Eds.), *Vocabulary Knowledge: Human Ratings and Automated Measures* (pp. 13–45). Amsterdam: John Benjamins.

Kellogg, R. T. (1990). Effectiveness of prewriting strategies as a function of task demands. *American Journal of Psychology*, *103*, 327–342.

Kellogg, R. T. (1996). A model of working memory in writing. In C. M. Levy & S. Ransdell (Eds.), *The Science of Writing: Theories, Methods, Individual Differences and Applications* (pp. 57–71). Mahwah, NJ: Erlbaum.

Kormos, J. (2011). Task complexity and linguistic and discourse features of narrative writing performance. *Journal of Second Language Writing*, *20*, 148–161.

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.

Leijten, M., & Van Waes, L. (2013). Keystroke logging in writing research: Using Inputlog to analyze and visualize writing processes. *Written Communication*, *30*, 358–392.

Lu, X. (2010). Automatic analysis of syntactic complexity in second language writing. *International Journal of Corpus Linguistics*, *15*, 474–496.

Macaro, E. (2014). Reframing task performance: The relationship between tasks, strategic behavior, and linguistic knowledge in writing. In H. Byrnes & R. Manchón (Eds.), *Task-Based Language Learning: Insights from and for L2 Writing* (pp. 53–77). Amsterdam: John Benjamins.

Malvern, D. D., & Richards, B. J. (1997). A new measure of lexical diversity. In A. Ryan & A. Wray (Eds.), *Evolving Models of Language* (pp. 58–71). Clevedon, UK: Multilingual Matters.

Manchón, R. (2014). The internal dimension of tasks: The interaction between task factors and learner factors in bringing about learning through writing. In H. Byrnes & R. Manchón (Eds.), *Task-Based Language Learning: Insights from and for L2 Writing* (pp. 27–52). Amsterdam: John Benjamins.

Mazgutova, D., & Kormos, J. (2015). Syntactic and lexical development in an intensive English for academic purposes programme. *Journal of Second Language Writing*, *29*, 3–15.

McCarthy, P., & Jarvis, S. (2010). MTLD, vocd-D, and HD-D: A validation study of sophisticated approaches to lexical diversity assessment. *Behavior Research Methods*, *42*, 381–392.

McNamara, D. S., Louwerse, M. M., Cai, Z., & Graesser, A. (2005, January 1). Coh-Metrix version 1.4 [Computer software]. Retrieved from http://cohmetrix.memphis.edu

Ong, J. (2014). How do planning time and task conditions affect metacognitive process of L2 writers? *Journal of Second Language Writing*, *23*, 17–30.

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

Plonsky, L., & Oswald, F. L. (2014). How big is "big"? Interpreting effect sizes in L2 research. *Language Learning*, *64*, 878–912.

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

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

Robinson, P. (Ed.). (2011). *Second Language Task Complexity: Researching the Cognition Hypothesis of Language Learning and Performance*. Amsterdam: John Benjamins.

Ruiz-Funes, M. (2014). Task complexity and linguistic performance in advanced college-level foreign language writing. In H. Byrnes & R. Manchón (Eds.), *Task-Based Language Learning: Insights from and for L2 Writing* (pp. 163–191). Amsterdam: John Benjamins.

Schilperoord, J. (1996). *It's About Time: Temporal Aspects of Cognitive Processes in Text Production*. Amsterdam: Rodopi.

Schoonen, R., Snellings, P., Stevenson, M., & van Gelderen, A. (2009). Towards a blueprint of the foreign language writer: The linguistic and cognitive demands of foreign language writing. In R. M. Manchón (Ed.), *Learning, Teaching, and Researching Writing in Foreign Language Contexts*. Bristol, UK: Multilingual Matters.

Shintani, N. & Ellis, R., (2013). *Exploring Language Pedagogy through Second Language Acquisition Research*. New York: Routledge.

Skehan, P. (1998). *A Cognitive Approach to Language Learning*. Oxford, UK: Oxford University Press.

Skehan, P. (Ed.). (2014). *Processing perspectives on task performance*. Amsterdam: John Benjamins.

Spelman Miller, K. (2000). Academic writers on-line: Investigating pausing in the production of text. *Language Teaching Research*, *4*, 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*, 433–453.

Stevenson, M., Schoonen, R., & Glopper, K. (2006). Revising in two languages: A multi-dimensional comparison of online writing revisions in L1 and FL. *Journal of Second Language Writing*, *15*, 201–233.

Tavakoli, P. (2014). Storyline complexity and syntactic complexity in writing and speaking tasks. In H. Byrnes & R. Manchón (Eds.), *Task-Based Language Learning: Insights from and for L2 Writing* (pp. 163–191). Amsterdam: John Benjamins.

Thorson, H. (2000). Using the computer to compare foreign- and native-language writing processes: A statistical and case study approach. *Modern Language Journal*, *84*, 55–70.

Tillema, M. (2012). *Writing in first and second language: Empirical studies on text quality and writing processes*. Unpublished doctoral thesis, Netherlands Graduate School of Linguistics, Utrecht, Netherlands.

Van Waes, L., & Leijten, M. (2015). Fluency in writing: A multidimensional perspective on writing fluency applied to L1 and L2. *Computers and Composition*, *38*, 79–95.

Wengelin, A. (2006). Examining pauses in writing: Theory, methods and empirical data. In K. H. Sullivan & E. Lindgren (Eds.), *Computer*

Keystroke Logging and Writing: Methods and Applications (pp. 107–130). Oxford, UK: Elsevier Science.

Appendix A
Descriptive and Inferential Statistics for Lexical Diversity and Syntactic Complexity on the Trinity ISE III Correspondence Task

Trinity Ratings	Simple (n =70)			Complex (n =76)			Comparison *t* test			
	M	SD	95%CI	M	SD	95%CI	t	p	95%CI	d
Task fulfillment	4	1.54	(1.72,2.26)	4.18	1.66	(1.82,2.34)	-0.88	.66	(-0.49,-31)	0.22
Organization & structure	3.94	1.6	(1.70,2.25)	4.48	1.4	(2.00,2.46)	-2.88	.15	(-.64,.10)	0.72
Language control	2.12	2	(.71,1.42)	1.94	1.88	(.65,1.29)	0.78	.70	(-39,.57)	0.18
Lexical Diversity										
K1 words	172.38	6.68	85.15,87.63	173.16	6.84	85.41,87.65	.98	.63	–1.96, 1.19	0.24
K2 words	8.12	2.46	3.68, 4.46	8.12	2.38	3.66, 4.44	0.04	.99	–.56, .57	<.02
Academic words	11.02	2.46	4.97, 5.97	10.56	4.1	4.61, 5.96	1.06	.60	–.63, 1.08	0.24
Off-list words	8.48	4.46	3.52, 4.96	8.16	4.04	3.44, 4.82	0.64	.75	–.84, 1.16	0.16
MTLD	157.72	40.86	72.57,85.40	160.42	38.14	74.37,86.66	0.58	.77	–10.59,7.90	0.14
D value	162.42	36.38	75.95,86.91	162.12	36.28	75.58, 87.41	0.08	.97	–8.34, 8.64	<.02
LSA	0.36	0.08	.17, .19	0.36	0.1	.16, .19	0.42	.83	–.02, .02	<.02
Syntactic complexity										
Words/t-unit	36.52	6.42	17.19,19.32	38.98	8.38	18.11,20.90	-2.8	.17	-3.00, .52	0.64
Clause/t-unit	3.76	0.64	1.78,1.99	3.68	0.66	1.73, 1.96	1.18	.56	-.11, .20	0.24
Modifiers per NP	1.68	0.28	.80, .88	1.62	0.26	.77,.85	1.78	.38	-.03, .09	0.44
Structural similarity	0.16	0.04	[.07, .08]	0.16	0.06	.07,.08	0.34	.87	-.01, .01	<.02

Appendix B
Stimulants for Simple and Complex Cases

Stimulant for Simple Case (200-250 words)

After a discussion about the importance of scientists and technology experts compared to musicians and artists, you have been asked to write an essay giving your opinions on the topic:

"Scientists and technology experts seem to be more valued by modern society than musicians and artists. Do you agree?"

In your essay, please address the issues below. You do not have to use all the examples suggested in the bullet points. Please select some of them and expand on those.

-If you agree with the statement, include the following information:

- What type of information can we learn about the importance of scientists and technology experts in the 20th century? For example, about their lifestyles, their mistakes or failures, their inventions.
- What are the advantages of learning about scientists and technology experts?

For example,

-we can better connect our own life to their inventions

-we can appreciate learning

-we can learn valuable lessons and not repeat past mistakes in the future

-If you disagree with the statement, include the following information:

-What kind of information can we learn about the importance of musicians and artists? For example, about their lifestyles, their habits, their customs, their achievements, their mistakes or failures

What are the advantages of learning about musicians and artists?

For example,

-we can make a comparison of the behavior the countries that neglect art and music and the other countries that value them

-we can appreciate the value of art and music in society

-we can learn valuable lessons and not repeat past mistakes in the future

Stimulant for Complex Case (200-250 words, 45 minutes)

After a discussion of the importance of scientists and technology experts compared to musicians and artists, you have been asked to write an essay giving your opinions on the topic:

"Scientists and technology experts seem to be more valued by modern society than musicians and artists. Do you agree?"

Appendix C, part 1

Reasons for Pausing: Summary of Stimulated Recall Comments for simple essay

Lea. / Loc.	Simple essay														
	Planning				Translation										
	Con.	Org.	Tot.	Total *	Lex.	Syn.	Coh.	Tot.		Moni.		Don't Re.		overall	
	N	N	N	%	N	N	N	N	%	N	%	N	%	N	
Lea. 1	28	2	30	20%	38	2	4	90	61%	22	15%	6	4%	148	100 %
ww	0	0	0	0%	2	0	0	2	1%	0	0%	2	1%	4	2%
bw	16	0	16	11%	34	0	2	74	50%	2	1%	4	3%	96	65%
bc	4	0	4	3%	2	0	0	8	5%	2	1%	0	0%	14	9%
bs	8	2	10	6%	0	2	2	6	4%	18	12%	0	0%	34	22%
Lea. 3	18	2	20	32%	12	8	0	32	52%	0	0%	10	16%	62	100 %
ww	0	0	0	0%	2	0	0	0	0%	0	0%	4	6%	4	6%
bw	6	0	6	10%	10	4	0	20	32%	0	0%	2	3%	28	45%
bc	0	0	0	0%	0	4	0	4	6%	0	0%	2	3%	6	9%
bs	12	2	14	23%	0	0	0	8	13%	0	0%	2	3%	24	39%
Lea. 5	28	14	42	36%	40	4	4	68	58%	4	3%	4	3%	118	100 %
ww	2	0	2	2%	0	0	0	2	2%	0	0%	2	2%	6	6%
bw	18	2	20	17%	28	2	4	56	47%	0	0%	2	2%	78	66%
bc	0	0	0	0%	2	2	0	8	7%	0	0%	0	0%	8	7%
bs	8	12	20	17%	0	0	0	2	2%	4	4%	0	0%	26	23%
Lea. 7	14	0	14	19%	30	4	4	44	61%	2	3%	12	17%	72	100 %
ww	0	0	0	0%	0	2	0	4	6%	0	0%	6	8%	10	14%
bw	10	0	10	14%	28	2	4	36	50%	0	0%	6	8%	52	72%
bc	0	0	0	0%	2	0	0	4	6%	0	0%	0	0%	4	6%
bs	4	0	4	6%	0	0	0	0	0%	2	3%	0	0%	6	9%
Tot	88	18	106	27%	120	18	8	234	59%	28	7%	32	8%	400	100 %
ww	2	0	2	1%	4	2	0	186	2%	0	0%	14	4%	24	7%
bw	50	2	52	13%	110	8	6	6	47%	2	0%	14	4%	254	65%
bc	4	0	4	1%	6	6	0	24	6%	2	1%	2	1%	32	8%
bs	32	16	48	13%	0	2	2	16	5%	24	7%	2	1%	90	26%

Lea = learner, Loc = pause location, ww = within words, bw = between words, bc = between clauses, bs = between sentences, bp = between paragraphs, Con =

content, Org =organization, Lex = lexical retrieval, Syn = syntactic encoding, Tot = total, Don't Re.= don't remember, moni= monitoring; *Due to rounding some totals do not add up to 100.

Appendix C, part 2

Reasons for Pausing: Summary of Stimulated Recall Comments for Complex Essay

Lea./ Loc.	Complex essay														
	Planning				Translation										
	Con.	Org.	Tot.		Lex.	Syn.	Coh.	Tot.	Moni.		Don't Re.		Total overall *		
	N	N	N	%	N	N	N	N	%	N	%	N	%	N	%
Lea. 2	38	8	46	42%	20	6	0	32	29%	6	5%	26	24%	110	100 %
ww	2	0	2	2%	4	2	0	6	5%	0	0%	12	11%	20	18%
bw	26	2	28	25%	16	4	0	22	20%	0	0%	12	11%	62	56%
bc	8	0	8	7%	0	0	0	4	4%	0	0%	2	2%	14	13%
bs	2	6	8	7%	0	0	0	0	0%	6	5%	0	0%	14	12%
Lea. 4	22	2	24	55%	10	0	0	16	36%	0	0%	4	9%	44	100 %
ww	2	0	2	5%	2	0	0	2	5%	0	0%	2	5%	6	15%
bw	10	0	10	23%	8	0	0	14	32%	0	0%	2	5%	26	60%
bc	0	0	0	0%	0	0	0	0	0%	0	0%	0	0%	0	0%
bs	10	2	12	27%	0	0	0	0	0%	0	0%	0	0%	12	27%
Lea. 6	48	2	50	60%	18	0	2	34	40%	0	0%	0	0%	84	100 %
ww	0	0	0	0%	0	0	0	4	5%	0	0%	0	0%	4	5%
bw	16	0	16	19%	14	0	0	18	21%	0	0%	0	0%	34	40%
bc	4	0	40	5%	0	0	0	4	5%	0	0%	0	0%	8	10%
bs	28	2	30	35%	4	0	2	8	10%	0	0%	0	0%	38	45%
Lea. 8	18	0	18	36%	10	0	0	12	24%	10	20%	10	20%	50	100 %
ww	0	0	0	0%	0	0	0	0	0%	0	0%	2	4%	2	4%
bw	8	0	8	16%	10	0	0	10	20%	0	0%	6	12%	24	48%
bc	6	0	6	12%	0	0	0	0	0%	0	0%	0	0%	6	12%
bs	4	0	4	8%	0	0	0	2	4%	10	20%	2	4%	18	36%
Tot	126	12	138	48%	58	6	2	94	33%	16	6%	40	14%	288	100 %
ww	4	0	4	1%	6	2	0	12	4%	0	0%	16	6%	32	11%
bw	60	2	62	22%	48	4	0	64	22%	0	0%	20	7%	146	51%
bc	18	0	18	6%	0	0	0	8	3%	0	0%	2	1%	28	10%
bs	44	10	54	18%	4	0	2	10	3%	16	5%	2	1%	82	27%

Lea = learner, Loc = pause location, ww = within words, bw = between words, bc = between clauses, bs = between sentences, bp = between paragraphs, Con = content, Org =organization, Lex = lexical retrieval, Syn = syntactic encoding, Tot = total, Don't Re.= don't remember, moni= monitoring; *Due to rounding some totals do not add up to 100.

Appendix D, part 1

Reasons for Revision: Summary of Stimulated Recall Comments for Simple Essay

Lea./ Loc.	Simple essay												
	Planning				Translation								
	Con.	Org.	Tot.		Lex.	Syn.	Coh.	Tot.		Don't Re.		Total overall *	
	N	N	N	%	N	N	N	N	%	N	%	N	%
Lea.1	46	0	46	20%	58	4	18	176	76%	10	4%	232	100%
bw	8	0	8	3%	4	0	2	16	7%	4	2%	28	12%
w	16	0	16	7%	30	2	10	94	41%	4	2%	114	50%
bc	14	0	14	6%	18	2	4	44	19%	2	1%	60	26%
ca	8	0	8	3%	6	0	2	22	9%	0	0%	30	12%
Lea.3	4	2	6	10%	8	6	18	52	84%	4	6%	62	100%
bw	0	0	0	0%	0	0	2	4	6%	0	0%	4	6%
w	2	0	4	3%	0	4	10	20	32%	2	3%	24	38%
bc	2	0	2	3%	6	2	4	20	32%	2	3%	24	38%
ca	0	2	0	3%	2	0	2	8	13%	0	0%	10	16%
Lea.5	24	0	24	20%	30	8	6	94	80%	0	0%	118	100%
bw	2	0	2	2%	0	0	0	6	5%	0	0%	8	7%
w	0	0	0	0%	28	2	2	64	54%	0	0%	64	54%
bc	18	0	18	15%	2	6	4	24	20%	0	0%	42	35%
ca	4	0	4	3%	0	0	0	0	0%	0	0%	4	3%
Lea.7	16	0	16	21%	32	6	2	58	74%	4	5%	78	100%
bw	0	0	0	0%	0	0	0	0	0%	0	0%	0	0%
w	4	0	4	5%	28	2	0	34	44%	4	5%	42	54%
bc	8	0	8	10%	4	0	2	14	18%	0	0%	22	28%
ca	4	0	4	5%	0	2	0	10	13%	0	0%	14	18%
Tot	90	2	92	19%	128	24	44	380	78%	18	4%	490	100%
bw	10	0	10	2%	4	2	4	26	5%	4	1%	40	8%
w	22	0	22	4%	86	10	22	212	43%	10	2%	244	49%
bc	42	0	42	9%	30	10	14	102	21%	4	1%	148	31%
ca	16	2	18	4%	8	2	4	40	8%	0	0%	58	12%

Lea = learner, Loc = pause location, bw = below word, w = one full word added/deleted/substituted, bc = below clause, ca = clause and above, Con =content, Org = organisation, Lex = lexical retrieval, Syn = syntactic encoding, Don't Re.= don't remember, Tot = total. *Due to rounding some totals do not add up to 100.

Appendix D, part 2
Reasons for Revision: Summary of Stimulated Recall Comments for Complex Essay

Lea./ Loc.	Complex essay												
	Planning				Translation								
	Con.	Org.	Tot.		Lex.	Syn.	Coh.	Tot.		Don't Re.		Total overall *	
	N	N	N	%	N	N	N	N	%	N	%	N	%
Lea.2	26	0	28	25%	24	14	0	80	70%	6	5%	114	100%
bw	2	0	2	2%	8	2	0	22	19%	2	2%	26	23%
W	4	0	4	4%	8	8	0	32	28%	2	2%	38	34%
bc	16	1	18	16%	6	42	0	22	19%	2	2%	42	37%
ca	4	0	4	4%	2	0	0	4	4%	0	0%	8	8%
Lea.4	14	0	14	33%	10	2	0	28	67%	0	0%	42	100%
bw	0	0	0	0%	0	0	0	0	0%	0	0%	0	0%
W	2	0	2	5%	6	0	0	14	33%	0	0%	16	38%
bc	4	0	4	10%	4	2	0	14	33%	0	0%	18	43%
ca	8	0	8	19%	0	0	0	0	0%	0	0%	8	19%
Lea.6	58	0	62	28%	24	24	22	158	71%	1	1%	222	100%
bw	4	0	4	2%	0	6	0	14	6%	0	0%	18	8%
w	24	0	24	11%	18	12	0	62	28%	0	0%	86	39%
bc	12	0	12	5%	4	4	22	68	31%	1	1%	82	37%
ca	18	4	22	10%	2	2	0	14	6%	0	0%	36	16%
Lea.8	18	0	22	19%	18	14	0	86	75%	6	5%	114	100%
bw	2	0	2	2%	4	2	0	12	11%	2	2%	16	15%
W	2	0	2	2%	8	6	0	36	32%	2	2%	40	36%
bc	6	0	6	5%	6	4	0	22	19%	2	2%	30	26%
ca	8	4	12	11%	0	2	0	16	14%	0	0%	28	25%
Tot	116	0	126	26%	76	54	22	352	72%	14	3%	492	100%
bw	8	0	8	2%	12	10	0	48	10%	4	1%	60	13%
W	32	0	32	7%	40	26	0	144	29%	4	1%	180	37%
bc	38	2	40	8%	20	14	22	126	26%	6	1%	172	35%
ca	38	8	46	9%	4	4	0	34	7%	0	0%	80	16%

Lea = learner, Loc = pause location, bw = below word, w = one full word added/deleted/substituted, bc = below clause, ca = clause and above, Con =content, Org = organisation, Lex = lexical retrieval, Syn = syntactic encoding, Don't Re. = don't remember, Tot = total. *Due to rounding some totals do not add up to 100.

ISSN 1110-2721

Occasional Papers
Vol. 64January (2018)