Effects of Post-Task Anticipation
during Online Collaborative Writing in L2

Vararin Charoenchaikorn

Chulalongkorn University Language Institute, Thailand
cvararin@hotmail.com

Article information	
Abstract	This study was motivated by Skehan's (1996, 1998) hypothesis about the potential of post-task anticipation to direct learners' attention towards form and previous online L2 collaborative writing research that found prioritization of meaning in this context. It investigated the effects of post-task anticipation on text length, accuracy and complexity, and learners' perceived effects of anticipation, in online collaborative writing. Ninety-eight Thai EFL learners of English on an online university course were divided into control and experimental groups. The control group performed the main online collaborative writing task without foreknowledge of a post-task, while the experimental group anticipated a peer language evaluation post-task. Texts written during the main task and questionnaire responses were assessed. Overall, statistical analyses revealed that post-task anticipation did not lead to increased attention to form. Independent-sample t-tests did not detect significant effects of post-task anticipation on text length or accuracy, and significantly more complex language was observed in the control group. Possible reasons for these findings are discussed. Despite the lack of positive outcomes for performance, questionnaire responses indicated that post-task anticipation might promote engagement during the main task.
Keywords	post-task, post-task anticipation, online collaborative writing, attention to form
APA citation:	Charoenchaikorn, V. (2022). Effects of post-task anticipation during online collaborative writing in L2. *PASAA, 64*, 68–98.

Introduction

With technological advancement, the utilization of online pedagogic writing tasks has become increasingly common. Many Web 2.0 writing tools, such as Wikis and Google Docs, have received much attention from language teachers and researchers as they facilitate online collaboration which may nurture writing development. To date, L2 researchers have examined online collaborative writing (CW) in multiple dimensions, ranging from interaction and writing processes during collaboration, writing quality and development, to students' perceptions (for a review, see Li, 2018). Many researchers have shown that L2 learners may prioritize meaning over linguistic form during online CW, which could hinder learning (Abrams, 2016; Elola & Oskoz, 2010; Godwin-Jones, 2018; Kessler, 2009; Mak & Coniam, 2008); hence, it may be necessary to direct learners' attention towards form in this context. The present study was motivated by this assumption and investigated whether such attention could be engineered by having learners anticipate a post-task. It has been hypothesized that, with foreknowledge of a form-focused post-task, learners potentially allocate more attention to form in the main task (Skehan, 1996, 1998). Although this hypothesis focuses on post-task anticipation (PA) during oral production tasks, this research operationalized an interpretation of this hypothesis and explored how PA influenced accuracy, complexity, and text length in online CW, and the perceived effects of this anticipation.

Literature Review

Online Collaborative Writing in L2

CW is defined as an activity in which all participants negotiate and work together in the entire writing process to produce a single text with shared ownership (Storch, 2013, 2019). It has been widely utilized in the L2 classroom and, with technological advancement, increasingly implemented in online contexts. According to Storch (2019), increased attention to CW results from the realization that it resembles team writing in the workplace (Mirel & Spilka, 2002) and the advent of Web 2.0 tools, such as Google Docs and Wikis, which facilitate the co-creation and sharing of texts and have brought about novel literacy practices. The main advantage of Web 2.0 applications is they allow learners to jointly compose texts without restrictions on time or space. Through these applications, learners can interact with their co-authors online during the whole writing process,

including the co-construction of text, revision, and editing (Li, 2018). For instance, in Google Docs, they can use the comment feature to make plans and revise together, discuss language use, and provide or receive feedback.

Essentially, the benefit of online CW can be viewed in terms of both sociocultural theory (Vygotsky, 1978, 1981) and the interaction hypothesis (Long, 1983, 1996), both of which assume the importance of interaction in promoting learning through knowledge co-construction and negotiation of meaning. CW can also afford opportunities for practice, receiving feedback and deliberating language output that may lead to language gains and consolidation of L2 knowledge (Storch, 2011; Wigglesworth & Storch, 2012). To date, studies have documented some advantages of online L2 CW. They have revealed that writing collaboratively online can result in significantly greater writing improvement (Bikowski & Vithanage, 2016) and better text quality in terms of content, organization, and coherence (Abrams, 2019; Strobl, 2014) when compared with online non-CW. In addition, online CW may lead to significantly greater pre- to post-writing test gains compared with offline CW (Wang, 2015). Many L2 researchers exploring learners' perceptions of online L2 CW have found that most learners reported positive experiences and benefits as regards, for example, writing development, learning opportunities, motivation, audience awareness, and content generation (Bikowski & Vithanage, 2016; Ducate, Anderson, & Moreno, 2011; Kost, 2011; Lee, 2010; Li & Zhu, 2013; Mak & Coniam, 2008; Wang, 2015).

However, despite these positive findings, many studies have found that learners tend to prioritize meaning over form or accuracy in online L2 CW, which may be due to its interactive nature. Studies utilizing both Google Docs (Abrams, 2016; Godwin-Jones, 2018; Kessler, Bikowski, & Boggs, 2012) and Wikis (Elola & Oskoz, 2010; Kessler, 2009; Mak & Coniam, 2008) have revealed that learners make more meaning- than form-related revisions/contributions, indicating more attention to meaning. They also show that language accuracy and complexity may not relate to CW patterns in an online context (Abrams, 2019; Elola & Oskoz, 2010; Strobl, 2014). For example, Abrams (2019) found that accuracy (errors/word), lexical complexity (the measure of textual lexical diversity, TLD) and syntactic complexity (C-unit length and clauses/C-unit) were not linked with collaborative patterns during composition in Google Docs. Elola and Oskoz (2010) also observed

non-significant differences between collaborative and individual writing in Wikis in accuracy, as measured by the percentage of error-free T-units, and complexity, as assessed by the percentage of words/T-unit and subordinate clauses/T-unit. Taken together, these findings suggest that CW in an online context might not be ideal for promoting attention to form or accuracy (Storch, 2011).

Task and Post-Task Anticipation

Educators have acknowledged the importance of tasks for L2 learning since the development of task-based language teaching (TBLT), a learner-based and communicative teaching method, in the 1980s (Van den Branden, Bygate, & Norris, 2009). Task utilization affords the opportunity for interaction that can be facilitative of L2 learning. According to Long (1991, 2015), the incidental shift in attention from meaning to linguistic form occurring during tasks, e.g., following communication breakdowns or corrective feedback, is essential for language development and enhances form-meaning mappings. Communicative tasks implemented in class may also prepare learners for the demands of tasks they will encounter in real life (Long, 2015).

Notwithstanding the possible benefits of tasks, Skehan (1996, 1998) hypothesized, based on VanPatten's (1994) input processing theory, that communicative tasks may lead to meaning prioritization. When left to perform tasks on their own, learners tend to focus on meaning rather than form and accomplish tasks by resorting to communication strategies rather than complex or accurate use of language. The assumptions underlying this hypothesis are that learners have limited attention capacity and meaning and form compete for their attention. Hence, Skehan suggests that attention to form must be engineered to maximize learning opportunities during L2 tasks and foster balanced development of all performance aspects—complexity, accuracy, and fluency.

To date, researchers have made several proposals concerning how attention to form can be achieved when tasks are utilized. Regarding task implementation, Skehan (1996, 1998, 2013) proposed that the utilization of a post-task could subtly maneuver learners into paying attention to form during the main task, potentially leading to increased noticing. He posits that, when learners are aware that they need to complete a post-task in which form is important, e.g., a task entailing a

language test, public performance or main task performance analysis, they may attend more to form during the main task, as that can be beneficial for the post-task stage (Skehan, 1996, 1998). The primary effect of this awareness is assumed to be on accuracy, rather than complexity (Skehan, 1998). According to Foster and Skehan (2013) and Skehan, Xiaoyue, Li, and Wang (2012), post-task anticipation (PA) may create a suitable condition for nurturing accuracy as it encourages learners to reallocate their cognitive resources to attend to form, possibly resulting in greater performance monitoring and more intensive access to knowledge in long-term memory. Essentially, this explanation assumes that PA has a *focusing* effect on performance (see Skehan, 2009). That is, it motivates learners to focus on accuracy during language formulation.

To date, a limited number of empirical studies have been conducted to test Skehan's PA hypothesis (Foster & Skehan, 2013; Li, 2014; Newton & Nguyen, 2019; Skehan & Foster, 1997). In their early work, Skehan and Foster (1997) explored PA effects on pre-intermediate learners' performance during three English-speaking tasks: a narrative, a decision-making and a personal information exchange task. Prior to these main tasks, learners in the experimental group were informed that two pairs would be chosen to perform the main tasks again in front of the class in the post-task stage. Results showed that the complexity of the control (-anticipation) and experimental (+ anticipation) groups was comparable for all tasks. However, the experimental group was significantly less fluent during the narrative task and more accurate in the decision-making task. A more recent study by Foster and Skehan (2013) confirms a significant effect on accuracy. It compared the main task performance of pre-intermediate students with and without anticipation of a post-task transcription activity. In the post-task stage, all learners in the experimental group were required to individually transcribe recordings of discussions they had during the main tasks, a narrative task and a decision-making task, and identify mistakes. Results revealed non-significant effects on fluency, but PA led to significantly enhanced complexity in the decision-making task and increased accuracy in both main tasks. The effects on accuracy were more pronounced in the decision-making task than in the narrative task. Taken together, the findings of these two studies suggest that PA may have more impact on accuracy than other performance aspects during oral tasks. Accuracy is the only performance aspect consistently found to be influenced by PA. Also, Foster and

Skehan (2013) detected a more noticeable effect on accuracy than other aspects. In another study utilizing post-task transcription tasks, Li (2014) investigated oral performance during narrative and decision-making tasks of intermediate learners and found no effects of PA on complexity, accuracy or lexical performance in the main task of the first task cycle. However, learners who had completed the post-task transcription task generated speech with significantly increased accuracy and complexity during the main task of the second cycle compared to those who had not completed the post-task. This might indicate the effects of post-task involvement rather than PA.

Corroborating the findings of Skehan and Foster (Foster & Skehan, 2013; Skehan & Foster, 1997), Newton and Nguyen (2019), who investigated language-related episodes (LREs), also found evidence of attention to form when learners had foreknowledge of a post-task. They examined LREs during the main tasks, operationalized as rehearsals for subsequent public performances, performed by 24 pairs of Vietnamese high-school learners of English. Two speaking tasks were utilized: a problem-solving task and a debating task. LREs were found to occur frequently during the main tasks, indicating attention to form, with significantly more LREs observed in rehearsals of the problem-solving task than in the debate task.

Extending the scope of investigation to an online context, Charoenchaikorn (2019) explored the effects of anticipation of two types of language correction post-tasks, individual and collaborative, on accuracy, revision and speed fluency during information-gap tasks. Unlike other studies, this research was done in the context of text-based synchronous computer-mediated communication. Not consistent with some research examining PA effects on oral performance, this research did not detect evidence of attention to form; no significant effects of PA or types of PA were observed in any performance measure. This suggests that the effectiveness of PA may depend on the task context.

Despite some evidence of the potential of PA to nurture attention to form, more research in this area is warranted to understand this influence. Only a handful of PA studies have been conducted to date. In addition, they explored only a few types of post-tasks – public performance, post-task transcription and

correction – and most investigated PA effects on oral performance. It is possible that types of main tasks and post-tasks and other task contexts affect this potential. For instance, PA effects during the main task performed in oral and written modes may differ; in terms of post-tasks, a post-task requiring learners to merely transcribe their friend's speech may not have as much effect as that entailing peer language evaluation and correction or publicizing all learners' performances, as in the latter cases learners anticipate that their mistakes will be scrutinized by their friends or that the whole class, as opposed to only one friend, will see their performance. In light of this, it is useful to explore PA further.

Research Questions

This study examines the potential of PA in L2 teaching and learning by exploring its effects on the performance of Thai EFL learners of English in a university context guided by the following research questions:

1. What is the effect of anticipating a peer language evaluation post-task on text length in online L2 CW?

2. What is the effect of anticipating a peer language evaluation post-task on accuracy in online L2 CW?

3. What is the effect of anticipating a peer language evaluation post-task on complexity in online L2 CW?

4. Does anticipating a peer language evaluation post-task motivate learners to put more effort into their work during online L2 CW?

Although Skehan's (1996, 1998) original hypothesis concerns the PA influence during oral production tasks, this research applies an interpretation of this hypothesis and investigates the influence on online CW. Possibly, knowledge of certain post-tasks also encourages learners to redirect their attention to form during written tasks. As discussed previously, L2 learners may attend more to meaning than linguistic form during online CW. PA could be useful to promote attention to form in this context.

In this study, the main task involved writing an opinion paragraph. The post-task was operationalized as a language evaluation post-task in which students' writing is evaluated by peers. This post-task is novel in that it has not been explored in previous PA research. As discussed in the previous section, while post-

task types may affect the effectiveness of post-task anticipation, PA studies have investigated only a few types of post-tasks. According to Skehan (1998) who made an assumption about the benefits of language analysis post-tasks, post-tasks that entail having the main task performance analysed, either by those completing the task themselves or their peers, might have an impact on main task performance; awareness of this analysis may alter how learners allocate their cognitive resources during the main task. Moreover, peer evaluation is supported by several theories, e.g., interactionist, collaborative learning, and sociocultural, and empirical evidence has demonstrated its effectiveness for L2 writing (for a review, see Yu & Lee, 2016).

Method

Design

Data were collected from 98 first-year Thai university students in four online English classrooms conducted via Zoom (Zoom Video Communications, 2020) video-conferencing software. All were enrolled on the same online university English course focusing on listening, reading, speaking, and writing skills required to be taken by every first-year student. Two classrooms ($n = 50$) were randomly assigned to the control condition (- PA), and the other two ($n = 48$) to the experimental condition (+ PA). Data were gathered from two classes of business students and two classes of science students to ensure comparability; each of these respective classes was randomly assigned to the experimental group. While the control group only performed the main CW task, participants in the experimental group also completed a language evaluation post-task. They were told, prior to the main task, about the post-task. Moreover, online questionnaires were administered to obtain information about the participants' background information, perception and experience during data collection. The texts they wrote and their questionnaire responses were analyzed to answer the research questions. Following international ethical standards, the research was approved by the ethics review committee of the university prior to data collection.

Participants

Initially, a total of 114 first-year students participated in the data collection session. However, 16 were excluded due to technical problems, e.g., Internet connection issues. Of the remaining 98 participants, 52 and 46 were studying in

the fields of business and science, respectively. Twenty-eight business and 22 science students were in the control groups and 24 students in each field were in the experimental group. Participants in each classroom had studied in the same online course for many weeks but never met in person. Twenty-nine (29.6%) were male, and 69 (70.4%) were female. All participants' L1 was Thai and their ages ranged from 17 to 20, with a mean of 18.4 (SD = 0.64). On average, they had learned English for 12.47 years (SD = 3.34, range = 3–18). Based on teacher assessment, they were at intermediate levels, B1–B2 in the Common European Framework of Reference for Languages (Council of Europe, 2001). They obtained a mean score of 61.01 out of 100 (SD = 13.86) for an English test administered nationwide upon high school completion, and 8.1 out of 12 (SD = 1.93) for an individual paragraph-writing assignment performed as part of the course they enrolled on. Most students (n = 54) reported using a computer to attend class on the data collection day. The rest used either a tablet (n = 34) or smartphone (n = 10). All participants reported spending at least one hour on their selected device per week, with 95 (96.94%) reporting spending at least three hours per week.

T-tests confirmed that the control and experimental groups were comparable for their number of years spent learning English, t(96) = -0.69, p = .491, BCa 95% CI [-1.79, 0.86], English test scores, t(96) = 0.83, p = .409, BCa 95% CI [-3.25, 7.72], and writing scores, t(96) = 1.68, p = .096, BCa 95% CI [-0.11, 1.41]. The Mann-Whitney U test performed on the background questionnaire data regarding the length of time participants spent on their device each week also showed comparability across groups, U = 1311.00, p = .287, z = 1.07, r = .11.

Main Task

To date, many definitions of tasks have been proposed by SLA researchers, e.g., real-world activities (Long, 1985), activities requiring the use of the target language to achieve a communicative goal (Willis, 1996) or involving information processing (Prabhu, 1987), or holistic activities entailing the use of multiple linguistic aspects to promote language learning (Samuda & Bygate, 2008). In this study, the main task was a writing task in which participants wrote a paragraph to express their opinion about the education system in Thailand. This type of writing is common in academic writing at the university level and the task involves the processing of linguistic information to communicate and learn the target language,

which is in accordance with many proposed task definitions. The writing topic was relevant to a unit they were covering in their English classes at the time, as this activity was done as part of the course. The prompt of the main task was "What changes would you like to see in the education system of your country? Why?" Participants were given 50 minutes to write a paragraph of at least 150 words in groups of two or three students to answer these questions. They were told not to use online resources and to discuss the questions with their group in a Zoom online meeting room to produce a text and write it as a Google Docs file. Google Docs (Google, 2020) is an online word processor that can be used for collaboration. On this platform, students can write on the same document at the same time, see peer contributions in real-time, and add comments to the text. Each group of students was provided with a link to a blank Google Docs file prepared prior to the data collection session.

Post-Task

The post-task was operationalized as a language evaluation task. During this task, participants in the experimental condition discussed in Zoom, with the same peer(s) as in the main task, to assess the text produced by another group. Each group was given an online evaluation form to complete in Microsoft Forms (Microsoft, 2020). It contained multiple items that required them to choose whether they 'totally agree', 'agree', 'neither agree nor disagree', 'disagree' or 'totally disagree' that their friends' text had good content, was well-organized and showed accurate and complex use of language, in terms of both lexis and sentence structure. In addition, each group was asked to rate the work from 0–5 and indicate both good points and any improvements that could be made. Students received peer feedback, i.e. the evaluation form completed by another group, following the post-task.

Questionnaires

Two online questionnaires, background and exit, written in the participants' L1 were administered through Microsoft Forms (Microsoft, 2020). The background questionnaire asked for participants' background information, such as age, sex, education, first language and familiarity with the device they were using. The exit questionnaire contained questions about the perceived difficulty and stressfulness of the main task, with two five-point Likert items enquiring whether participants

agreed that the task was challenging and stressful. Moreover, it probed the relationship among group members and whether they had enough time for the main task performance. The exit questionnaire for the experimental group additionally asked about the perceived usefulness and effects of the post-task. Participants were asked to choose whether they 'totally agree', 'agree', 'neither agree nor disagree', 'disagree' or 'totally disagree' that the post-task was useful for learning English, PA motivated them to put more effort into their work, and PA made them worried during the main task. They were also required to give reasons for their answers to these three questions.

Data Collection Procedures

The data collection took place once in each of the four online classrooms during normal class time. The writing task was part of online classroom activities. Participants in all classes had become familiar with online learning via Zoom. Prior to data collection, the course had been conducted using the platform for more than three weeks.

The data collection began with all participants in each class being in the same conference room (i.e. the main room) to receive instructions about the main task; both the control group and the experimental group received the same instructions. At this point, the experimental group was also told that they would perform a language evaluation post-task afterwards, the text they wrote during the main task would be assessed by another group based on the criteria shown on the online evaluation form presented to them, and they would receive peer feedback, i.e. the form completed by their friends, after the post-task. However, they did not know beforehand which group would evaluate their work. Then, participants were sent to separate online breakout rooms to perform the main task with their group. Zoom breakout rooms are private conference rooms useful for group work during an online class. Participants can be sent to different breakout rooms to hold small private group meetings and, once the work is done, they can return to the main room. In this research, two or three participants were allocated to each breakout room using the programme function that allows automatic generation of breakout rooms. Group allocation was, thus, not predetermined. For the purposes of a possible future study, some breakout rooms were recorded during the main task, based on participants' consent, using the recording function of Zoom. Upon

completion of the main task, the control group returned to the main conference room and was sent online questionnaires to complete individually. The experimental group returned to receive instructions for the post-task. Participants carried out the post-task in the same breakout room with the same group as the main task. Each group was randomly provided with a link to the document written by another group to evaluate. Once finished, the experimental group proceeded to complete the questionnaires individually and received peer feedback.

Data Analysis
1) Written Text

Participants' written output was examined in terms of length, accuracy and complexity. Text length was operationalized as the number of words in the final text. Accuracy measures were grammatical, lexical and total errors/100 words. This ratio measure (errors/100 words) has been utilized in various SLA studies (e.g., Barkaoui & Knouzi, 2018; Lee, Joo, Moon, & Hong, 2007; Mehnert, 1998; Révész, Michel, & Lee, 2017; Suzuki & Kormos, 2020). In a comparative study, Inoue (2016) found that it aligned best with raters' judgements of accuracy when compared to other oft-used measures of accuracy, errors/AS-unit and percentage of error-free clauses. In this study, errors were coded using Atlas.ti 9 (Scientific Software Development GmbH, 2020) data analysis software. Grammatical errors concerned errors relating to grammar, such as verb form, sentence structure, subject-verb agreement, articles and the grammatical number of nouns. Lexical errors involved erroneous lexical choices. Spelling and punctuation mistakes were not counted as errors. To ensure coding reliability, the second coder, who is a native speaker of English, recoded approximately 20% of the data ($n = 10$), randomly selected across two groups. Cohen's kappa for error identification and categorization was 0.94 and 0.89, respectively, indicating strong inter-coder agreement.

Lexical and syntactic complexity were explored using the measures summarized in Tables 1 and 2. Before the texts were subjected to automated language complexity analysis software, word spacing, spelling and punctuation mistakes were corrected to ensure accurate analyses. Lexical complexity was examined using various measures utilised in previous SLA research. According to Michel (2017), lexical complexity has usually been conceptualized in terms of

diversity, sophistication and density. In this research, diversity was gauged by the measure of textual and lexical diversity (MTLD), an index representing the mean length of sequential word strings that maintain a given type/token ratio (McCarthy & Jarvis, 2010). This value is not influenced by text length and is, thus, a more valid measure than those impacted by length, e.g., the basic type/token ratio (McCarthy & Jarvis, 2010). A higher value of MTLD indicates an increased proportion of different words, or types, in the text. Lexical sophistication was gauged by two measures: log frequency of content words and proportion of academic words. Log frequency of content words, a measure reflecting lexical rarity (Jarvis, 2013), was computed based on the CELEX lexical database (Baayen, Piepenbrock, & Gulikers, 1995). The proportion of academic words (Coxhead, 2000) was adopted because it relates to the type of lexis relevant to the setting of the research. These words are specific to academic contexts and do not include the 2,000 most frequently used words in English. Lexical density was measured by means of the ratio of content words to total words in the text. Apart from diversity, sophistication and density, an additional measure of disparity was included based on Jarvis' (2013) proposal which outlines six components of lexical complexity. Semantic disparity involves the degree of differentiation between types of lexis (Jarvis, 2013). It was assessed using Latent Semantic Analysis (LSA) following Jarvis' (2013) suggestion and the conduct of previous research (e.g., Révész, Kourtali, & Mazgutova, 2017; Révész, Michel, & Lee, 2017). This analysis is based on a statistical model of semantic meaning that reveals conceptual similarity, i.e., semantic overlap, between sentences by comparing each sentence to every other sentence in the text. Low LSA index values indicate increased complexity in terms of disparity. The indices of MTLD and LSA and log frequency of content words were computed using Coh-Metrix 3.0 (Graesser, McNamara, & Kulikowich, 2011; Graesser, McNamara, Louwerse, & Cai, 2004; McNamara, Graesser, McCarthy, & Cai, 2014), a system for text analysis. The proportion of academic words and the ratio of content/ total words were obtained from VocabProfiler (Cobb, 2020), an online vocabulary analysis tool.

Table 1

Lexical Complexity Measures

Diversity
MTLD
Sophistication
Log frequency of content words
Proportion of academic words
Density
Content/ total words
Disparity
LSA index

Following Norris and Ortega's (2009) proposal, syntactic complexity was gauged not only by means of global measures but also those capturing clausal complexity, through subordination and coordination, and phrasal complexity. This allows for the investigation of different dimensions of complexity that may relate to different stages of L2 development (Norris & Ortega, 2009). A global measure adopted was the length of T-units operationalized as words/T-unit, a common syntactic complexity measure in SLA research that has been found to indicate differences in L2 proficiency and development (Bulté & Housen, 2014; Ortega, 2003; Stockwell & Harrington, 2003). A T-unit is defined as the main clause and any subordinate clauses attached to or embedded in it (Hunt, 1965). Another general measure is the syntactic similarity index – an index assessing the similarity of syntactic structures in a text by comparing all possible combinations of sentence pairs. High syntactic similarity index values show increased similarity and, hence, reduced syntactic complexity (Crossley, Greenfield, & McNamara, 2008). Coordinate and subordinate complexity were evaluated through T-units/sentence and clauses/T-unit, respectively. Finally, phrasal complexity was measured by words/clause, a measure recommended by Norris and Ortega (2009) for assessing phrasal complexity. However, because clause length can be affected by adding more adjuncts and other expansions at the level of clauses, it may not be a "pure measure" of phrasal complexity (Bulté & Housen, 2012, p. 38). Hence, a complementary measure of phrasal complexity, complex nominals/T-unit, was utilized. These nominals include 1) noun clauses, 2) nouns with possessive, adjective, appositive, participle, prepositional phrase or relative clause, and 3)

gerunds or infinitives in the subject position (Cooper, 1976). In this research, all indices of syntactic complexity were obtained from a text analysis tool, Synlex (Lu, 2010), except the syntactic similarity index which was computed using Coh-Metrix 3.0.

Table 2

Syntactic Complexity Measures

Global complexity
Words/ T-unit
Syntactic similarity index
Clausal complexity
I. Coordination
T-units/ sentence
II. Subordination
Clauses/ T-unit
Phrasal complexity
Words/ clause
Complex nominals/ T-unit

2) Questionnaire

The responses of the experimental group to the exit questionnaire item relating to the perceived effect of PA were examined. In the question that asked participants to choose whether they 'totally agree', 'agree', 'neither agree nor disagree', 'disagree' or 'totally disagree' that PA motivated them to put more effort into their work during the main task, a frequency score was calculated for each of the five possible answers. Then, the reasons participants wrote to justify their answer to this question were coded manually in Atlas.ti for themes, and the themes were grouped based on whether they indicated the effect of PA.

3) Statistical Analyses

Prior to running statistical analyses, the data were checked to ensure a normal distribution and significant outliers were removed. One extreme value was excluded from the analysis of text length, and two from that of a measure of phrasal complexity, words/clause. Pearson's correlation analyses to test the relationship between different measures adopted were also run. The results

confirmed that these measures were relatively independent of each other; the correlations ranged from weak to moderately strong (r = .01–.72), with an exception being a very strong and significant correlation between the rates of grammatical and total errors ($p < .001$, $r = .89$)[1]. Moreover, because two out of the 24 and 23 groups were formed by three students in the control and experimental conditions, respectively, Mann-Whitney U tests were run to ensure comparability across group size. The results showed that groups of two and three students were comparable for all performance measures (length, accuracy and complexity), $p >$ .05, $r = |.02|–|.28|$.

To answer research questions 1–3, a series of independent-sample t-tests were performed in SPSS Version 22 to test for differences in text length, accuracy and complexity across groups (control vs experimental). Plonsky and Oswald's (2014) benchmarks for the interpretation of d in L2 research were adopted to evaluate effect size. Absolute values of 0.40, 0.70 and 1.00 were the thresholds for small, medium and large effect sizes for between-group comparisons, respectively. For the perceived effects of PA, descriptive statistics for the questionnaire responses were calculated to answer research question 4.

Results

Effect on Text Length

On average, the control group wrote 161.04 words ($n = 24$, $SD = 40.79$) and the experimental group 180.09 words ($n = 22$, $SD = 34.31$) within the time given, 50 minutes. An independent t-test found this difference to be non-significant, $t(44)$ = -1.71, $p = .095$, BCa 95% CI [-40.18, 2.83], and the effect size was in the small to medium range, $d = 0.47$.

Effect on Accuracy

Descriptive statistics for accuracy measures presented by group are provided in Table 3. On average, the experimental group produced grammatical, lexical and total errors more frequently than the control group. However, independent t-tests (Table 4) revealed no significant difference across groups for

---

[1] Despite a very strong and significant correlation, a decision was made to report both rates because total error rate shows an overall picture of accuracy.

all these measures. The effect size was small for frequency of grammatical and total errors, but in the small to medium range for frequency of lexical errors.

Table 3

Descriptive Statistics for Accuracy Measures

(errors/ 100 words)	Control			Experimental		
	n	M	SD	n	M	SD
Grammar	24	8.42	4.59	23	8.93	4.48
Lexis	24	6.41	2.22	23	7.38	2.32
Total	24	14.84	4.89	23	16.31	5.25

Table 4

Effect of Group on Accuracy

(errors/ 100 words)	p	d	t	df	BCa 95% CI	
					Lower	Upper
Grammar	.702	0.11	-0.39	45	-2.76	1.65
Lexis	.152	0.43	-1.46	45	-2.32	0.27
Total	.324	0.30	-1.00	45	-4.11	0.10

Effect on Complexity

Table 5 provides descriptive statistics for complexity measures. A higher proportion of academic words and ratio of content/total words and lower log frequency and LSA index value in the experimental group suggest an increase in lexical density, sophistication and disparity in this group compared to the control group. On the other hand, the control group produced more lexically diverse text as indicated by the higher MTLD value. However, for all these measures of lexical complexity, independent t-tests found a non-significant result and a rather small effect size, $d \leq |0.41|$, as shown in Table 6.

Regarding syntactic complexity, all measures revealed a higher degree of complexity in the control group compared to the experimental group (Table 5). That is, on average, the control group produced more words/T-unit, T-units/sentence, clauses/T-unit, words/clause and complex nominals/T-unit, and generated text with reduced syntactic similarity. The outcomes of independent t-tests in Table 6

show the significant effect of group on all measures of global complexity (words/T-unit and syntactic similarity index), the measure of subordination (clauses/T-unit) and the measure of phrasal complexity (complex nominals/T-unit). For these measures, the effect size was roughly in the medium range.

Table 5

Descriptive Statistics for Complexity Measures

	Control			**Experimental**		
	n	M	SD	n	M	SD
Lexical complexity						
MTLD	24	65.83	15.01	23	65.80	16.67
Log frequency of content words	24	2.50	0.14	23	2.48	0.12
Proportion of academic words	24	4.07	2.38	23	4.41	2.37
Content/ total words	24	0.51	0.04	23	0.52	0.05
LSA index	24	0.18	0.05	23	0.16	0.05
Syntactic complexity						
Words/ T-unit	24	18.19	5.12	23	15.23	2.73
Syntactic similarity index	24	0.09	0.03	23	0.11	0.03
T-units/ sentence	24	1.12	0.14	23	1.08	0.09
Clauses/ T-unit	24	1.88	0.52	23	1.58	0.34
Words/ clause	23	9.56	2.11	22	9.55	1.64
Complex nominals/ T-unit	24	2.18	0.60	23	1.85	0.48

Table 6

Effect of Group on Complexity

	p	d	t	df	BCa 95% CI Lower	BCa 95% CI Upper
Lexical complexity						
MTLD	.995	-0.002	0.01	45	-9.53	9.63
Log frequency of content words	.562	-0.14	0.58	45	-0.06	0.10
Proportion of academic words	.624	0.14	-0.49	45	-1.70	1.06
Content/ total words	.256	0.38	-1.15	45	-0.04	0.01
LSA index	.176	-0.41	1.37	45	-0.01	0.05
Syntactic complexity						
Words/ T-unit	.017	-0.58	2.49	35.40	0.61	5.45
Syntactic similarity index	.039	0.58	-2.12	45	-0.039	-0.002
T-units/ sentence	.222	-0.30	1.24	39.36	-0.03	0.12
Clauses/ T-unit	.023	-0.58	2.37	40.04	0.06	0.57
Words/ clause	.987	-0.005	0.02	43	-1.01	0.99
Complex nominals/ T-unit	.047	-0.54	2.05	45	-0.01	0.66

Perceived Motivating Effect

This effect was evaluated through participants' responses to the question which asked whether they 'totally agree', 'agree', 'neither agree nor disagree', 'disagree' or 'totally disagree' that PA motivated them to put more effort into their work during the main task. Of 48 participants in the experimental condition, only two answered 'totally disagree' and none 'disagree'. Twelve (25%) chose 'neither agree nor disagree'. Most participants (70.83%) seemed to perceive the effect of PA, with 20 reporting 'agree' and 14 'totally agree'.

The reasons participants provided to justify their answers concerned several themes. Participants who did not perceive the motivating effect of PA wrote that they were not worried about the post-task ($n = 4$), or that they always try to produce the best possible text regardless of whether there is a post-task ($n = 5$). For example, one participant reported "I just wrote normally. I didn't feel I had to be concerned about showing my work to other people", and another "I think we should always write as accurately as possible. It doesn't matter whether our work will be assessed by the teacher or other friends".

With respect to the reasons supporting the effectiveness of PA, some responses related to text organization ($n = 1$), lexical variety ($n = 3$) and formality ($n = 4$). In particular, participants reported that PA encouraged them to pay more attention to ideas organization and use more varied lexical items and formal language. For instance, they wrote "I tried to make the content more well-organized"; "I attempted to use a variety of lexical items because I wanted my work to be as good as possible"; and "For vocabulary, I usually use common words but it [PA] encouraged me to include more formal ones". In addition, six participants wrote that PA motivated them to make their language, both structure and lexis, clear and easy to understand so that their friends could understand it during the post-task. For example, one wrote "I changed the words so that they are easier to understand" and another "I tried to make my text not too complicated. If our text is confusing, other people may misunderstand our ideas". Moreover, 11 students reported effects on language accuracy. That is, knowledge of the post-task motivated them to write more accurately by using correct spelling, sentence structures and lexis. One such response was "I tried my best to write with correct grammatical structures because I wanted to make the fewest mistakes possible. If the text is not well-written, people will think that I can't write. This won't affect only me but my partner too". Finally, eleven participants reported overall increased attention to language. With PA, these learners became more focused on their language. For example, two participants wrote "I paid more attention to the words and sentences I wrote" and "I thought more about language before typing".

Discussion

This research found significant effects of PA on neither text length nor accuracy. However, the language of the control group was significantly more complex than that of the experimental group, as assessed by multiple measures. The non-significant finding pertaining to text length indicates that learners may not write more in a CW task when anticipating a language evaluation post-task. To a certain extent, this is in line with previous studies (Charoenchaikorn, 2019; Foster & Skehan, 2013; Skehan & Foster, 1997) that found no or an unclear effect of PA on speaking fluency and speed fluency during text chat.

Regarding accuracy, although approximately a quarter of the participants in the experimental group reported that PA motivated them to write more accurately

in the questionnaire, the non-significant effects observed suggest that PA might not be effective in promoting attention to form and accuracy in the context of this study. This finding corresponds with that of Li (2014), who found no significant effects of the anticipation of post-task transcription on accuracy in oral performance in the baseline data. Because the main task was timed, it is possible that this time limit restricted participants from sufficiently and effectively improving their accuracy. Although all groups completed their writing within the given time, the collaborative nature of the main task that required them to discuss and agree on what they would write may have necessitated more time. Indeed, in the questionnaire, approximately a fifth ($n = 20$) of participants reported that they would have preferred to have more time, ranging between 10–15 minutes, to revise or polish their writing or discuss more with their friends.

When comparing the findings of this study to those of previous research, it is rather surprising that some previous studies (Foster & Skehan, 2013; Skehan & Foster, 1997) demonstrate PA effects on accuracy during speaking tasks. The increased time pressure for language production during speaking should lead to a lower PA effect on accuracy during speaking than writing. Thus, other explanations, apart from the time limit, are also plausible for non-significant effects on accuracy, and these might cause discrepancies across studies. First, PA may impact accuracy more strongly in face-to-face oral performance tasks than in online written modes; therefore, no effects of PA were detected on accuracy in this study or on accuracy and revision during text chat (Charoenchaikorn, 2019). It could be that the sense of anonymity and impersonality in an online environment makes learners feel freer to make mistakes during the main task, resulting in a lack of influence of PA on accuracy. In the case of the present research, learners in online classes did not see each other in person, so the influence of the post-task might be remote. This idea is linked to the hypothesis of Foster and Skehan (2013, p. 254) who suggested that the influence of a post-task may be weakened when its "threat" is remote, for example, when only a few students are selected to complete a post-task instead of the whole classroom. In this research, the sense of anonymity and impersonality could have increased such remoteness and reduced post-task effects. A second explanation involves the nature of CW. A common finding of L2 research on CW is that learners tend to prioritize meaning over form (Abrams, 2016; Elola & Oskoz, 2010; Godwin-Jones, 2018; Kessler, 2009;

Kessler et al., 2012; Mak & Coniam, 2008). Perhaps the anticipation of a peer evaluation task is not enough to direct their attention towards form in this context. Third, the type of post-task utilized in this study – peer evaluation – might not be as effective in fostering accuracy or attention to form during the main task as those adopted in the previous studies demonstrating PA effects on accuracy (Foster & Skehan, 2013; Skehan & Foster, 1997) – post-task transcription and public performance. Finally, the present study utilized a collaborative task. If each participant had sole responsibility for the language produced during the main task, individual differences, e.g., in terms of fear of negative evaluation, might have influenced the effect of PA more, and the results could have been different.

With regard to complexity, the results obtained from several syntactic complexity measures show significantly reduced complexity in the experimental group than the control group. These findings are surprising and not in line with Skehan's (1998) hypothesis about the potential of PA to promote attention to form, or the findings of Skehan and Foster (Foster & Skehan, 2013; Skehan & Foster, 1997) and Li (2014) that show an unclear or lack of PA effect on syntactic complexity in oral performance tasks. In the present research, some participants reported attempting to make their language, both structure and lexis, easy to understand to facilitate peer evaluation post-task; this is consistent with the observation of Newton and Nguyen (2019), who found that learners may be primarily oriented towards preparing for the post-task rather than the objective of the main task they are performing. Since participants were told that their text would be read and evaluated by their friends prior to the main task, this potentially predisposed them to be concerned about comprehensibility while writing. It might be that this influence of the peer evaluation post-task led to significantly reduced syntactic complexity in the experimental group. When language is made easy to understand, complexity could be compromised. Another plausible reason for the differences detected across groups is accuracy. Although there were no clear effects of PA on performance as regards accuracy, about a quarter of the students in the experimental condition reported attending to accuracy when having foreknowledge of the post-task. For these learners, the preoccupation with accuracy may occur at the expense of complexity. Based on the trade-off hypothesis (Skehan, 1998, 2009), processes relating to fluency, accuracy and

complexity compete for attention; when any of these aspects tax cognitive resources that are limited in capacity, performance in other areas will suffer.

Despite the lack of positive effects on performance discussed thus far, the questionnaire responses show that a peer evaluation post-task potentially motivates learners to put more effort into the main task, with 70.83% of the participants in the experimental group perceiving this effect. These participants reported attending to several language aspects due to PA. This result corroborates that of Newton and Nguyen (2019). In their study, learners anticipating public performance tasks engaged in frequent LREs during the main tasks, suggesting that anticipation might have led to learners' greater engagement; teachers also reported that post-tasks encouraged learners to make greater learning endeavours during the main tasks. Taken together, these findings suggest that PA may promote task engagement, regardless of its influence on performance.

Although not explored in the research questions, responses pertaining to the PA effect on worry showed that approximately 70% of the participants did not agree that PA made them worried when performing the main task. Most of these learners reported that they were not worried because they believed the post-task offered a learning opportunity (e.g., "Both positive and negative feedback given during the post-task can be used to improve my work in the future"). Related to this, a follow-up Mann-Whitney U analysis shows that there was comparability in the scores of the control and experimental groups for perceived stressfulness of the main task, $U = 1134.50$, $p = .622$, $z = -0.49$, $r = -.05$, suggesting that both groups were as stressed during this task. These seem to indicate that, while PA might push participants to make more effort, the post-task utilized did not cause much worry during the main task. Indeed, the outcome of a follow-up Spearman's correlation analysis shows a weak and non-significant correlation between the scores for the perceived PA effect on effort and worry ($p = .107$, $r_s = .24$). In addition, out of 48 participants in the experimental group, there were as many as 22 (45.83%) reporting that they agreed that the post-task made them put more effort into the task but did not agree that it made them worried. Overall, these findings indicate that learners were possibly motivated during the main task not because they were anxious about post-task evaluation. Instead, many embraced

this as a learning opportunity, which perhaps could lead them to put more effort into their work.

Pedagogical Implications, Limitations, and Future Research

The findings of the present research have two main pedagogical implications. First, when compared to previous research findings, the findings reported in this study show that PA effects might be different across modes, task conditions and post-tasks and multiple factors may impact these effects. For instance, the effect of PA on complexity found could have been caused by the post-task nature. Hence, there are many aspects of tasks and task implementation that should be carefully considered when educators aim to maximise the benefits of PA in their classrooms. Ultimately, these decisions should be informed by sufficient research evidence that has yet to be provided by future studies. The second important implication concerns the findings pertaining to perception. This research provides evidence that a post-task entailing peer language evaluation probably motivates learners to engage more during online CW and students perceive the learning opportunity of this post-task. This suggests that the adoption of this post-task may be useful for promoting engagement. Although the present research was conducted in a university setting and an online written mode, it is possible that this advantage of the anticipation of a language evaluation post-task might be extended to the main task implemented in other settings, such as a face-to-face speaking task.

This research, however, is not without limitations. First, only one type of main task was utilized. Previous research (Foster & Skehan, 2013; Skehan & Foster, 1997) has shown that PA effects may depend on the type of main task. Hence, more than one task type should be adopted in future studies. Second, because the study was conducted in normal classroom settings, learners' proficiency was not ascertained prior to data collection to ensure ecological validity. Researchers might consider measuring proficiency and comparing the influence of PA across proficiency levels. It could be that this influence varies based on proficiency. The effect on accuracy may be stronger in a group with higher language ability as they have more L2 resources available to draw on. Third, this research only evaluates PA effects by assessing textual output. Research that examines interaction during the main task could be useful to gain insights into how

PA impacts attention to form and negotiation during the main CW task. Fourth, as this research explores online L2 CW, the results are not generalizable to other contexts such as individual writing and offline CW. To the best of my knowledge, the effects of PA have not been investigated in these contexts.

Owing to the lack of PA research, more studies in this field are warranted. Future research can be done by replicating previous works or revealing the impact of different types of PA on varying task modes. Regarding post-task types, PA effects are probably influenced by post-task nature. As is evident in this study, learners' orientation towards a peer evaluation post-task during the main task encouraged them to make their language easy to understand, possibly resulting in reduced complexity. Thus, sufficient research on each type of post-task is necessary to understand their impact. Furthermore, PA research comparing its effects in different main task conditions might prove useful. As discussed previously, time limitation during the main task potentially reduces the effects of PA on written performance. Different implementations of the main task might influence PA effects. Moreover, studies could be conducted to explore PA effects during individual writing and how they differ from those during CW. In addition, although speculative, how learners view the learning opportunity of the post-task might influence PA effects, and this could be examined further. Finally, positive findings pertaining to the benefit of PA in promoting motivation should be confirmed in future studies. Other PA research to date has primarily focused on performance. Utilizing self-report data to obtain information about learners' perceived effects may allow researchers to better capture the nuances of the effects of PA.

Conclusion

This research has investigated the potential of PA to enhance L2 performance. It operationalises an interpretation of Skehan's (1996, 1998) hypothesis and extends the scope of investigation of previous PA works by exploring the effects of a peer evaluation post-task during online L2 CW. The study is motivated by the scarcity of PA research and a common finding of online L2 CW studies which shows that learners may primarily focus on meaning in this context. It found evidence that the anticipation of a peer evaluation post-task does not cause worry but does encourage learners to make greater endeavours during the

main online CW task. In addition, this anticipation might not positively impact performance. This lack of positive influence could be due to several reasons, such as the type of post-task, mode and main task implementation. Further studies are necessary to confirm these findings and advance research in this field.

About the Author

Vararin Charoenchaikorn is a language educator and researcher with a PhD in Linguistics. Her research interests include task-based language teaching, second language acquisition, computer-assisted language learning, and second language writing.

References

Abrams, Z. (2016). Exploring collaboratively written L2 texts among first-year learners of German in Google Docs. *Computer Assisted Language Learning*, *29*(8), 1259–1270.

Abrams, Z. (2019). Collaborative writing and text quality in Google Docs. *Language Learning & Technology*, *23*(2), 22–42.

Baayen, R. H., Piepenbrock, R., & Gulikers, L. (1995). *The CELEX lexical database* [CD-ROM]. Linguistic Data Consortium.

Barkaoui, K., & Knouzi, I. (2018). The effects of writing mode and computer ability on L2 test-takers' essay characteristics and scores. *Assessing Writing*, *36*, 19–31.

Bikowski, D., & Vithanage, R. (2016). Effects of web-based collaborative writing on individual L2 writing development. *Language Learning & Technology*, *20*(1), 79–99.

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). John Benjamins.

Bulté, B., & Housen, A. (2014). Conceptualizing and measuring short-term changes in L2 writing complexity. *Journal of Second Language Writing, 26*, 42–65.

Charoenchaikorn, V. (2019). *L2 revision and post-task anticipation during text-based synchronous computer-mediated communication (SCMC) tasks* (Doctoral dissertation). Lancaster University, UK.

Cobb, T. (2020). *Web Vocabprofile*. https://www.lextutor.ca/vp/

Cooper, T. C. (1976). Measuring written syntactic patterns of second language learners of German. *The Journal of Educational Research*, *69*, 176–183.

Council of Europe (2001). *Common European Framework of Reference for Languages: Learning, teaching, assessment*. Cambridge University Press.

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

Crossley, S. A., Greenfield, J., & McNamara, D. S. (2008). Assessing text readability using cognitively based indices. *TESOL Quarterly*, *42*, 475–493.

Ducate, L., Anderson, L., & Moreno, N. (2011). Wading through the world of wikis: An analysis of three wiki projects. *Foreign Language Annals*, *44*(3), 493–522.

Elola, I., & Oskoz, A. (2010). Collaborative writing: Fostering foreign language and writing conventions development. *Language Learning & Technology*, *14*(3), 51–71.

Foster, P., & Skehan, P. (2013). Anticipating a post-task activity: The effects on accuracy, complexity, and fluency of second language performance. *The Canadian Modern Language Review*, *69*(3), 249–273.

Godwin-Jones, R. (2018). Second language writing online: An update. *Language Learning & Technology*, *22*(1), 1–15.

Google. (2020). *Google Docs*. https://www.google.com/docs/

Graesser, A. C., McNamara, D. S., & Kulikowich, J. (2011). Coh-Metrix: Providing multilevel analyses of text characteristics. *Educational Researcher, 40,* 223-234.

Graesser, A. C., McNamara, D. S., Louwerse, M. M., & Cai, Z. (2004). Coh-Metrix: Analysis of text on cohesion and language. *Behavior Research Methods, Instruments, and Computers*, *36*, 193–202.

Hunt, K. W. (1965). *Grammatical structures written at three grade levels* (Research Report No. 3). National Council of Teachers of English.

Inoue, C. (2016). A comparative study of the variables used to measure syntactic complexity and accuracy in task-based research. *The Language Learning Journal*, *44*(4), 487–505.

Jarvis, S. (2013). Defining and measuring lexical diversity. In S. Jarvis & M. Daller (Eds.), *Vocabulary knowledge: Human ratings and automated measures* (pp. 13–44). John Benjamins.

Kessler, G. (2009). Student-initiated attention to form in wiki-based collaborative writing. *Language Learning & Technology, 13*, 79–95.

Kessler, G., Bikowski, D., & Boggs, J. (2012). Collaborative writing among second language learners in academic web-based projects. *Language Learning & Technology*, *16*(1), 91–109.

Kost, C. (2011). Investigating writing strategies and revision behaviour in collaborative writing projects. *CALICO Journal*, *28*(3), 606–620.

Lee, H., Joo, K., Moon, J., & Hong, Y. (2007). EFL learner performance variation as the effect of interlocutor type. In S. Van Daele, A. Housen, F. Kuiken, M. Pierrard, & I. Vedder (Eds.), *Complexity, accuracy and fluency in second language use, learning and teaching* (pp. 127–138). KVAB.

Lee, L. (2010). Exploring wiki-mediated collaborative writing: A case study in an elementary Spanish course. *CALICO Journal*, *27*(2), 260–276.

Li, M. (2018). Computer-mediated collaborative writing in L2 contexts: An analysis of empirical research. *Computer Assisted Language Learning*, *31*(8), 882–904.

Li, M., & Zhu, W. (2013). Patterns of computer-mediated interaction in small writing groups using wikis. *Computer Assisted Language Learning*, *26*(1), 62–81.

Li, Q. (2014). Get it right in the end: The effects of post-task transcribing on learners' oral performance. In P. Skehan (Ed.), *Processing perspectives on task performance* (pp. 129–154). John Benjamins.

Long, M. H. (1983). Native speaker/non-native speaker conversation and the negotiation of comprehensible input. *Applied Linguistics*, *4*, 126–141.

Long, M. H. (1985). A role for instruction in second language acquisition: Task-based language teaching. In K. Hyltenstam & M. Pienemann (Eds.), *Modelling and assessing second language acquisition* (pp. 77–99). Multilingual Matters.

Long, M. H. (1991). Focus on form: A design feature in language teaching methodology. In K. de Bot, R. Ginsberg, & C. Kramsch (Eds.), *Foreign language research in cross-cultural perspective* (pp. 39–52). John Benjamins.

Long, M. H. (1996). The role of the linguistic environment in second language acquisition. In W. C. Ritchie & T. K. Bhatia (Eds.), *Handbook of second language acquisition* (pp. 413–468). Academic Press.

Long, M. H. (2015). *Second language acquisition and task-based language teaching*. John Wiley & Sons.

Lu, X. (2010). Automatic analysis of syntactic complexity in second language writing. *International Journal of Corpus Linguistics, 15*(4), 474–496.

Mak, B., & Coniam, D. (2008). Using wikis to enhance and develop writing skills among secondary school students in Hong Kong. *System, 36*, 437–455.

McCarthy, P. M., & 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., Graesser, A. C., McCarthy, P. M., & Cai, Z. (2014). *Automated evaluation of text and discourse with Coh-Metrix*. Cambridge University Press.

Mehnert, U. (1998). The effects of different lengths of time for planning on second language performance. *Studies in Second Language Acquisition*, *20*, 83 – 108.

Michel, M. (2017). Complexity, accuracy, and fluency in L2 production. In S. Loewen & M. Sato (Eds.), *The Routledge handbook of instructed second language acquisition* (pp. 50–68). Taylor & Francis.

Microsoft. (2020). *Microsoft Forms*. https://forms.office.com/

Mirel, B., & Spilka, R. (Eds.). (2002). *Reshaping technical communication: New directions and challenges for the 21st century*. Lawrence Erlbaum.

Newton, J., & Nguyen, B. T. T. (2019). Task repetition and the public performance of speaking tasks in EFL classes at a Vietnamese high school. *Language Teaching for Young Learners*, *1*(1), 34–56.

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

Ortega, L. (2003). Syntactic complexity measures and their relationship to L2 proficiency: A research synthesis of college-level L2 writing. *Applied Linguistics, 24*, 492–518.

Plonsky, L., & Oswald, F. L. (2014). How big is "Big"? Interpreting effect sizes in L2 research. *Language Learning*, *64*(4), 878–912.

Prabhu, N. S. (1987). *Second language pedagogy*. Oxford University Press.

Révész, A., Kourtali, N. E., & 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. (2017). *Investigating IELTS academic writing task 2: Relationships between cognitive writing processes, text quality, and working memory* (IELTS Research Reports Online Series). British Council, Cambridge English Language Assessment and IDP: IELTS Australia.

Samuda, V., & Bygate, M. (2008). *Tasks in second language learning*. Palgrave Macmillan.

Scientific Software Development GmbH. (2020). *Atlas.ti 9*. https://atlasti.com/

Skehan, P. (1996). A framework for the implementation of task-based instruction. *Applied Linguistics*, *17*, 38–62.

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

Skehan, P. (2009). Modelling second language performance: Integrating complexity, accuracy, fluency, and lexis. *Applied Linguistics*, *30*(4), 510–532.

Skehan, P. (2013). Nurturing noticing. In J. M. Bergsleithner, S. N. Frota, & J. K. Yoshioka (Eds.), *Noticing and second language acquisition: Studies in honor of Richard Schmidt* (pp. 169–180). University of Hawai'i, National Foreign Language Resource Center.

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

Skehan, P., Xiaoyue, B., Li, Q., & Wang, Z. (2012). The task is not enough: Processing approaches to task-based performance. *Language Teaching Research*, *16*(2), 170 –187.

Stockwell, G., & Harrington, M. (2003). The incidental development of L2 proficiency in NS-NNS email interactions. *CALICO Journal, 20*, 337–359.

Storch, N. (2011). Collaborative writing in L2 contexts: Processes, outcomes, and future directions. *Annual Review of Applied Linguistics*, *31*, 275–288.

Storch, N. (2013). *Collaborative writing in L2 classrooms*. Multilingual Matters.

Storch, N. (2019). Collaborative writing. *Language Teaching*, *52*(1), 40–59.

Strobl, C. (2014). Affordances of Web 2.0 technologies for collaborative advanced writing in a foreign language. *CALICO Journal*, *31*(1), 1–18.

Suzuki, S., & Kormos, J. (2020). Linguistic dimensions of comprehensibility and perceived fluency: An investigation of complexity, accuracy, and fluency in second language argumentative speech. *Studies in Second Language Acquisition*, *42*, 143–167.

Van den Branden, K., Bygate, M., & Norris, J. M. (2009). *Task-based language teaching: A reader*. John Benjamins.

VanPatten, B. (1994). Evaluating the role of consciousness in second language acquisition: Terms, linguistic features and research methodology. *AILA Review*, *11*, 27–36.

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.

Vygotsky, L. S. (1981). The genesis of higher mental functions. In J. V. Wertsch (Ed.), *The concept of activity in Soviet psychology* (pp. 144 – 188). M. E. Sharpe.

Wang, Y. C. (2015). Promoting collaborative writing through wikis: A new approach for advancing innovative and active learning in an ESP context. *Computer Assisted Language Learning*, *28*(6), 499–512.

Wigglesworth, G., & Storch, N. (2012). What role for collaboration in writing and writing feedback. *Journal of Second Language Writing*, *21*, 364–374.

Willis, J. (1996). *A framework for task-based learning*. Longman.

Yu, S., & Lee, I. (2016). Peer feedback in second language writing (2005–2014). *Language Teaching, 49*(4), 461–493.

Zoom Video Communications. (2020). *Zoom*. https://zoom.us/