Journal of Second Language Writing 53 (2021) 100824

Contents lists available at ScienceDirect

Journal of Second Language Writing

journal homepage: www.elsevier.com/locate/jslw

Exploring the relationships among cognitive and linguistic resources, writing processes, and written products in second language writing

Minkyung Kim [a,*], Yu Tian [b], Scott A. Crossley [b]

[a] Nagoya University of Commerce and Business, Sagamine-4-4 Komenokicho, Nisshin, Aichi, 470-0193, Japan
[b] Georgia State University, USA

ARTICLE INFO

Keywords:
Second language writing
Linguistic knowledge
Working memory
Attention
Language burst
Pausing behavior

1. Introduction

For second language (L2) students[1], L2 writing skills are crucial for successful academic performance in L2 medium higher education (e.g., Evans et al., 2015). To produce stronger L2 texts, L2 learners need to have higher levels of linguistic and cognitive resources (e.g., L2 vocabulary knowledge and working-memory capacity [WMC]) to help coordinate a range of writing processes (e.g., planning and lexical retrieval), and write fluently (Hayes, 2009; Kellogg, 1996). While L2 writers likely depend on both cognitive and linguistic resources as well as writing processes, links among these variables are less clear as are their interactions with written products including text length and writing quality. The purpose of this study is to examine the relationships among cognitive and linguistic resources, writing processes, and writing products in an L2 English undergraduate writing context. In investigating writing,

* Corresponding author.
E-mail address: minkyung_kim@nucba.ac.jp (M. Kim).
[1] We define L2 students as those that speak a language other than a native language.

https://doi.org/10.1016/j.jslw.2021.100824
Received 30 March 2020; Received in revised form 3 May 2021; Accepted 3 May 2021
Available online 25 May 2021

we adopt a cognitive perspective in which writing is considered a process through which a writer creates meaning (Hayes & Berninger, 2014). Specifically, cognitive and linguistic resources included attention, working memory, L2 vocabulary knowledge, general knowledge, and L2 reading skills. Writing processes included mean P-burst lengths (mean characters produced between pauses) and mean pause times. Writing products included words per minute, text length, and text quality. Findings of this study may shed new light on how linguistic and cognitive resources and writing processes are related and how they may also be linked to L2 writing products among L2 English undergraduate students.

2. Hayes and Berninger's (2014) writing model

Adopting a cognitive perspective, the current study is informed by the writing-specific cognitive model proposed by Hayes and Berninger (2014). We chose this model because it is a more sophisticated version of the Hayes-Flower model of writing (1980), which has been influential in L1 writing research. However, it has not strongly informed theoretical motivations in L2 writing contexts. Based on the Hayes and Berninger model (2014), which incorporates various levels of cognitive processing (i.e., resource-, process-, and control-levels), this study attempts to provide information about how L2 writing can be influenced by cognitive resources and processes. More specifically, the Hayes and Berninger model (2014) includes three major levels of cognitive processing: the resource-, process-, and control-levels. First, the bottom, or resource, level represents general cognitive resources that writers draw on while composing. It consists of attention, WMC, reading, and long-term memory. Attention refers to "the ability to maintain focus on a task in the face of distraction" (Hayes & Berninger, 2014, p. 4). Working memory is a memory system that stores and processes information, which is important in retaining the relevant information and, simultaneously, turning ideas into written forms. Long-term memory is a complex resource which contains not only knowledge of language, including vocabulary and discourse schema, but also knowledge of the world, facts, episodes, and experiences. Reading is important during writing because writers typically read the text they have produced and reread it for revision and edits. Second, the middle, or process, level represents operation of, and interaction between, cognitive processes, including writing processes and the task environment. At the process level, writing processes include four main processes: a proposer, a translator, an evaluator, and a transcriber. The proposer suggests a package of ideas. The input of the proposer comes from various sources and resources, such as long-term memory and the text produced so far. The translator transforms ideas taken from the proposal into language strings of verbal forms (Hayes & Chenoweth, 2007). The transcriber then turns the language strings produced by the translator into written text. The evaluator judges the adequacy of all of the writing processes. Lastly, the top, or control, level represents factors that direct operations at the process level, including the task initiator and the planner (i.e., setting goals). Among these three levels, this study focuses on the resource level and the process level. We did not consider the task environment or the control level because they were set up by researchers (i.e., writing in a computer lab on a given prompt). In L2 writing contexts, previous studies have examined the roles that cognitive and linguistic resources, and writing processes play in L2 writing.

3. Cognitive and linguistic resources, writing processes, and written products

In this section, based on Hayes-Berninger model (2014), we present previous L2 studies which focused on the resource level (Section 3.1) and the process level (Section 3.2), followed by the current study (Section 3.3).

3.1. *Cognitive and linguistic resources and L2 writing*

The resource level of cognitive processing in the Hayes-Berninger model (2014) consists of attention, WMC, long-term memory, and reading. From a broader information-processing perspective, attention is described in two main ways: attention as selection of information for further processing, and attention as effort devoted to maintaining performance on a task (Sanders, 1998). Previous studies with a more focus on attention as effort to sustain a task have examined the role of attention in L2 learning and performance (e.g., Gass, 2011). For example, failure to maintain attention to L2 spoken communication likely has a negative impact on monitoring of output (Kormos, 1999). More recently, Gass (2011) reported that for English-speaking adult learners of Italian, learning gains from feedback during oral interactions were related to attention capacity (as measured by a Stroop task), such that high gainers had better attentional control than low gainers. In relation to L2 writing, however, to our knowledge, few if any studies have examined the role of attention in L2 writing. Potentially, attention as effort may help L2 writers sustain their performance on L2 writing tasks that typically place greater cognitive and linguistic demands.

WMC is generally defined as a limited capacity of a central executive processor that stores, retrieves, and manipulates the activated information (Engle, 2002). WMC has been widely researched in L2 research with increasing evidence that functions of WMC affect the efficiency of L2 learning and use (e.g., Linck, Osthus, Koeth, & Bunting, 2014; Williams, 2012). In the context of L2 writing where various writing processes are affected by the availability of working memory resources (e.g., Kormos, 2012), components of working memory seem to be important. However, previous studies have reported mixed results for WMC in L2 writing (Kormos & Sáfár, 2008; Michel, Kormos, Brunfaut, & Ratajczak, 2019; Révész, Michel, & Lee, 2017). For example, Kormos and Sáfár (2008) found that for ninth graders in Hungary, L2 (English) writing scores (measured using the writing section of Cambridge First Certificate Examination) were significantly correlated with scores of a non-word span task (recalling non-words after listening; $r = .48$), but not with scores of a backward digit span task (i.e., recalling a series of digits in reverse order). Révész et al. (2017) examined the roles of WMC (measured by digit span, non-word span, color shape task, block backward [recalling blocks in the reverse order of how they had been highlighted], stop signal task, and operation span) in 30 adult L2 English learners' writing (measured using the opinion-writing section of the IELTS), but none of the WMC scores reached a significant correlation with holistic writing scores. However, Révész et al. (2017)

found that WMC was related to online writing behaviors (e.g., better performance on the color shape task was strongly correlated with shorter between-sentence pause times; *rho* = .59). More recently, Michel et al. (2019) found that for 94 sixth and seventh graders in Hungary, a composite score of WMC (measured by forward and backward digit spans and symmetry span task) was not significantly related to L2 English writing scores (measured with the four writing tasks of the TOEFL Junior Comprehensive test) except for the academic editing task. Overall, these findings indicate a limited role for WMC on L2 writing scores, but suggest an important role for WMC on writing-specific processes such as pausing behavior and editing.

In addition to WMC, according to the Hayes and Berninger model (2014), long-term memory is mainly composed of general knowledge and knowledge of language. The use of general knowledge (i.e., retrieving relevant information from long-term memory) is considered important in producing texts (Hayes & Berninger, 2014; Hayes & Flower, 1980). Higher-levels of general knowledge likely facilitate flexible access to context-relevant ideas, which may in turn enhance writing processes, particularly for generating ideas and developing texts. In L2 writing contexts, while knowledge of writing topics is considered important (Bachman & Palmer, 1996; He & Shi, 2012), to our knowledge, little research has linked general knowledge to L2 writing performance. For linguistic knowledge presumably stored in long-term memory (Hayes & Berninger, 2014), vocabulary knowledge can be used as a proxy measure because vocabulary is considered key to learning an L2 (e.g., Ellis, 1997) and writing is a process of meaning making through the use of a range of words (Berninger, 2000). Much research has reported important roles for vocabulary knowledge in successful L2 writing (e.g., Milton, Wade, & Hopkins, 2010; Stæhr, 2008). For instance, Milton et al. (2010) reported that for L2 adult learners of English in U.K., vocabulary size scores were strongly correlated with writing scores (measured by the writing section of the IELTS; $r = .76$).

With respect to the effects of reading on writing, proficient readers, who are skilled at constructing a mental representation of the text and selecting important elements from the text, are also likely to effectively use these skills during writing to shape the mental representations for the texts they compose and select relevant content (Spivey, 1990). L2 research generally agrees that L2 reading and writing influence each other (e.g., Belcher & Hirvela, 2001; Hulstijn, 2015) with empirical evidence reporting that L2 reading and writing scores show moderate-to-strong correlations in Chinese adult learners of English ($r = .49$; Carson, Carrell, Silberstein, Kroll, & Kuehn, 1990) and Korean ninth graders learning English ($r = .73$; Pae, 2019). In L2 writing assessment, the role of reading in writing has been widely examined with the focus on reading-to-write tasks (e.g., Plakans & Gebril, 2017). For example, Plakans and Gebril (2017) found that for the reading-listening-writing section of the TOEFL, coherent representations of the reading and listening passages in test-takers' essays were related to higher writing scores. However, to our knowledge, few if any studies have examined the effect of L2 reading skills along with other cognitive skills (e.g., WMC) on predicting L2 writing performance.

In sum, a number of previous studies have examined the effects of various cognitive and linguistic resources on L2 writing. Beyond cognitive and linguistic resources that writers may use during L2 writing, online writing processes are also important. In the next section, previous research on L2 writing processes is presented.

3.2. Processes, fluency, and L2 writing

Writing processes have been widely examined in L2 writing (e.g., Silva, 1993). For example, studies have described general L2 writing processes, reporting that L2 writers produce texts through planning, translation, revision, and monitoring processes in a similar manner as L1 writers do (e.g., Raimes, 1987). However, as compared to L1 writers, L2 learners tend to retrieve and process linguistic knowledge for encoding ideas less automatically and produce texts less fluently (Chenoweth & Hayes, 2001).

In examining online writing processes, two of the important facets include language bursts and pause behaviors (e.g., Hayes & Berninger, 2014; Kaufer, Hayes, & Flower, 1986; Limpo & Alves, 2017). A language burst is defined as a chunk of letters or words produced between two consecutive pauses (Kaufer et al., 1986). In the Hayes and Berninger model (2014), the translator is considered a key source of language bursts, such that bursts are associated with the capacity of the translator for searching for appropriate linguistic forms to encode ideas (Hayes, 2009). Thus, burst length reflects "the capacity of the translator to handle complex language structures" (Chenoweth & Hayes, 2001, p. 94). Pause behaviors indicate periods of graphomotor inactivity during writing, being often expressed in terms of mean length of pauses (Kaufer et al., 1986; Limpo & Alves, 2017; Révész et al., 2017). Mean pause times are related to the period in which the translator may undergo cognitive processes, such as planning and searching for words (Limpo & Alves, 2017; Révész et al., 2017). Language bursts and pauses are interrelated, such that the writer repeats the process of producing one language burst and then having a pause. Importantly, language bursts and pause behaviors are also related to writing fluency because they are associated with the speed of writing. Greater burst lengths and shorter pause times are associated with greater fluency (e.g., Van Waes & Leijten, 2015). Thus, burst length and pausing behavior reflect not only writing processes, but also writing fluency.

In L2 contexts, fluency has been traditionally defined as "the extent to which the language produced in performing a task manifests pausing, hesitation, or reformulation" (Ellis, 2003, p. 342). Fluency is related to learners' access to, and control over, their L2 knowledge in language use. Thus, pausing behaviors directly reflect pausing and hesitation during online writing, while burst lengths reflect a status of not pausing. In this aspect, an investigation with keystroking data for L1 and L2 writing (Van Waes & Leijten, 2015) suggested a multi-dimensional nature of process-related writing fluency as composed of four main subcomponents: production, process variance, revision, and pausing behavior.

Beyond the conceptualization of writing fluency at the process level, writing fluency can also be defined at the product-level. In writing literature, there are at least two commonly used writing fluency measures. One measure is length of text produced in timed writing, while another is a production rate measured by dividing text length by writing time (i.e., words per minute [WPM]; Chenoweth & Hayes, 2001). Both of text length and WPM have been used as valid measures of writing fluency in previous studies. For example, Knoch, Rouhshad, Oon, and Storch (2015) used essay length in timed essays as a fluency measure, while Limpo and Alves (2017) used WPM as a fluency measure.

Recently, an increasing number of studies have examined online writing behaviors in L2 writing (Michel et al., 2020; Révész et al., 2017; Spelman Miller, Lindgren, & Sullivan, 2008; also see the special issue edited by Révész & Michel, 2019). Spelman Miller et al. (2008) examined the role of writing behavior on expository writing in high school L2 (English) writers in Sweden and found that writing scores were predicted by burst lengths, but not by pausing indices. The above reviewed study by Révész et al. (2017) reported that writing scores were significantly correlated with burst lengths (*rho* = .38) and pause frequencies per 100 words (*rho* = −.36). Examining the potential links between writing behaviors and the end products of writing would merit further consideration.

3.3. Current study: research question

To summarize, previous studies have examined linguistic and cognitive resources, processes, and fluency in relation to L2 writing. However, previous research has focused on a few focal variables. Consequently, a more comprehensive understanding of the relationships among cognitive and linguistic resources, online writing processes, and writing products would provide a more robust understanding of L2 writing. Also, as evident from the previous literature review, while fluency can be measured at both the process level and the product level, the relationship between the two in L2 writing is less clear. To fill these gaps, the current study examines the relationships of cognitive and linguistic resources (i.e., attention, WMC, long-term memory, and reading) and writing processes (i.e., language bursts and pause behaviors) with three end products of L2 writing–words produced per minute (WPM), text length, and text quality. Thus, we considered language bursts and pause behavior as fluency measures at the process-level, while WPM and text length were considered fluency measures at the product-level. The study is guided by the following research question (see Section 4.4.3 for hypothesized models):

What are the relationships among cognitive and linguistic resources, writing processes, and writing products in undergraduate L2 writing?

4. Method

4.1. Participants

Participants were 100 undergraduate students (67 female) from a research-oriented university located in the southern United States. Participants were from various countries, including Colombia, Ethiopia, India, Iran, Mexico, Pakistan, Venezuela, Viet Nam, and Zimbabwe. On average, students were 20.53 years old (*SD* = 2.82), started learning English at an age of 7.59 (*SD* = 4.37), studied English in both immersion and classroom settings for 12.94 years (*SD* = 4.12), and studied English in immersion instruction settings for 7.96 years (*SD* = 6.27). See supplementary materials A for more details.

4.2. Instruments

4.2.1. Background survey

Questions in the background survey concerned participants' L1s, age, major, academic year, gender, citizenship, age of initial English learning, years of learning English, and years living in an English-speaking country. We asked participants' citizenship to know their country of origin.

4.2.2. Attention

As suggested by Hayes and Berninger (2014), attention was measured using a Stroop test in which participants were asked to name font colors in English within a given time limit. Using English in the Stroop test was considered acceptable because participants were matriculated undergraduate students in an English-medium university, and did not report having difficulties in naming basic color words (e.g., red) in English. Two conditions were used: congruent and incongruent. Five ink colors were used: red, green, brown, blue and purple. In the congruent condition, participants named font colors that appeared with the symbols "@@@@" repeatedly in the five ink colors. In the incongruent condition (i.e., interference condition), participants were asked to name the font colors for words whose meanings did not match the font color. For example, participants saw the word 'blue' in a green font color and were instructed to say 'green'. The order of presenting the conditions was random. Each condition had 20 items. Within each condition, items were randomly presented. Each item lasted two seconds. This test took approximately five minutes to complete. For each condition, participant's average response times were calculated in milliseconds. When errors occurred or no response was recorded, their response times were not calculated. The average accuracy rates for the congruent and incongruent conditions were 99.35 and 98.25, respectively. Following Ludwig, Borella, Tettamanti, and de Ribaupierre (2010), a normed naming interference index, in which lower scores (i.e., quicker responses) means better attentional capacity, was calculated as follows:

$$\frac{\text{response times in the incongruence condition} - \text{response times in the congruence condition}}{\text{response times in the congruence condition}}$$

4.2.3. Working memory capacity

To measure WMC, a written letter running span task (which was developed to measure the capacity to store and rehearse information and variation in verbal WMC) with automated, built-in instructions and practice sessions was used (Kim, Payant, & Pearson, 2015). We chose this test because verbal WMC may be used in writing during the translation process to store and manipulate verbal information. During the running span task, participants were presented with a random series of letters (e.g., FGHJKQ) in written form, and then asked to choose the last *n* letters in the same order as presented (target length = 3, 4, 5, 6, or 7 letters). For example, after being presented with five letters (e.g., HJLQR), participants were asked to recall the final three letters in correct order (i.e., LQR). Using English letters was considered acceptable because participants were undergraduate students in an English-medium university and reported no difficulties in recognizing and rehearsing English letters. This test took approximately 10 min to complete. The total number of correctly remembered letters regardless of whether the letters were remembered in the presented order was scored. Running span tasks, such as a span task used in this paper, are found to closely relate to language proficiency and skills (Linck et al., 2014).

4.2.4. Western-based general knowledge

Students' general knowledge was assessed using a 30-item test that asked about students' knowledge about Western-based science, literature, and history (Roscoe, Crossley, Snow, Varner, & McNamara, 2014). The use of this test has been validated in L1 research with high reliability (Roscoe et al., 2014). Each item had four choices with one correct answer. An example item is *Who is the author of the mystery fiction Sherlock Holmes?* Questions were presented in a random order. One limitation is that the test was likely biased in favor of participants who studied in English-speaking environments because some questions were associated with English and American literature and history. Cronbach's alpha for the general knowledge test ($k = 30$) was .67. While a traditional threshold for Cronbach's alpha as indicative of acceptable reliability is .70, alpha values above .60 are also considered acceptable or reasonable, especially in the administration of instruments that test a broad range of distinct knowledge like our general knowledge test (Taber, 2018).

4.2.5. English vocabulary knowledge

Vocabulary knowledge was assessed using two different vocabulary sections of the Gates-MacGinitie Reading Skill tests (Fourth Edition; MacGinitie, MacGinitie, Maria, & Dreyer, 2002): Level 10/12, Forms S and T. The Gates-MacGinitie Reading Skill tests were chosen because they are standardized tests that have two different, comparable forms and have been widely used in both L1 and L2 contexts (e.g., Crossley, Yang, & Mcnamara, 2014). Each of the two Gates-MacGinitie vocabulary tests had 45 multiple-choice questions. Participants were given 20 min to complete the vocabulary test. The tests were counterbalanced across participants, such that 50 participants took the Form S, while 50 took the Form T. To compare scores from the two different test forms, a normed scale for extended scale scores (ESSs) developed based on all test items from Level 3 through 10/12 and Adult Reading (MacGinitie et al., 2002) was used. The Cronbach's alpha values for the Form S and the Form T were .88 and .87, respectively.

4.2.6. English reading comprehension skills

For assessing English reading comprehension skills, two forms of the Gates-MacGinitie Reading Skill tests (Fourth Edition; MacGinitie et al., 2002) were used: Level 10/12, Forms T and S. The Gates-MacGinitie reading tests comprised 48 multiple-choice questions with passages from various domains, including narratives, autobiographies, and academic texts. Participants were given 40 min to complete the test. To compare scores from the two different test forms, a normed scale for ESSs (MacGinitie et al., 2002) was used. The tests were counterbalanced across participants. The Cronbach's alpha value for both of the Form S and the Form T was .89.

4.2.7. English writing ability

English essays were written in response to SAT-based prompts. The SAT is a common assessment of college readiness in the United States. Although SAT writing is a type of pseudo-academic five-paragraph essay, we selected SAT prompts because they lead to persuasive writing, which can demonstrate the writer's ability to make a coherent argument to persuade readers (Gardner & Nesi, 2013). The ability to produce a well argumented essay is crucial in academic contexts in higher education. Two SAT-based prompts were used to control for the prompt effects. One was about competition and the other was about appearance (see supplementary materials B for prompts). Students were given 25 min to complete the writing task. Students were asked to write one essay in a response to one of the two prompts. Prompts were randomly assigned. The writing test forms were counter-balanced, such that 50 participants wrote essays about competition, while the other 50 wrote essays about appearance. Spelling and grammar checks were not available. During writing, participants' keyboard strokes were recorded using the keystroke logging program Inputlog (Leijten & Van Waes, 2013).

4.3. Procedure

Each participant was seated in a laboratory room equipped with one desktop computer. Participants attended one session that lasted around two hours. They first signed an informed consent form, and then completed the background survey. Participants then completed reading comprehension, writing, vocabulary, attention, WMC, and general knowledge tests. The order of these six test batteries was counterbalanced. The running span task and the Stroop test were administered using E-Prime 2.0 software (Schneider, Eschman, & Zuccolotto, 2012). The other test batteries and the background survey were administrated using the Qualtrics Research Suite software (Qualtrics, Provo, UT). While carrying out all of the tasks, participants were not allowed to use any reference tools, including dictionaries. After completion, participants were remunerated.

4.4. Data analysis

4.4.1. Measures of writing processes

From keystroke logging data, we used two writing-process-related measures that have been commonly used in previous studies

(Limpo & Alvès, 2017; Révész et al., 2017; Barkaoui, 2019): mean P-burst length (i.e., mean number of characters produced between pauses; Limpo & Alves, 2017) and mean pause time. Following previous studies (Chenoweth & Hayes, 2001; Limpo & Alves, 2017; Révész et al., 2017), we set the threshold pause time to two seconds for both P-burst length and pause time because it reflects the minimal period that is required for transforming ideas into meaningful language strings (see supplementary materials C for more details). That is, threshold levels below two seconds may be more related to lower-level cognitive processes, such as spelling and quick lexical access (Limpo & Alvès, 2017). P-burst lengths included all of the characters produced between two seconds pauses. While we measured typing skills, which tended to relate to P-burst lengths, we did not include them in the analysis (see supplementary materials D for more details). Instead, we presumed that P-burst lengths reflect typing skills to some degree (Alves, Castro, de Sousa, & Strömqvist, 2007).

4.4.2. Measures of writing products

Using the writing data collected, three variables at the product level were calculated: words per minute (WPM), text length (i.e., word counts in final texts), and text quality (i.e., scores). WPM was calculated by dividing text length by writing time using individuals' total writing minutes, rather than the given time of 25 min. It should be mentioned that WPM tended to vary as a function of typing skills (see supplementary materials D for more details). Text quality was assessed by two trained raters using a six-point holistic rating scale developed for the SAT (see supplementary materials E). The rating scale holistically assessed essay quality on a range of one to six. The two raters were doctoral students in applied linguistics who were familiar with university writing. After a training session, the raters scored the essays collected for this study independently. If two ratings differed by more than one point, the raters discussed the ratings so that the disagreement between the raters was one point or less. Prior to adjudication, inter-rater reliability was acceptable with a Cohen's Kappa value of .77. Average scores between the raters were calculated for each essay.

4.4.3. Statistical analysis

To examine the relationships among cognitive and linguistic resources, writing processes, and writing products in undergraduate L2 writing, a path analysis was conducted using the *Lavaan* package (Rosseel, 2012) in *R* (R Core Team, 2018). Path analysis uses a correlation matrix as input and is similar to linear regression analysis in that path analysis estimates regression weight (Alwin & Hauser, 1975). However, unlike regression analysis, path analysis can include structural and hierarchical relationships by having multiple dependent variables as well as variables that are hypothesized to affect dependent variables indirectly through one or more intervening variables. It should be noted that the directionality specifications of a path model are assumed and hypothesized prior to data analysis based on clear rationales for each path (Kline, 2015). See supplementary materials F for a full description of statistical analysis.

For modeling, we took two main steps: (1) predicting three product-based variables (i.e., WPM, text length, and text quality) in

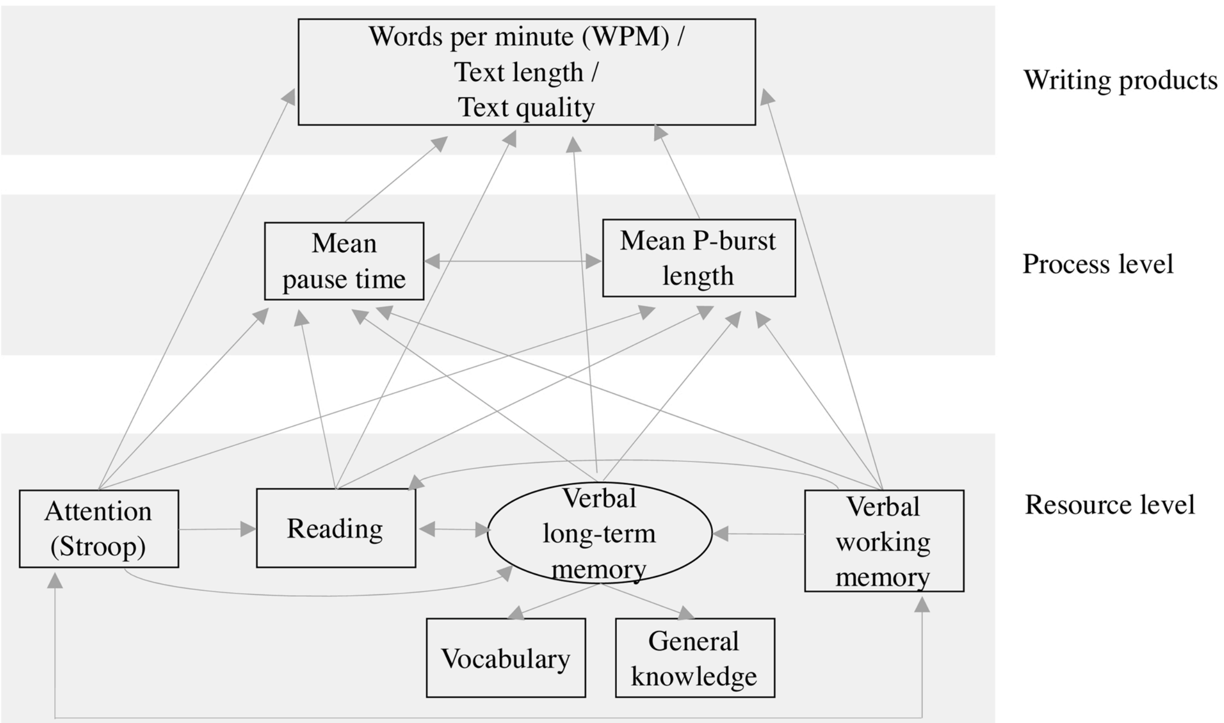

Fig. 1. Hypothesized models that predict WPM, text length, and text quality.
Note: Rectangles indicate observable variables, while an oval indicates a latent variable.

separate models using resource-level and process-level variables; and (2) building a final model that includes variables which significantly predicted the three product-based variables. Below, these steps are presented in detail.

4.4.3.1. Predicting three product-based variables. Three separate models were tested to predict the three product-based variables (i.e., WPM, text length, and text quality) using the resource-level and process-level variables found in Hayes and Berninger (2014) under the assumption that they would influence product-based variables. Separate models were first constructed to lessen the complexity of the models. Fig. 1 shows hypothesized models. Predictive paths (whether one variable would primarily influence another) were drawn in single-direction arrows, while covariance paths (where two variables would influence each other) were drawn in double-direction arrows. A latent variable is shown in an oval.

In Fig. 1, the resource level included attention, verbal WMC, reading ability, vocabulary knowledge, and general knowledge. A latent variable of verbal long-term memory informed by vocabulary and general knowledge was formed because the Hayes and Berninger's model (2014) suggests that both linguistic and general knowledge are part of long-term memory. While the relationships among the resources were not explicitly stated, we assumed that attention and verbal WMC would operate as foundational cognitive resources, while verbal long-term memory and reading ability would function as higher-order resources. This assumption is based on Kim and Park (2019), who proposed that WMC and attention are conceptualized as foundational cognitive functions. WMC and attention are necessary for many general learning tasks, including vocabulary learning and discourse comprehension (e.g., Linck et al., 2014). Indeed, L2 research has found that individuals with greater WMC tend to perform better on learning foreign vocabulary in laboratory settings (Williams & Lovatt, 2003) and have better L2 reading comprehension skills (Jeon & Yamashita, 2014). Thus, verbal long-term memory and reading were hypothesized to be influenced by attention and verbal WMC. In addition, we hypothesized that attention and verbal WMC would influence each other at the foundational cognitive level, while reading and working-memory would influence each other at the higher-order level.

In the process level, we focused on writing processes using measures of mean P-burst length and mean pause times. A covariance path was drawn between P-burst length and pause time because both were considered part of the writing processes and may influence each other. In addition, given that the P-burst length may be influenced by the writer's language experience and cognitive resources (Chenoweth & Hayes, 2001), prediction paths from the resource-level variables to the process-level variables were added.

In the product level, writing products were measured by WPM, text length, and text quality. These product-based variables were predicted by all of the variables at the process and resource levels. The rationales for these predictions were partly supported by previous studies which have shown that greater cognitive and language resources tend to be related to greater essay quality (e.g., Carson et al., 1990; Chenoweth & Hayes, 2001; He & Shi, 2012; Kormos & Sáfár, 2008) and that longer P-burst lengths and shorter pause times were related to greater WPM and better texts (e.g., Alves & Limpo, 2015; Limpo & Alves, 2017).

4.4.3.2. Building a final model. Based on the models that predicted the three product-based variables (i.e., WPM, text length, and text quality), we further hypothesized the hierarchical relationships (see Fig. 2), such that WPM would positively influence both length and quality, and length would positively influence essay quality. Specifically, greater WPM would lead to greater essay length and better essays, and greater length would link to better quality.

5. Results

Table 1 displays descriptive statistics and correlations among the measured variables. Due to technical issues, one student's Stroop results and two students' keystroke logging data were not recorded, and thus handled as missing data. In addition, one student's mean pause time indicated an extreme outlier (25.02 s with a *z*-score of 7.72), which resulted in a high kurtosis value of 12.92. After removing this student's mean pause time to handle it as missing data, the kurtosis value of pause times in the remaining participants was acceptable (3.48). For writing scores, an independent samples *t*-test indicated that there was no significant score difference between international students (Mean = 2.94, *SD* = 1.11) and non-international students (Mean = 2.88, *SD* = 1.17): $t(80.8) = -.28$, $p = .78$.

The three hypothesized models shown in Fig. 1 failed to converge. The examination of covariances among the variables revealed that the lack of the convergence was likely due to strong correlations of vocabulary and general knowledge with reading scores ($.56 \leq rs \leq .62$). To solve the convergence issue, given that vocabulary, general knowledge, and reading are related to English literacy (e.g., reading helps increase both vocabulary and general knowledge), a latent variable of *literacy-related resources* was created, which was informed by vocabulary, general knowledge, and reading scores. With the *literacy-related resources* latent variable formed, each model had an acceptable fit to the data: the WPM model (χ^2 [10] = 16.69, $p = .08$, $\chi^2/df = 1.67$, CFI = .94, SRMR = .05), the text length model

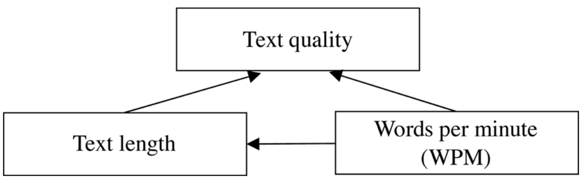

Fig. 2. Hypothesized hierarchical relationships among WPM, text length, and text quality.

Table 1
Descriptive statistics and correlations.

Variable	1	2	3	4	5	6	7	8	9	10	*n*	Mean	*SD*	Min.	Max.	Skewness	Kurtosis
1. Stroop (attention)	1										99	.20	.16	.01	.96	1.62	4.26
2. Running span (verbal WMC)	–.12	1									100	41.40	11.07	13	63	−.01	−.71
3. Vocabulary	.00	.27	1								100	555.59	27.61	495	612	.03	−.68
4. Western-based general knowledge	–.11	.34	.56	1							100	16.89	4.18	7	25	−.09	−.66
5. Reading	–.09	.31	.62	.59	1						100	567.61	26.39	513	653	.37	.45
6. Mean P-burst length in characters	.14	.07	.22	–.07	.03	1					98	50.28	32.94	16.04	200.50	1.90	4.44
7. Mean pause time in seconds	.20	–.01	–.21	–.09	–.18	–.06	1				97	6.25	2.43	3.14	16.07	1.81	3.48
8. Words per minute (WPM)	.03	.17	.26	.03	.15	.82	–.31	1			98	17.47	7.49	3	42.54	.83	.51
9. Essay length	–.02	.22	.29	.25	.45	.19	–.33	.40	1		100	292.13	109.97	64	551	.16	−.26
10. Essay score	–.03	.04	.30	.36	.49	.07	–.35	.22	.56	1	100	2.92	1.13	1	5.50	.26	−.40

Note. $|r| \geq .20$, $p < .05$; $|r| \geq .27$, $p < .01$.

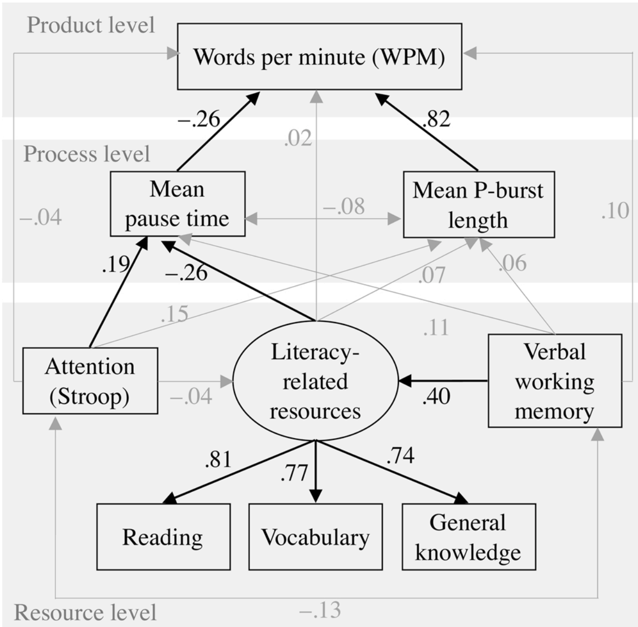

Fig. 3. Path model that predicts WPM.
Note: Statistically significant paths ($p < .05$) are indicated by black lines, while nonsignificant paths are indicated by grey lines.

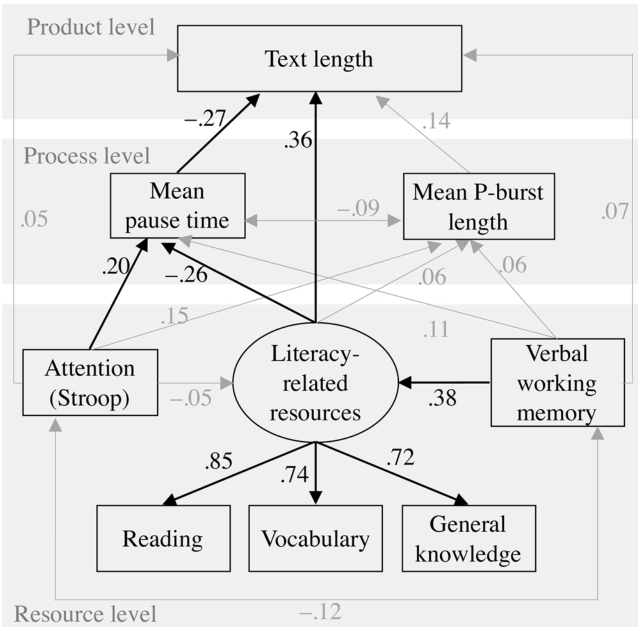

Fig. 4. Path model that predicts text length.
Note: Statistically significant paths ($p < .05$) are indicated by black lines, while nonsignificant paths are indicated by grey lines.

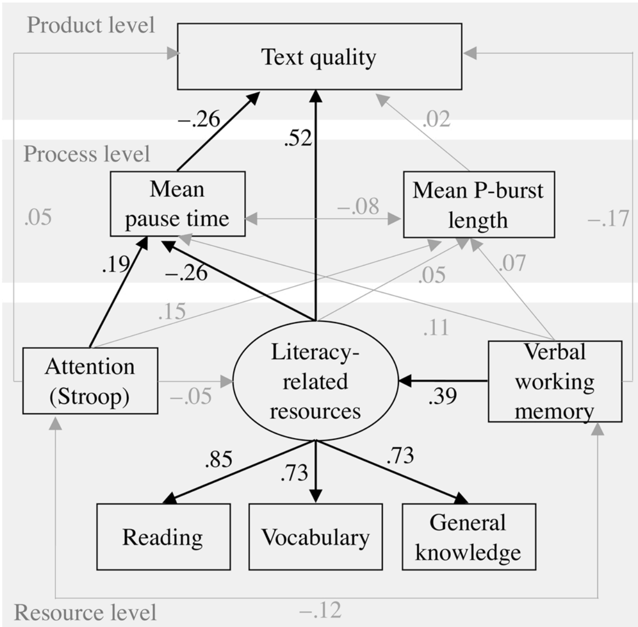

Fig. 5. Path model that predicts text quality.
Note: Statistically significant paths ($p < .05$) are indicated by black lines, while nonsignificant paths are indicated by grey lines.

(χ^2 [10] = 21.38, p = .02, χ^2/df = 2.14, CFI = .93, SRMR = .04), and the text quality model (χ^2 [10] = 21.44, p = .02, χ^2/df = 2.14, CFI = .93, SRMR = .04).

Standardized path coefficients of the three models are shown in Fig. 3 (predicting WPM), Fig. 4 (predicting text length), and Fig. 5 (predicting text quality). At the resource level, the latent variable of *literacy-related resources* was significantly related to verbal WMC (.38 ≤ βs ≤ .40), but not to attention. Attention and verbal WMC were unrelated. At the process level, P-burst length and pause time were also unrelated. Across the resource and process levels, mean pause times were significantly positively related to the Stroop scores (.19 ≤ βs ≤ .20) and significantly negatively related to the latent variable of *literacy-related resources* (βs = −.26), but not to verbal WMC. This indicates that longer pause times were associated with longer latencies during the Stroop test (i.e., lower attentional capacity) and lower levels of literacy-related resources. Mean P-burst length was not related to attention, verbal WMC, or the latent variable of *literacy-related resources.* At the product level, all of the WPM, text length, and text quality were significantly negatively related to mean pause times (−27. ≤ βs ≤ −.26). The latent variable of *literacy-related resources* showed significant positive effects on

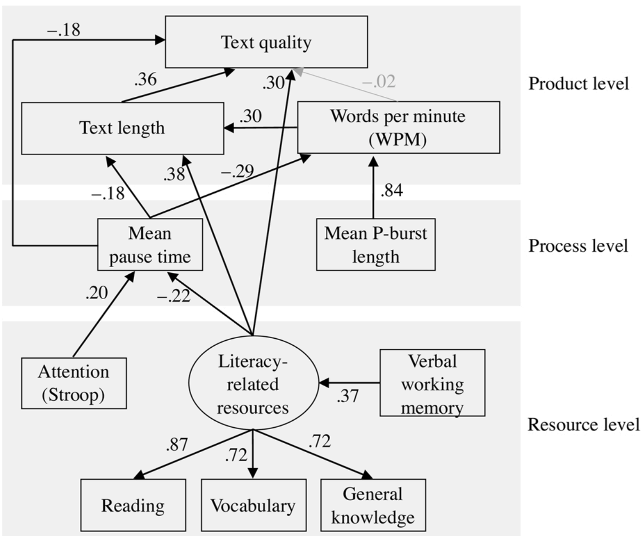

Fig. 6. Final model.
Note: Statistically significant paths ($p < .05$) are indicated by black lines, while nonsignificant paths are indicated by grey lines.

text length (β = .36) and text quality (β = .52), but not on WPM. Mean P-burst length showed significant positive effects on WPM (β = .82), but not on text length or text quality. Attention and verbal WMC were not related to WPM, text length, or text quality. Lastly, 76.5 % of variance in WPM, 30.2 % of variance in text length, and 35.8 % of variance in text quality were explained, respectively.

Using significant paths only in the three models, a final model was created to predict text quality using the remaining product level items and the process and resource level items. The final model had an acceptable fit: χ^2 (27) = 41.03, p = .04, χ^2/df = 1.52, CFI = .96, SRMR = .06 (Fig. 6). Fig. 6 shows standardized path coefficients of the final model. In addition, at the product level, WPM had a significant positive effect on text length (β = .30), but not on text quality. Text length had a significant positive effect on text quality (β = .36). The final model explained 40.9 % of variance in text quality.

In addition, based on the final model, as a post hoc analysis, we examined the potential indirect effects on L2 writing scores among the significant paths. Specifically, we investigated a total of nine indirect paths that began from verbal WMC, attention, and mean length P-burst (which were not predicted by any other variables) and linked all the way through to the end variable of text quality. However, none of the indirect paths was significant (see supplementary materials G for the full results).

6. Discussion

The research question addressed the relationships among cognitive and linguistic resources, writing processes, and writing products in undergraduate L2 writing. We first discuss findings in terms of the relationships within and across the resource-, process-, and product-levels, followed by overall discussion and implications.

6.1. Relationships of variables within the resource, process, and product levels

At the resource level, findings showed that L2 vocabulary, general knowledge, and L2 reading skills were strongly correlated, forming a single latent variable related to L2 literacy. While not in line with Hayes and Berninger's model (2014) that assumed the presence of the long-term memory separate from reading skills, our finding suggests general knowledge, L2 vocabulary knowledge, and reading skills were closely related. With respect to the relationships of resource-level variables, verbal WMC measured by the running span tended to be related to the latent variable of *literacy-related resources* within the context of this study. This result likely supports the importance of WMC in vocabulary knowledge (e.g., Williams & Lovatt, 2003), reading skills (e.g., Jeon & Yamashita, 2014), and general knowledge (e.g., Hambrick & Engle, 2002). On the other hand, attention measured using the Stroop test did not tend to relate to verbal WMC measured by the running span test, which indicates that the two measures may represent two different cognitive functions. It should also be noted that because we used one L2-based measure for each of attention and WMC, results may differ if other tests are used. Finally, no relationship between attention as measured by the Stroop test and the latent variable of *literacy-related resources* in the context of this study was found. This may reflect that attention, or performance on the Stroop test more specifically, is not strongly related to L2 literacy.

At the process level, mean pause time did not tend to relate to mean P-burst length when measured with a pause threshold level of 2000 ms. This might be partly due to a mutually exclusive distribution of production and pausing behavior (i.e., production and pausing do not occur simultaneously). This finding supports a multi-dimensional nature of process-related writing fluency (Van Waes & Leijten, 2015), indicating that production and pausing behavior may represent different aspects of writing fluency, and thus should be considered together to examine writing processes.

At the product level, a greater number of words produed in the final text per minute led to longer texts. Given that both WPM and text length in timed writing are related to speed fluency, this result seems expected. On the other hand, writing scores were predicted by essay length, but not by WPM, such that lengthier essays were also rated higher. These findings indicate that producing a larger number of words in a shorter period (which is also likely related to stronger typing skills) may lead to longer texts, but may not necessarily entail better writing.

6.2. Relationships of variables between the resource and process levels

Across the resource and process levels, two significant paths only were found. These paths were related to mean pause times. Results indicated that shorter mean pause times were related to shorter response times during the Stroop test and higher levels of literacy-related resources. This finding supports the notion that L2 writers with greater attentional capacity and greater L2 literacy-related resources tended to have shorter mean pause times during writing processes. It may be that better attentional capacity helps L2 writers direct more resources towards writing and remain focused allowing them to express their messages as fluently as possible within the given time (Hayes & Berninger, 2014). Also, having greater L2 literacy-related resources may help L2 writers better generate ideas, search for lexical items, and read (and re-read) the text they have produced, which in turn may lead to more fluent writing processes with shorter mean pause times.

On the other hand, P-burst length, which is considered important in searching for appropriate linguistic forms to encode ideas (Hayes, 2009), did not show any relationship with verbal WMC, attention, or L2-literacy-related resources. This finding may be partly due to the participants' relatively higher level of L2 proficiency (i.e., matriculated undergraduate students), such that their processes of translating ideas into L2 forms (operationalized by P-burst length with a pause threshold of 2000 ms) may take place rather automatically, not be influenced by L2-literacy-related resources, and not tax the processing and storage functions of verbal WMC or attentional capacity. This finding is in line with Révész et al. (2017) who reported no relationship between phonological short-term memory measures (similar to the verbal WMC measure in this study) and P-burst lengths with a 2000 ms pause threshold in L2 adult learners. It is also worthy of mentioning that Révész et al. (2017) reported a significant correlation between phonological short-term memory and the greater use of frequent words (*rho* = .60). Likewise, we found a significant though weak correlation between P-burst length and L2 vocabulary test scores (*r* = .22). In this regard, mean P-burst lengths with a 2000 ms pause threshold may be related to L2 vocabulary knowledge and use. Furthermore, there is a possibility that P-burst lengths with a 2000 ms pause threshold may be reflective of typing skills, rather than a pure measure of translating processes (i.e., turning ideas into verbal forms). Finally, P-burst lengths measured in a fine-grained manner with different pause thresholds (e.g., 200 ms and 500 ms), at different locations in texts (e.g., within-words and between-words), and at different time points (e.g., beginning, middle, and end; Barkaoui, 2019; Michel et al., 2020; Révész, Michel, & Lee, 2019) may lead to different relationships between P-burst lengths and resource-level variables. However, more research on this matter would be needed.

6.3. Relationships of variables between the resource and product levels

Across the resource and product levels, L2-literacy-related resources only were related to two product-related variables, i.e., text length and text quality, such that higher levels of L2-literacy-related resources were related to longer texts and higher-rated essays. This finding corroborates previous research that has reported the importance of L2 knowledge (e.g., Milton et al., 2010; Stæhr, 2008) and L2 reading (e.g., Belcher & Hirvela, 2001) in L2 writing, while adding a new finding that supports the importance of Western-based general knowledge in L2 writing. Specifically, it seems that higher levels of L2 knowledge, general knowledge, and L2 reading skills may help L2 writers to be free of constraints on various writing processes (e.g., the excessive attention drawn to linguistic concerns at the expense of idea generation and the difficulties in finding appropriate lexical and syntactic forms more automatically). This in turn may lead to fluent writing expressed in terms of longer texts and better text quality expressed in terms of writing scores.

Verbal WMC and attention variables did not show any of direct relations with the product-level variables (i.e., WPM, text length, and text quality). This result is similar to Révész et al. (2017) who reported no correlations between WMC measures and writing scores in L2 adult learners. The lack of the relationships between verbal WMC and attention (as measured in this study) with the end products of L2 writing can be accounted for with at least two reasons. First, cognitive processing demands during the verbal WMC and attention tests may mismatch those during L2 writing. That is, verbal WMC functioning and attentional resources, which are more related to storing and processing information occurring at shorter moments of time (less than one minute), may not impact text length or writing quality occurring for a longer period at the discourse level. Second, verbal WMC and attention may be relevant more closely to writing processes of younger learners who increasingly develop verbal WMC and attentional capacities (Kormos & Sáfár, 2008; Michel et al., 2019; Roy et al., 2018). Furthermore, for writing processes of adults with cognitive maturity, when producing argumentative writing, other abilities, such as critical thinking, argumentation, and evidence providing, may be important. Thus, more research on the relationships between cognitive resources and L2 adult writing is needed.

6.4. Relationships of variables between the process and product levels

Between the process and product levels, mean pause times significantly predicted the three product-based variables, which indicates that shorter mean pause times were linked to the greater production of words per minute, longer texts, and higher-rated essays. This result supports previous studies that have reported the links between shorter pause times and greater WPM in L1 children (Limpo & Alves, 2017) and between less frequent pauses and higher writing scores in L2 adult learners (Révész et al., 2017). On the other hand, mean P-burst length significantly predicted WPM only, which seems in line with previous research that reported the facilitative effect of P-burst length in WPM (Limpo & Alves, 2017). However, the lack of links between P-burst length with text length and text quality was puzzling given the important role of P-burst length in models of writing (Chenoweth & Hayes, 2001; Hayes & Berninger, 2014) and findings of previous studies that reported the facilitative effects of P-burst length in writing scores (Limpo & Alves, 2017; Révész et al., 2017; Spelman Miller et al., 2008). A possible explanation is that due to the low-stake nature of the writing test in our study, some students may have adopted a stream-of-conciousness writing stragegy leading to several very long P-bursts that produced a shorter essay. That is, producing several larger chunks of written language with fewer interruptions may not necessarily end up with longer texts or better essay quality. Another possible explanation is that the heterogeneity of our participants, who came from a variety of L1 backgrounds,[2] may have led to a greater variety of behaviors when producing written language after pauses longer than two seconds. However, an in-depth analysis on P-burst lengths across participants is needed to support this hypothesis.

6.5. Overall discussion and implications

Taken together, we report two main findings. First of all, our findings indicate that at the resource-level, a latent variable of L2-literacy-related variables (i.e., L2 vocabulary, L2 reading, and general knowledge related to literature, history, and science) was formed and was the only significant variable that predicted L2 writing quality. We also found that foundational cognitive resources (attention and verbal WMC) did not affect product-level variables, but results hinted at their indirect effects, though not significantly. Attention positively impacted mean pause times (i.e., greater attentional capacity, shorter pause times), which in turn influenced L2 writing quality (i.e., shorter pause times, higher-rated essays). Verbal WMC positively influenced literacy-related resources, which in

[2] Previous studies have generally recruited participants with the same L1s. See, for instance, the Portuguese native speakers in Limpo and Alves (2017), the Swedish native students in Spelman Miller et al. (2008), and the Chinese students in Révész et al. (2017).

turn positively affected L2 writing quality. These findings have theoretical implications in relation to the writing model of Hayes and Berninger (2014), such that in L2 writing contexts, particularly for adult L2 learners, the model can further highlight literacy-related variables over foundational cognitive resources. We also find that verbal long-term memory measured by L2 vocabulary knowledge and general knowledge is closely related with L2 reading skills. Furthermore, these findings have pedagogical implications. For instance, to help L2 students in tertiary academic contexts improve their L2 writing, instructors can direct their instructional targets toward L2 vocabulary, L2 reading, and general knowledge, which are teachable goals. Teachers could focus specifically on topic specific vocabulary and background readings that may help students develop the necessary knowledge to be successful in the writing context. Teachers' concerns about L2 students attention levels or working memory capacity during writing should also be mitigated because they were not found to be closely related to L2 writing in this study.

Second, our findings show relationships of writing fluency at process- and product-levels in relation to L2 writing quality, which, to our knowledge, have been rarely examined. Specifically, mean P-burst length predicted WPM, but not text length. This finding partly supports the role of P-burst length in writing as expressed in the model of Hayes and Berninger (2014) and the notion that "the longer the bursts, the greater the fluency" (Limpo & Alves, 2017, p. 309). As a reminder, our findings indicate the longer the P-bursts, the greater the fluency in terms of WPM, but the fluency at the entire text level (measured by text length) is not greater. In addition, speed fluency measures of writing (P-burst lengths and WPM, which are also related to typing skills) did not predict L2 writing quality, which suggests that they may not relate to L2 writing quality in the context of this study. On the other hand, mean pause length, which can represent breakdown fluency, was important in predicting L2 writing quality, supporting the importance of pausing behavior in timed L2 writing (Barkaoui, 2019; Révész et al., 2017). These findings have pedagogical implications as well. For instance, to help L2 students in tertiary academic contexts produce better essays in timed settings, instructors can inform students of the importance of fluent writing in terms of shorter pause times and producing longer texts during the given limited time. Teachers could work with students to generate ideas before writing so they could practice writing with fewer pauses while producing longer texts. Teachers could also be alert to long pauses in students' writing processes, examine reasons underlying these pauses, and provide appropriate scaffolding (e.g., language or strategies support) if necessary. This study implies that students might need more support to reduce their pause lengths than to produce longer P-bursts or more words per minute in their text production.

Additionally, while the final model explained 40.9 % of variance in writing scores, it did not explain over half the potential variance. The remaining variance may be related to other factors important for predicting writing scores, such as lexical and syntactic features, coherence, argumentation, critical thinking, and grammatical errors.

7. Conclusion

Overall, this study reports the relationships among cognitive and linguistic resources, writing processes, and writing products in undergraduate L2 writing. Findings from this study support the notion that linguistic and cognitive resources and writing processes as suggested in the Hayes and Berninger model (2014) are important for adult L2 writers to better produce argumentative essays.

Findings of this study should be interpreted alongside its limitations, which may offer indications for future studies. The limitations are provided as follows:

- This study was based on a cross-sectional design and correlational data. Thus, the basis for inferences was correlational and does not directly support casual inferences. To better examine causality among variables, research designs can include experimental designs in which the casual variable is a manipulated variable and longitudinal designs in which causal variables and outcomes are measured on more than one occasion over time (Kline, 2015).
- We focused on a single argumentative essay in a specific and a small group of undergraduate L2 students in a US university. Further research could replicate reported results by including various proficiency levels, different writing task types, different writing prompts, larger sample sizes, and different age groups.
- To calculate mean pause time and P-burst length, we set our pause threshold to two seconds, following previous research, but there might be different approaches to setting a threshold, such as individualized thresholds that consider each individual's typing speed, which a future study can adopt. Mean pause times and P-burst lengths can also be examined considering the temporal dimension of writing (e.g., writing processes captured at different writing stages) and different textual locations (e.g., between words and between paragraphs), and using different pause thresholds (Barkaoui, 2019; Michel et al., 2020).
- For writing-process-related measures, we included only two variables. Future studies can include other process-related measures, such as revisions and process variation to examine how they are related to resource- and product-level variables. In addition, when examining L2 writing process, we adopted a single approach of using keystroke logging data. Future study could use multiple data collection methods including keystroke logging, verbal protocols, and eye-tracking (Michel et al., 2020; Révész & Michel, 2019).
- Because the Stroop test and the simple running span test were measured in English, it may not be the most accurate description of WMC and attention for L2 learners. Alternatively, language-independent, non-verbal tests, such as the Corsi block-tapping test (Kessels, Van Zandvoort, Postma, Kappelle, & De Haan, 2000) and digit span tests, may be used to measure WMC. In addition, it would be of interest to investigate how WMC when measured by complex span tests (which require both cognitive processing and storage, such as operation span tasks; Linck et al., 2014) is associated with L2 writing.
- Instruments used in our study, including the Western-based general knowledge test, Gates-MacGinitie Reading Skill tests, and SAT-based prompts, are likely biased towards Western-based literature and achievements. In addition, we used a Western-based knowledge test, which may imply a false assumption that there is a unified set of general knowledge across different cultures.

The knowledge tested in our study also differs from academic knowledge that is considered important for L2 writers' success in higher education. Future studies can develop and use culture-neutral general knowledge tests or/and academic knowledge tests.
- Adopting a descriptive approach to exploring L2 writing, we collected data in a laboratory setting, which differs from a classroom-based setting. Future studies may use variables which were found to be related to the quality of L2 writing in this study, such as mean pause times and L2-literacy-related resources, as targets of intervention in classroom settings. Future studies can also examine L2 writing at the resource-, process-, and product-levels when L2 students have access to various resources (e.g., dictionaries) and produce multiple drafts in classroom settings.

Acknowledgements

This manuscript is based on the first author's dissertation research, which was funded by The International Research Foundation for English Language Education (TIRF), Educational Teaching Service (ETS), and Georgia State University.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.jslw.2021.100824.

References

Alves, R. A., & Limpo, T. (2015). Progress in written language bursts, pauses, transcription, and written composition across schooling. *Scientific Studies of Reading, 19*(5), 374–391. https://doi.org/10.1080/10888438.2015.1059838.

Alves, R. A., Castro, S. L., de Sousa, L., & Strömqvist, S. (2007). Influence of typing skill on pause-execution cycles in written composition. In M. Torrance, D. Galbraith, & L. Van Waes (Eds.), *Recents developpements in writing-process research* (pp. 55–65). Dordrecht, Netherlands: Kluwer. https://doi.org/10.1163/9781849508223_005.

Alwin, D. F., & Hauser, R. M. (1975). The decomposition of effects in path analysis. *American Sociological Review, 40*(1), 37–47. https://doi.org/10.2307/2094445.

Bachman, L., & Palmer, A. (1996). *Language testing in practice: Designing and developing useful language tests*. Oxford: Oxford University Press.

Barkaoui, K. (2019). What can L2 Writers' pausing behaviour tell us about their L2 writing processes? *Studies in Second Language Acquisition, 41*(3), 529–554. https://doi.org/10.1017/S027226311900010X.

Belcher, D. D., & Hirvela, A. (Eds.). (2001). *Linking literacies: Perspectives on L2 reading-writing connections.*. Ann Arbor, MI: University of Michigan Press.

Berninger, V. (2000). Development of language by hand and its connections to language by ear, mouth, and eye. *Topics in Language Disorders, 20*(4), 65–84. https://doi.org/10.1097/00011363-200020040-00007.

Carson, J. E., Carrell, P. L., Silberstein, S., Kroll, B., & Kuehn, P. (1990). Reading-writing relationships in first and second language. *TESOL Quarterly, 24*(2), 245–266. https://doi.org/10.2307/3586901.

Chenoweth, N., & Hayes, J. R. (2001). Fluency in writing: Generating text in L1 and L2. *Written Communication, 18*, 80–98. https://doi.org/10.1177/0741088301018001004.

Crossley, S. A., Yang, H. S., & Mcnamara, D. S. (2014). What's so simple about simplified texts? A computational and psycholinguistic investigation of text comprehension and text processing. *Reading in a Foreign Language, 26*(1), 92–113. https://scholarspace.manoa.hawaii.edu/handle/10125/66686.

Ellis, N. (1997). Vocabulary acquisition: Word structure, collocation, word-class, and meaning. In N. Schmitt, & M. McCarthy (Eds.), *Vocabulary: description acquisition and pedagogy* (pp. 122–139). Cambridge, UK: Cambridge University Press.

Ellis, R. (2003). *Task-based language learning and teaching*. Oxford: Oxford University Press.

Engle, R. W. (2002). Working memory capacity as executive attention. *Current Directions in Psychological Science, 11*(1), 19–23. https://doi.org/10.1111/1467-8721.00160.

Evans, N. W., Anderson, N. J., & Eggington, W. G. (Eds.). (2015). *ESL readers and writers in higher education: Understanding challenges, providing support*. New York, NY: Routledge.

Gardner, S., & Nesi, H. (2013). A classification of genre families in university student writing. *Applied Linguistics, 34*(1), 25–52. https://doi.org/10.1093/applin/ams024.

Gass, S., & Lee, J. (2011). Working memory capacity, inhibitory control, and proficiency in a second language. In M. Schmid, & W. Lowie (Eds.), *Modeling bilingualism: From structure to chaos* (pp. 59–84). Amsterdam, The Netherlands: John Benjamins. https://doi.org/10.1075/sibil.43.06gas.

Hambrick, D. Z., & Engle, R. W. (2002). Effects of domain knowledge, working memory capacity, and age on cognitive performance: An investigation of the knowledge-is-power hypothesis. *Cognitive Psychology, 44*(4), 339–387. https://doi.org/10.1006/cogp.2001.0769.

Hayes, J. R. (2009). From idea to text. In R. Beard, D. Myhill, J. Riley, & M. Nystrand (Eds.), *The SAGE handbook of writing development* (pp. 65–79). London, UK: Sage.

Hayes, J. R., & Berninger, V. W. (2014). Cognitive processes in writing: A framework. In B. Arfe, J. Dockrell, & V. W. Berninger (Eds.), *Writing development in children with hearing loss, dyslexia, or oral language problems* (pp. 3–15). Oxford, UK: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199827282.003.0001.

Hayes, J. R., & Chenoweth, N. A. (2007). Working memory in an editing task. *Written Communication, 24*(4), 283–294. https://doi.org/10.1177/0741088307304826.

Hayes, J., & Flower, L. (1980). Identifying the organization of writing processes. In L. W. Gregg, & E. R. Steinberg (Eds.), *Cognitive processes in writing* (pp. 3–30). Hillsdale, NJ: Erlbaum.

He, L., & Shi, L. (2012). Topical knowledge and ESL writing. *Language Testing, 29*(3), 443–464. https://doi.org/10.1177/0265532212436659.

Hulstijn, J. (2015). *Language proficiency in native and non-native speakers: Theory and research*. Amsterdam: John Benjamins.

Jeon, E. H., & Yamashita, J. (2014). L2 reading comprehension and its correlates: A meta-analysis. *Language Learning, 64*(1), 160–212. https://doi.org/10.1111/lang.12034.

Kaufer, D. S., Hayes, J. R., & Flower, L. (1986). Composing written sentences. *Research in the Teaching of English, 20*(2), 121–140.

Kellogg, R. (1996). A model of working memory in writing. In M. Levy, & S. Ransdell (Eds.), *The science of writing: Theories, methods, individual differences, and applications* (pp. 57–72). Mahwah, NJ: Lawrence Erlbaum Associates.

Kessels, R. P., Van Zandvoort, M. J., Postma, A., Kappelle, L. J., & De Haan, E. H. (2000). The Corsi block-tapping task: Standardization and normative data. *Applied Neuropsychology, 7*(4), 252–258. https://doi.org/10.1207/S15324826AN0704_8.

Kim, Y. S. G., & Park, S. H. (2019). Unpacking pathways using the direct and indirect effects model of writing (DIEW) and the contributions of higher order cognitive skills to writing. *Reading and Writing, 32*(5), 1319–1343. https://doi.org/10.1007/s11145-018-9913-y.

Kim, Y., Payant, C., & Pearson, P. (2015). The intersection of task-based interaction, task complexity, and working memory: L2 question development through recasts in a laboratory setting. *Studies in Second Language Acquisition, 37*(3), 549–581. https://doi.org/10.1017/S0272263114000618.

Kline, R. B. (2015). The mediation myth. *Basic and Applied Social Psychology, 37*(4), 202–213. https://doi.org/10.1080/01973533.2015.1049349.

Knoch, U., Rouhshad, A., Oon, S. P., & Storch, N. (2015). What happens to ESL students' writing after three years of study at an English medium university? *Journal of Second Language Writing, 28*, 39–52. https://doi.org/10.1016/j.jslw.2015.02.005.

Kormos, J. (1999). Monitoring and self repair in L2. *Language Learning, 49*(2), 303–342. https://doi.org/10.1111/0023-8333.00090.

Kormos, J. (2012). The role of individual differences in L2 writing. *Journal of Second Language Writing, 21*(4), 390–403. https://doi.org/10.1016/j.jslw.2012.09.003.

Kormos, J., & Sáfár, A. (2008). Phonological short term-memory, working memory and foreign language performance in intensive language learning. *Bilingualism Language and Cognition, 11*, 261–271. https://doi.org/10.1017/S1366728908003416.

Leijten, M., & Van Waes, L. (2013). Keystroke logging in writing research: Using inputlog to analyze and visualize writing processes. *Written Communication, 30*(3), 358–392. https://doi.org/10.1177/0741088313491692.

Limpo, T., & Alves, R. A. (2017). Written language bursts mediate the relationship between transcription skills and writing performance. *Written Communication, 34*(3), 306–332. https://doi.org/10.1177/0741088317714234.

Linck, J. A., Osthus, P., Koeth, J. T., & Bunting, M. F. (2014). Working memory and second language comprehension and production: A meta-analysis. *Psychonomic Bulletin & Review, 21*(4), 861–883. https://doi.org/10.3758/s13423-013-0565-2.

Ludwig, C., Borella, E., Tettamanti, M., & de Ribaupierre, A. (2010). Adult age differences in the Color Stroop Test: A comparison between an item-by-item and a blocked version. *Archives of Gerontology and Geriatrics, 51*(2), 135–142. https://doi.org/10.1016/j.archger.2009.09.040.

MacGinitie, W. H., MacGinitie, R. K., Maria, K., & Dreyer, L. G. (2002). Gates-MacGinitie Reading tests. *Technical report for Forms S & T* (fourth edition). Itasca, IL: Riverside.

Michel, M., Kormos, J., Brunfaut, T., & Ratajczak, M. (2019). The role of working memory in young second language learners' written performances. *Journal of Second Language Writing, 45*, 31–45. https://doi.org/10.1016/j.jslw.2019.03.002.

Michel, M., Révész, A., Lu, X., Kourtali, N. E., Lee, M., & Borges, L. (2020). Investigating L2 writing processes across independent and integrated tasks: A mixed-methods study. *Second language research*. https://doi.org/10.1177/0267658320915501. Advanced online publication.

Milton, J., Wade, J., & Hopkins, N. (2010). Aural word recognition and oral competence in a foreign language. In R. Chacón-Beltrán, C. Abello-Contesse, M. Torreblanca-López, & M. López-Jiménez (Eds.), *Further insights into non-native vocabulary teaching and learning* (pp. 83–98). Clevedon, UK: Multilingual Matters.

Pae, T.-I. (2019). A simultaneous analysis of relations between L1 and L2 skills in reading and writing. *Reading Research Quarterly, 54*(1), 109–124. https://doi.org/10.1002/rrq.216.

Plakans, L., & Gebril, A. (2017). Exploring the relationship of organization and connection with scores in integrated writing assessment. *Assessing Writing, 31*, 98–112. https://doi.org/10.1016/j.asw.2016.08.005.

Qualtrics Research Suite. (2016). *Computer software]. Provo, Utah: qualtrics.com*.

R Core Team. (2018). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

Raimes, A. (1987). Language proficiency, writing ability, and composing strategies: A study of ESL college student writers. *Language Learning, 37*(3), 439–468. https://doi.org/10.1111/j.1467-1770.1987.tb00579.x.

Révész, A., Michel, M., & Lee, M.-J. (2017). Investigating IELTS academic writing task 2 – Relationships between cognitive writing processes, text quality, and working memory. *Australia: British council, Cambridge English language assessment and IDP*. https://www.ielts.org/teaching-and-research/research-reports/ielts_online_rr_2017-3.

Methodological advances in investigating L2 writing processes [Special issue]. In Révész, A., & Michel, M. (Eds.), *Studies in second language acquisition, 41*, (2019). https://doi.org/10.1017/S0272263119000366, 3.

Roscoe, R. D., Crossley, S. A., Snow, E. L., Varner, L. K., & McNamara, D. S. (2014). Writing quality, knowledge, and comprehension correlates of human and automated essay scoring. In W. Eberle, & C. Boonthum-Denecke (Eds.), *Proceedings of the 27th International Florida Artificial Intelligence Research Society (FLAIRS) Conference* (pp. 393–398). Palo Alto, CA: AAAI Press.

Rosseel, Y. (2012). lavaan: an R package for structural equation modeling. *Journal of Statistical Software, 48*(2), 1–36. https://doi.org/10.18637/jss.v048.i02.

Roy, A., Kefi, M. Z., Bellaj, T., Fournet, N., Le Gall, D., & Roulin, J. L. (2018). The Stroop test: A developmental study in a French children sample aged 7 to 12 years. *Psychologie Francaise, 63*(2), 129–143. https://doi.org/10.1016/j.psfr.2016.08.001.

Sanders, A. (1998). *Elements of human performance*. Mahwah, NJ: Lawrence Erlbaum Associates.

Schneider, W., Eschman, A., & Zuccolotto, A. (2012). *E-prime user's guide*. Pittsburgh, PA: Pschology Software Tools, Inc.

Silva, T. (1993). Toward an understanding of the distinct nature of L2 writing: The ESL research and its implications. *TESOL Quarterly, 27*(4), 657–677. https://doi.org/10.2307/3587400.

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*, 433–454. https://doi.org/10.2307/40264477.

Spivey, N. N. (1990). Transforming texts: Constructive processes in reading and writing. *Written Communication, 7*(2), 256–287. https://doi.org/10.1177/0741088390007002004.

Stæhr, L. S. (2008). Vocabulary size and the skills of listening, reading and writing. *Language Learning Journal, 36*(2), 139–152. https://doi.org/10.1080/09571730802389975.

Taber, K. S. (2018). The use of Cronbach's alpha when developing and reporting research instruments in science education. *Research in Science Education, 48*(6), 1273–1296. https://doi.org/10.1007/s11165-016-9602-2.

Van Waes, L., & Leijten, M. (2015). Fluency in writing: A multidimensional perspective on writing fluency applied to L1 and L2. *Computers and Composition, 38*(Part A), 79–95. https://doi.org/10.1016/j.compcom.2015.09.012.

Williams, J. N. (2012). Working memory and SLA. In S. Gass, & A. Mackey (Eds.), *Handbook of second language acquisition* (pp. 427–441). New York: Routledge.

Williams, J. N., & Lovatt, P. (2003). Phonological memory and rule learning. *Language Learning, 53*(1), 67–121. https://doi.org/10.1111/1467-9922.00211.

Dr. Minkyung Kim is an applied linguist working as an assistant professor at Nagoya University of Commerce and Business, Japan. Her research focuses on second language (L2) reading and writing and language assessment.

Yu Tian, a PhD student in the Department of Applied Linguistics & ESL at Georgia State University. His primary research interests include second language writing, natural language processing and psycholinguistics. He is particularly interested in using psycholinguistic methods and natural language processing tools to study both the writing process and written products of second language learners.

Dr. Scott Crossley is a Professor of Applied Linguistics and Learning Sciences at Georgia State University. Professor Crossley's primary research focus is on natural language processing and the application of computational tools and machine learning algorithms in language learning, writing, and text comprehensibility. His main interest area is the development and use of natural language processing tools in assessing writing quality and text difficulty.