Writing in a non-alphabetic language using a keyboard: Behaviours, cognitive activities and text quality

Xiaojun Lu

Thesis submitted to University College London

for the degree of Doctor of Philosophy

April, 2020

Declaration

I, Xiaojun Lu confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Abstract

Cognitive processes in second language (L2) writing have received growing attention for the past three decades. Yet, few studies have examined the processes involved in non-alphabetic language writing. Jointly adopting Kellogg's (1996) and Rijlaarsdam and Van Den Bergh's (1996) writing models as the theoretical frameworks, this thesis investigates the separate and joint effects of writing stages, genre and proficiency on the cognitive processes in which L2 writers of a non-alphabetic language engage, and how writing behaviours predict L2 text quality.

Thirty-two L2 users of Chinese wrote two argumentative and two narrative essays using the Pinyin input method. Their keystrokes during writing were captured. All participants described their thoughts about pauses and revisions prompted by the recording of their last writing performance. A cloze test was used to establish participants' level of L2 proficiency. Data was also obtained from 32 first language (L1) users for a fuller picture of non-alphabetic language writing processes. Writing behaviours were analysed in terms of speed fluency, pausing and revision, and the stimulated recall comments were classified as planning, translation or monitoring. Text quality was determined via holistic rating.

Mixed-effects regressions and qualitative analyses yielded the following results. First, a small stage effect emerged for the cognitive processes in non-alphabetic language writing, with a slight difference between L1 and L2. Second, L2 writers engaged in different behaviours and cognitive activities when writing in different genres, while L1 writers only behaved differently. Third, L2 proficiency affected speed fluency and certain pausing and revision behaviours, and this effect tended to be mediated by stages. Finally, L2 text quality was predicted by speed fluency and pauses between smaller textual units, and these relationships were modulated by stages, genre and proficiency. L1 text quality was only predicted by pauses between larger textual units in certain stages; the strength of the relationships was genre-dependent.

Impact statement

The current thesis explores the cognitive processes involved in non-alphabetic language, that is, Chinese first and second language (L1, L2) writing, and how the writing processes predict text quality. The findings of the study shed some useful light on theory building and research methodology in writing studies, and carry some potential impacts on L2 writing instruction.

Firstly, this thesis provides novel insights to the existing theories and frameworks on cognitive writing processes which have been developed mainly based on data collected from alphabetic language writers. By investigating writing in Chinese that employs a logographic writing system, the evidence presented in the current thesis puts a clear emphasis on the process of transcription when modelling writing in non-alphabetic languages, as transcription is assumed to be automated and receives little attention in alphabetic language writing.

Apart from the implications for writing theories, the findings of the present study attest the appropriateness of employing multiple data collection methods and triangulating different data sources to examine writing processes. The joint use of quantitative (i.e. keystroke logging files) and qualitative data (i.e. stimulated recall protocols) in the thesis allows more valid and reliable interpretations about the cognitive processes in which writers engage.

Last but not least, the results of this lab-based study can, to some extent, inform L2 writing instruction. For example, the cognitive writing processes, revealed in the observable patterns of pausing and revision behaviours, tend to indicate the source of writing difficulties, which may help L2 writing instructors to decide what strategies to teach based on learners' needs. In addition, the findings on the relationship between writing processes and text quality may be useful in identifying the writing strategies that can potentially lead to more successful writing outcomes.

Acknowledgments

This thesis would not have been possible without the generous support of many people over the past four years.

First and foremost, I would like to express my deepest gratitude to my principle supervisor, Professor Andrea Révész, for her thoughtful suggestion and comments at every stage of my PhD project. She has taught me about both the traditional and more novel research methods for data collection for studies on writing, as well as how to conduct statistical analysis for my study, improve my academic writing in English, and clearly deliver presentations based on my research findings in conferences and seminars. Her encouragement, inspiration, guidance and generosity in sharing her expertise in second language acquisition (SLA) was invaluable, which from time to time made my journey through the doctoral study more enjoyable than painful.

I also wish to thank my subsidiary supervisor, Dr Ana Pellicer-Sánchez, for providing me with useful knowledge on the eye tracking methodology. Every time when I got stuck in the eye tracking lab, Ana was always the first person who I could turned to for a solution.

I am also thankful to all members and student members of the London Second Language Acquisition Research Forum for organising/attending the annual PhD conferences and departmental-level doctoral seminars. Many thanks for your challenging questions and constructive feedback that have pushed my project forward.

My special thanks go to the learners of L2 Chinese and the Chinese students who have participated in my study. I greatly appreciate their responsibility and patience for completing this time- and energy-consuming experiment. I am grateful to Mr Donghai Hou, Ms Yaxian Lang, Ms Jiayi Lin and Ms Mengning Xie for inviting me to the language exchanging events in the School of Oriental and African Studies to meet potential participants. A sincere note of thank also goes to Ms Tian Gan for advertising my research in various locations in London, and to Mr Taso Papadopoulos for circulating my project information online.

I would also like to thank China Scholarship Council for the generous financial support in the past four years, and the Institute of Education, University College London for the grant to the 2019 Annual Conference of American Association for Applied Linguistics.

Last but not least, I wish to acknowledge my beloved family. I am indebted to my mom and grandma for being away for long. I am grateful to them for their unwavering and unconditional support and love from the other side of the world. I would also like to extend my heartfelt gratitude to my dearest friends in both London and China. Thank you for your companionship, assistance and encouragement throughout this long journey. 谢谢你们！

Table of contents

List of tables

List of figures

Chapter 1 Introduction

1.1 Motivation and background of the study

I developed my interest in investigating writing, particularly in Chinese as a second language (L2), when teaching Chinese to speakers of other languages at a university in China. I was once assigned to run a selective module on writing for high-intermediate learners. Only seven students registered, with one of them never showing up again after the first two sessions. More disappointingly, a colleague of mine ended up cancelling the module on writing for low-intermediate learners due to the lack of students. During teaching, I noticed that my students were very reluctant to write, despite the fact that they took an active part in pre- and post-writing activities. Their negative attitude towards writing was not only associated with the typical difficulties that L2 writers normally encounter, such as formulating grammatically correct sentences or generating ideas, but also related to the hard time they had 'drawing' Chinese characters stroke by stroke on paper. As a result, it usually took them at least 45 minutes to produce a hand-written essay with 150 Chinese characters (equivalent to an English text with 75 to 100 words), even though they had been given time to pre-plan the ideas and vocabulary needed for writing. I tried various ways to help them overcome their difficulties and boost their motivation for writing, but my students showed little progress by the end of the module. This worried me, because producing efficient writing is an important skill to develop, especially in an L2 which is considered a critical and strategic foreign language in today's multilingual world.

I turned to existing literature for solutions to address my concerns; however, it seemed that little research had been done on writing in L2 Chinese (H. Liu & Ling, 2012). Most earlier studies concerning L2 Chinese writing are non-empirical, either introducing writing models and theories (e.g. Li, 2015; Xue, 2013), describing the pedagogy and materials used in writing courses (e.g. Feng, 2006; Ma, 2014) or commenting on how to teach L2 Chinese writing based on the author's personal experience (e.g. Chen, 2003; Tong, 1996). Empirical studies are scarce, with a majority of them focusing on texts written by L2 learners of Chinese. Among these studies, most are devoted to analysing errors in L2 Chinese written discourse (e.g. Y. Liu, 2007; Xing, 2016) and comparing the essays produced by L1 and L2 Chinese writers from a certain perspective, such as the degree of formality and elegance in writing (Mo, 2016) and the use of theme-rhyme structures (L. Wang & Liu, 2015). A small number of researchers have in addition looked into how L2 Chinese text quality is influenced by task differences, such as task complexity (Hu & Lu, 2016; Yuan, 2010), pre-task planning conditions (Fei, 2015; Liao, 2018) and the presentation of task requirements (S. Liu & Cao, 2015). Several other studies have examined L2 Chinese texts to validate the measures for linguistic complexity and accuracy in written Chinese (An, 2015; Jiang, 2013; Jin, 2007; Wu, 2016a, 2016b).

There is no doubt that studying the end products of writing tasks can provide useful information about writing performance, development and assessment. Nevertheless, knowledge of the processes in which writers engage can help to identify good writing strategies and diagnose difficulties in writing from another angle (Révész & Michel, 2019). Unfortunately, compared to L2 research on writing in alphabetic languages, such as English, much less is known about text production processes in L2 Chinese.

Since Flower and Hayes (1981) proposed their cognitive model of writing, there has been a growing interest in describing and understanding the processes involved in composing a text. So far, it has been generally agreed that a decent amount of effort is required to produce a meaningful piece of writing even in one's first language (L1). In the process of text production, writers do not only engage in the activities of translating the output of thoughts into linguistic form, but also considering rhetorical goals so as to meet genre requirements and the needs of the reader, and/or clarify their understanding of the topic through a synthesis of the content (e.g. Bereiter & Scardamalia, 1987; Galbraith, 1999). These activities are assumed to take place recursively within the writer's limited working memory capacity (Kellogg, 1996), resulting in constant interactions between cognitive activities, such as planning, translation and monitoring, throughout the writing process.

Apparently, writing a text in an L2 is even more demanding. While it requires little attention for L1 writers to produce linguistically accurate sentences, L2 writers, especially those with relatively lower proficiency in the target language, may have to allocate more cognitive resources to selecting proper lexical items, formulating grammatically correct clauses and using appropriate cohesive devices to enhance the transition between paragraphs. Besides, L2 writing tends to present greater challenges when the orthographic system of the L2 (e.g. a non-alphabetic logographic writing system like Chinese) shares little similarity to that of the writer's L1 (e.g. an alphabetic phonetic writing system like English). In particular, given the non-transparent sound and form mapping of symbols, L2 writers are likely to encounter difficulties when transcribing (i.e. handwriting or typing) the text, a process which is generally believed to be fully automated in adults' L1 and L2 writing. In addition, L2 writers may experience an increased cognitive load when involved in the process of re-reading for monitoring. As there is no direct link between the pronunciation and written form of a symbol, it can be cognitively demanding and time-consuming to read through a text comprising logographic symbols. This leaves an important gap in the literature concerning how a less familiar writing system may influence L2 writing processes.

In addition, although writing being a dynamic process is acknowledged in most theoretical frameworks (e.g. Bereiter & Scardamalia, 1987; Flower & Hayes, 1981; Galbraith, 1999; Kellogg, 1996), the common practice employed to investigate writing processes, and L2 writing processes in particular, relies on calculating the frequency and percentage of a writing activity for the whole writing session without accounting for the time at which a given activity occurs (Rijlaarsdam & Van Den Bergh, 1996). The small number of L2 empirical investigations that have considered the role of time in the writing processes indicate that, similar to L1 writers, L2 writers differ in the timing of their writing activities, suggesting the importance of including the variable of time when studying L2 writing processes (e.g. Barkaoui, 2016; Roca de Larios et al., 2008; Xu & Qi, 2017).

Apart from exploring the role of time in writing, there has been increasing attention paid to how the process of text production is associated with writer- and task-related factors. A few researchers have looked into the relationship between L2 proficiency and cognitive writing processes (e.g. Barkaoui, 2019; Chenoweth & Hayes, 2001; Tillema, 2012; Tiryakioglu et al., 2019). The findings show that higher- and lower-proficiency L2 writers do indeed engage in different cognitive processes during text production; however, little agreement has been reached upon whether and how these proficiency-related features are influenced by the variable of time in writing. Others have investigated how L2 writing processes may be affected by task-related factors, such as task type (e.g. Barkaoui, 2015), task complexity (e.g. Révész et al., 2017) and task condition (e.g. Khuder & Harwood, 2015). Surprisingly, however, limited effort has been devoted to examining the extent to which L2 writing processes are influenced by genre, a 'central and remarkably productive concept in L2 writing studies' (Tardy, 2011, p. 2). Previous L2 writing studies on genre have primarily focused on analysing certain aspects of written texts (e.g. lexical or grammatical units) to describe the differences in rhetorical patterns between genres (e.g. Lu, 2011; Yoon & Polio, 2016).

Another area that has started to attract attention from L2 writing researchers is the relationship between writing processes and text quality. Studying this link can provide useful knowledge about what leads to successful writing, which has potential value for L2 writing instruction. Although process-product relations have generated a substantial amount of literature in the field of L1 writing (e.g. C. Beauvais et al., 2011; Deane, 2014), much less is known about this link in L2 writing, despite some initial attempts (e.g. Spelman Miller et al., 2008; Stevenson et al., 2006). Furthermore, empirical evidence about L1 writing also shows that text quality does not only relate to how frequently a writing activity occurs in the text production process, it also depends on the point in the writing process where the activity takes place (e.g. Rijlaarsdam & Van Den Bergh, 1996). However, it was not until recently that the temporal dimension was taken into account when studying the relationship between writing processes and L2 text quality (e.g. Gánem-Gutiérrez & Gilmore, 2018; Tillema, 2012; Van Weijen, 2008; Xu, 2018).

1.2 Focus of the study

Against this background, the main goal of this thesis is to expand existing L2 writing research by examining the processes in which L2 writers engage when composing in a non-alphabetic language. The focus will be on L2 writers of Chinese, to date an understudied population. More specifically, the thesis aims to investigate the behaviours (i.e. speed fluency, pausing and revision) and associated cognitive activities of writers of Chinese when performing argumentative and narrative writing tasks. Unlike the majority of earlier work on L2 Chinese, the present study does not only conceptualise writing as a dynamic process, but also analyses the process from a dynamic perspective by taking into account how writing behaviours and the cognitive activities associated with them may change across different stages of writing (e.g. beginning, middle, end). The current study also attempts to look at how genre and L2 proficiency might influence the behaviours and associated cognitive activities of L2

Chinese writers through the whole of, and at different stages of, the writing process.

This thesis also aims to explore the extent to which writing behaviours predict the quality of writing outcomes in L2 Chinese. It takes another step forward by investigating process-product links in relation to the stages of writing, genre and writers' L2 proficiency to see how contextual and individual factors may modulate the strength of these relationships. The relationship between writing process and L2 text quality is under-researched but of great importance to L2 writing. Understanding the links between process and product can yield some insights into how visible behaviour patterns (e.g. pausing and revision), which are assumed to reflect cognitive processes, may improve text quality. Besides, examining the influence of contextual and individual factors on process-product relations can provide more specific information about the effective cognitive processes of a certain group of L2 writers when performing a certain type of writing task. This may inform L2 instructors of strategies that are beneficial to teach.

The current study also makes some methodological contributions. It involves the joint use of quantitative and qualitative data sources, which allows for generating more valid and reliable conclusions on L2 Chinese writers' cognitive processes during text production. Quantitative data on writers' behaviours during task performance were collected using a keystroke logging technique, while qualitative data on writers' cognitive activities associated with writing behaviours were gathered via a stimulated recall procedure.

Finally, it is worth highlighting that given the limited amount of research available on writing in Chinese, this study additionally uses data collected from L1 writers of Chinese as a baseline to obtain a more comprehensive understanding of the cognitive processes and process-product relationships involved in non-alphabetic language writing. However, it should be noted that the purpose of including data from L1 writers is to determine if the observed trends are unique to writing in L2 as compared to L1 Chinese. In other words, I am interested in whether the temporal distribution of writing activities, genre differences and process-product links is similar in L1 and L2 writing, rather than making direct comparisons of writing behaviour measures between the two groups.

1.3 Overview of the thesis

The remainder of the thesis is organised into five chapters. Chapter 2 conducts a review of relevant literature. It starts with a detailed description of the most influential cognitive models of writing, followed by a critical summary of previous empirical studies on the role of time, genre and L2 proficiency in the processes and process-product relationships involved in L2 writing. Next, it examines the methodologies that have been utilised to explore L2 writing processes. The final part of the chapter identifies research gaps and presents the specific research questions of the study. Chapter 3 describes the research methodology employed to collect and analyse data. In this chapter, the overall design of the study is first described, followed by an introduction to the participants and the instruments that were used to address the research questions. The chapter ends with an explanation of the coding procedures and statistical analyses. Chapter 4 presents the results of the study. First, it reports the results obtained from quantitative analyses, followed by the findings of qualitative analyses. Chapter 5 discusses the quantitative and qualitative results in light of the research questions and existing literature. Finally, in Chapter 6, the main findings of the thesis are summarised, and the implications of the study for theory, methodology and pedagogy are considered. The chapter concludes with a careful consideration of the limitations of the current study and points out some possible directions for future research.

Chapter 2 Review of the literature

Chapter 2 is dedicated to a review of the theoretical and empirical work on the cognitive process of writing. This chapter begins with an overview of several cognitive models of writing, including the earliest model on writing process of Flower and Hayes (1981) and its revised versions (Hayes, 1996, 2012), Scardamalia and Bereiter's (1987) knowledge-telling and knowledge-transforming models of writing, Kellogg's (1996) model of writing process, Galbraith's (1999, 2009a, 2009b) dual-process model of writing and Rijlaarsdam and Van den Bergh’s (1996) temporal model of writing. The next section of the chapter provides a critical review of previous empirical studies with a main focus on how stage and genre influence second language (L2) writing processes as well as the process-product relationships in L2 writing. The section that follows introduces the methods used in previous studies to elicit writers' cognitive processes during writing; the strengths and limitations of these methods are also discussed in this section. Finally, the rationale and the research questions of the study are presented.

2.1 Theoretical background: cognitive models of the writing process

This section first provides a detailed description of the most influential cognitive models of the writing process. Given that these models have generated substantial empirical work, the description is followed by a brief summary of the key empirical findings in relation to the models in first and second language (L1, L2) writing. Although some models, such as Galbraith's (1999, 2009a, 2009b) dual-process model of writing, may not be explicitly incorporated in the rest part of the thesis, a discussion of these models are still included in this section in order to offer the general context in which L1 and L2 writing research has been built upon. Based on the review, I will explain my choice of the theoretical framework of the thesis at the end of the section.

2.1.1 Flower and Hayes' model of the writing process

Perhaps the best-known model in L1 writing process-oriented research is Flower and Hayes' cognitive model of the writing process (Flower & Hayes, 1981). This model was established through analysing the writers' think-aloud protocols by using a coding scheme developed by Hayes and Flower's (1980) for categorising mental activities involved in writing. In the model, writing activities are grouped into three central elements as illustrated in Figure 2.1.

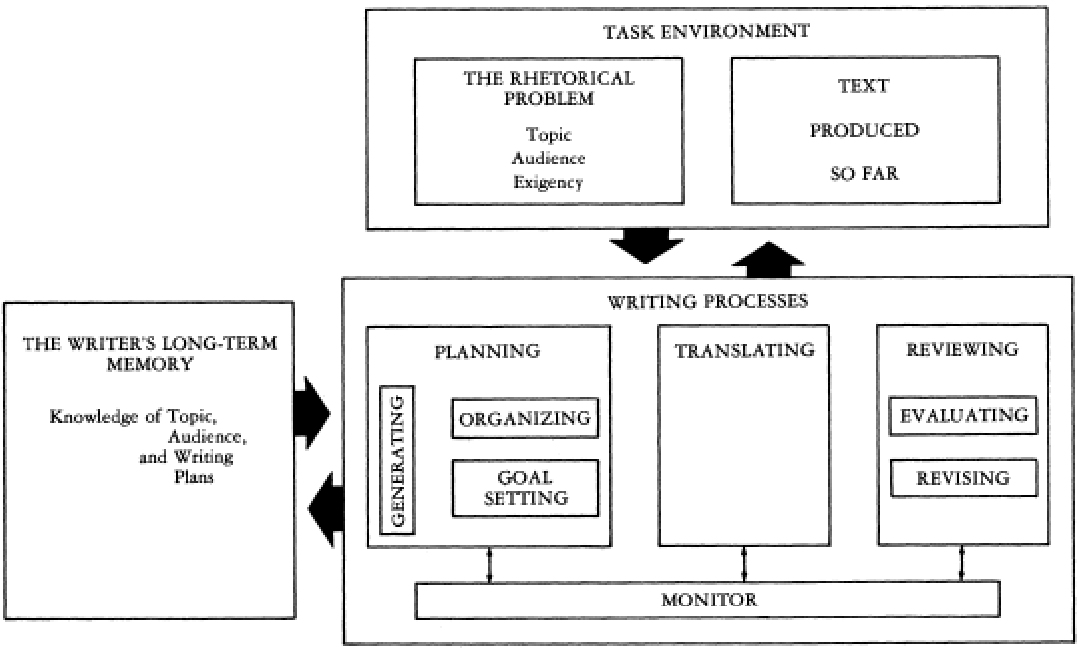

Figure 2.1 Flower and Hayes' model (1981)

The first one is the writer's long-term memory. It stores the writer's knowledge, such as knowledge of the topic, audience, and writing plans. Different from short-term memory, long-term memory is mostly stable and has its ‘internal organisation of information’ (p. 371). According to Flower and Hayes, there are two ways in which long-term memory may constrain writing processes. Firstly, it can be demanding to find the cue for the retrieval of useful knowledge. The second one relates to the need to reorganise or adapt the retrieved information in order to meet the rhetorical requirements.

The second element, the task environment, includes ‘all of those things outside the writer’s skin’ (p. 369). The rhetorical problem is the most important constituent at the beginning of writing. It encompasses the intended topic, audience and motivating cues. As writing proceeds, another constituent, the text-produced-so-far, enters the task environment. The growing text directs the process of writing by interacting with the writer’s knowledge in long-term memory and the way that the rhetorical problem is defined.

The third element, which is also the most-discussed element in Flower and Hayes’ work, includes the writing processes. Writing processes are comprised of three basic cognitive activities, planning, translating and reviewing, which are hierarchically structured and contain sub-processes. Planning is ‘the act of building an internal representation of the knowledge’ (p. 372) to be used for writing. Three sub-processes, generating ideas, organising ideas, and goal setting, is associated with the act of planning. The sub-process of generating ideas includes the retrieval of information from the writer’s long term memory. The generated ideas can be either in forms of standard written languages, fragmented thoughts, or even non-linguistic symbol systems, such as images and feelings. Organising ideas is a stage in which the writer establishes meaningful structures of the ideas by identifying categories and searching for subordinate and superordinate ideas of the given topic. In this stage, the writer also makes decisions on the representation and ordering of the text guided by the third sub-process of planning, goal setting. Flower and Hayes suggest that the writer creates the goals by bearing the need of the potential readers in mind. The goals are not static; they can be developed and revised when new ideas are retrieved and organised. Thus, the act of goal setting takes place throughout the process of writing for integrating the purpose and content. The process of translating is referred to as an activity of 'putting ideas into visible language' (p. 373). In other words, translating is responsible for converting the ideas, which are generated in planning and represented in various symbol systems, into a linear piece of written text with elaborate syntax. Reviewing involves two sub-processes, evaluating and revising. The writer reads what has been produced so far and evaluates whether or not the text satisfies the rhetorical requirements. Revising can take place when an error or anything contracts the intended goals is detected. Additionally, it is suggested that evaluating and revising have a special feature of interrupting any other processes in writing. The writer can evaluate and revise when engaging in the process of planning. It means the unwritten thoughts or statements undergo reviewing as well. These writing processes are coordinated by the monitor, referred to by Flower and Hayes as 'a writing strategist' (p. 374). The monitor is responsible for determining when the writer shifts from generating ideas to translating, and when reviewing is needed, etc.

In Hayes' (1996) revised model, the three components are replaced by two components, the individual environment and the task environment. The individual environment consists of two initially separated elements in the 1981 model, long-term memory and writing processes, as well as two new elements, the writer's motivation and working memory. The task environment contains the physical environment (including the text produced so far and the composing medium) and the social environment (including the audience and collaborators). There are three major differences between the old model and the new model. Firstly, the new model highlights the role of working memory and places emphases on the motivational and affective factors in the writing process. Secondly, the new model includes the visual-spatial features of the written texts. Finally, the cognitive activities are reorganised into more general categories. Planning is subsumed under reflection; translation becomes a component of text production; reviewing is replaced by text interpretation.

The latest modification of the 1981 model was made by Hayes in 2012. In the model (Hayes, 2012), the writing process is made up of four major elements, the proposer, the translator, the transcriber and the evaluator. The proposer is responsible for the generation of ideas. These ideas then enter the translator and are converted into linguistic forms. The stage, in which linguistic forms are put down on the paper or the screen, is monitored by the transcriber. The evaluator plays a role in all three stages to detect errors and initiate revision. In other words, revision applies not only to the text that has been written down but also to the retrieved ideas and translated language strings. The writing process interacts with the task environment which contains collaborators and critics, transcribing technology, task material and written plans, and the text-written-so-far. The task environment and the writing process are integrated into a larger component, the process level, which is regulated by the control level (consisting of motivation, goals, current plans and writing schemas), and supported by the resource level (constituted by attention, long-term memory, working memory and reading).

The most important feature of these models is the stress on the non-linearity and dynamism of the writing process in which cognitive activities present recursive interplay with one another (Whalen, 1993). In other words, cognitive activities can 'occur at any time' (Flower & Hayes, 1981, p. 367) during the process of writing and the occurrence is 'not fixed in a rigid order' (p. 375). This feature differentiates the models from the traditional perception that views writing as product-based, which presents a three-staged linear process from pre-writing, writing to rewriting (e.g. Rohman, 1965). The differences between individuals in the writing processes are ascribed to 'the different configurations of production rules' (Galbraith, 2009a, p. 9) mediated by the knowledge stored in long-term memory, the deployment of the knowledge in accordance with the task environment, and the writer's motivation and affect of writing. Flower and Hayes' research has also established the most frequently used terms in the process-oriented writing research, especially in studies aimed at portraying a general picture of writing processes or investigating certain macro processes and their sub-processes (Roca de Larios et al., 2002). The models have enabled a characterisation of differences between more and less skilled writers (Galbraith, 2009a; Roca de Larios et al., 2016) in empirical studies.

Empirical evidence (Hayes & Flower, 1986) showed that expert writers generated a far more elaborate network of goals than their novice counterparts in terms of the process of planning. They modified and even restructured the network during text production through responses to new ideas and the text-produced-so-far. Skilled writers also selected and constructed ideas with the guidance of the rhetorical goals of the whole text, whereas unskilled writers were more likely to simply state what they knew about the topic and generated relevant ideas. In regard to the reviewing process, experienced writers, in general, were more able to identify and solve problems than novice writers, and tended to address both global and local problems of the text. However, novice writers were more likely to treat the process of reviewing as a sentence-level task in which attention was largely paid to whether the words and phrases were appropriately used (Hayes, 1996; Hayes et al., 1987).

Many L2 writing researchers also drew on these models (Flower & Hayes, 1981; Hayes, 1996, 2012) to describe the distinction between skilled and unskilled writers, and consistent findings have been reported across a number of studies (e.g. Cumming, 1989; Raimes, 1985, 1987; Skibniewski, 1988; Victori, 1999). It was revealed that, similar to that in L1 writing, experienced L2 writers planned at both structural and linguistic levels of the text, and usually spent a longer time in planning than less skilled writers. They also revised more effectively and made a better balance between global-level and local-level revisions. On the contrary, less experienced L2 writers planned less. They also had less reflection and modification on what had been planned while writing (Cumming, 1989). Besides, they reviewed less with most revisions related to the local level, especially linguistic concerns of the text.

2.1.2 Scardamalia and Bereiter's model of the writing process

The distinction between expert and novice writers is more explicitly addressed in Scardamalia and Bereiter's (1987) model of writing, in which the knowledge-telling and knowledge-transforming dichotomy of text production is proposed. The two approaches are contrasted in terms of 'how knowledge is brought into the writing process' and 'what happens to knowledge in that process' (p. 143).

In the knowledge-telling model (Figure 2.2), the writing process of novice writers is described as a think-say or what-next process. Novices retrieve tend-to-be relevant information from long-term memory guided by topic and genre identifiers according to the assignment representation, and then write it down directly onto the paper. The retrieved information provides writers with additional cues that serve as bases for further searches in long-term memory, which assists in increasing the coherence of the text. However, the increased coherence results from neither deliberate nor conscious application of content and discourse knowledge. In other words, novice writers rely on a set of activities that do not require much engagement of cognitive effort. As a result, texts produced by novice writers fit the topical and genre requirements of the writing task but only at the superficial level.

By contrast, in the knowledge-transforming model (Figure 2.3), expert writers are involved in a far more complex problem-solving process resulting from a dialectical interaction between the content problem space and the rhetorical problem space. The transformation of knowledge takes place when 'the problem encountered in the rhetorical space is translated into sub-goals to be achieved in the content space and vice versa' (p. 147). Although the actual process of information retrieval is much similar to the knowledge-telling process, the expert also constantly addresses the constraints set out by analysing the rhetorical problems during writing, which leads to the elaboration of sub-goals both in terms of content and rhetoric. Therefore, writers engaging in the knowledge-transforming process are assumed to produce texts much tailored to the readers' need, since the texts are not a mere reflection of the writers' knowledge in long-term memory but also an adaptation of their thoughts to the communicative goals (Galbraith, 2009a).

The contrast between knowledge-telling and knowledge-transforming approaches to writing has informed the interpretation of a number of differences between novice and expert writers. As suggested by the knowledge-telling model, novice writers start writing immediately when a first item retrieved from long-term memory meets the requirements concerning the topic and the genre. In contrast, the planning process of expert writers is more complex. By re-analysing data from Burtis et al. (1983), Bereiter and Scardamalia (1987) found that less skilled writers did little overall planning and goal setting, and their mental representation largely resembled what they actually wrote down onto the paper. As for expert writers, the think-aloud protocols of the planning process included not only the part that appeared in the written products but also an interweaving network of goal statement, idea generating, problem-solving strategies, comments and so forth. These findings are similar to what Hayes and Flower (1986) described in their work.

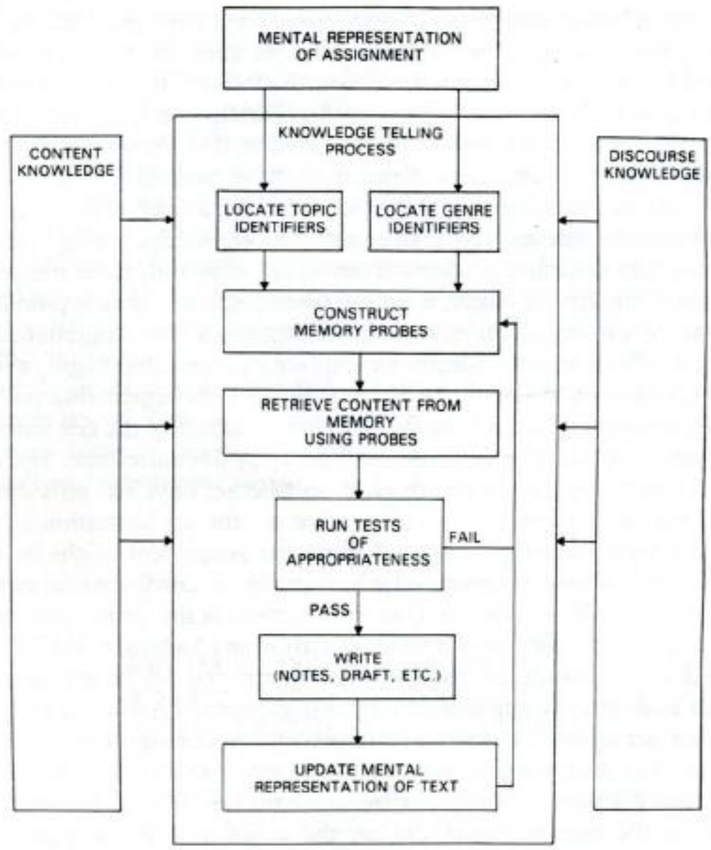

Figure 2.2 Scardamalia and Bereiter's knowledge-telling model (1987)

Another typical feature of less skilled writers reported in an empirical study conducted by Bereiter and Scardamalia (1987) is that they revise less. This has also been noted by researchers utilising Flower and Hayes' (1981) model. However, Scardamalia and Bereiter (1987) further argued that even though pedagogical interventions could increase the chances of revision, the revision process of novice writers hardly involved adjustments to goals and main ideas, and resulted in no real transcending of the knowledge. Unlike novice writers, expert writers continuously revised and rethought owing to the interplay between the content problem space and the rhetorical problem space during the process of writing.

Consequently, the text produced by more and less skilled writers differs in various ways. In the written texts by the novice, sentences were merely stuck to the topic, whereas in the texts by the expert, there was a more complex linking of ideas (Mccutchen & Perfetti, 1982). Furthermore, although the texts by non-expert writers might meet the structural requirements of the certain literary type, the texts were not very likely to succeed in achieving the rhetorical goal of that type. In addition, expert writers recognised information according to the need of readers, while novice writing was more like ‘writer-based prose’ (Flower & Hayes, 1981, p. 371), which simply recorded the writer’s flow of thoughts.

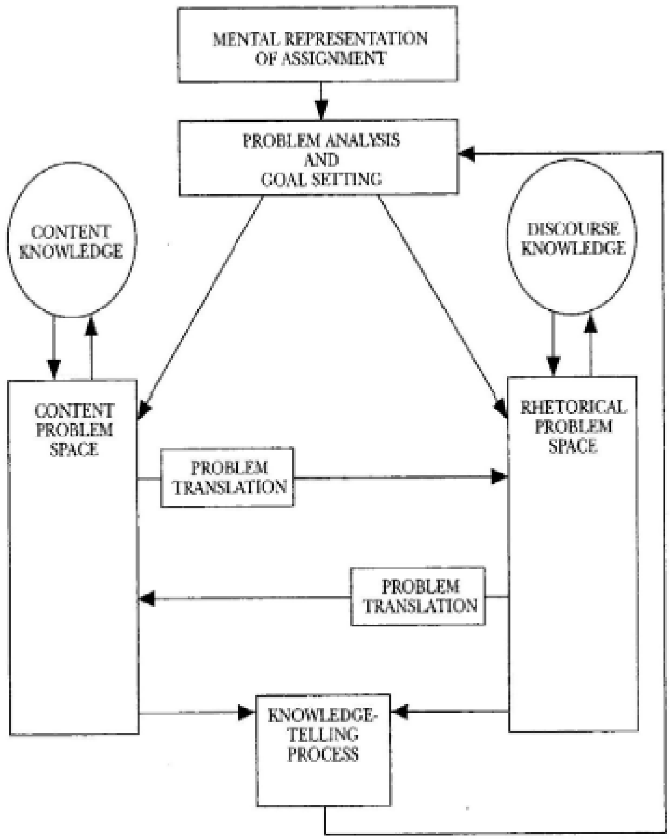

Figure 2.3 Scardamalia and Bereiter’s knowledge-transforming model (1987)

The distinctive features of the novice and the expert assumed by the models of

Scardamalia and Bereiter (1987) have been reported by some L2 writing researchers. For instance, Brook (1985, cited from Krapels, 1990) found that unskilled L2 writers carried out planning and translating simultaneously. Danzak (2011) noted that the writing process of L2 English learners at the secondary level appeared to demonstrate features of the knowledge-telling model. The learners activated knowledge relevant to the topic and the genre by simply taking cues from the prompt of the writing task, and wrote into the text almost everything they retrieved about the given topic. Hirose and Sasaki (1994) suggested that the revision process of good L2 writers was more goal-directed, and took place at various levels, while bad writers only revised at the sentence level. The findings of a comparative case study on two English L2 writers by Cumming (1995) showed that the Arab writer is a knowledge teller, and the French writer is a knowledge transformer. The former did not ‘refine the knowledge, use it to achieve new goals, or transform his thinking’ (p. 379), whereas the latter constantly engaged in solving mental problems which she considered as a way to clarify her own thinking.

It should be noted that, similar to Flower and Hayes' (1981) model, Scardamalia and Bereiter's (1987) model emphasises higher-order thinking processes in writing, such as generating goals and ideas. Consequently, differences between more and less skilled writers are mainly captured in terms of goal setting and content generating activities. On the other hand, both Flower and Hayes' (1981) and Scardamalia and Bereiter's (1987) models provide little information about lower-order thinking processes, such as linguistic encoding, and how lower-order processing may affect higher-order processing in text production. However, there has been accumulated empirical evidence showing that converting ideas into written messages can be cognitively demanding, especially in L2 writing (e.g. Roca de Larios et al., 1999). This issue has been addressed in the ensuing writing models, such as previously described Hayes' 1996 and 2012 writing models (Hayes, 1996, 2012), and Kellogg's model of writing which is discussed in the next section.

2.1.3 Kellogg's model of the writing process

Kellogg's (1996) model of the writing process stems from the assumption that speaking and writing involve common processes. It derives from Levelt's (1989) model of speech production in which the speaking process is seen as consisting of the sub-processes of conceptualising, formulating and articulating, and is coordinated by a monitor. In Kellogg's cognitive model (Figure 2.4), the writing process is defined into three basic systems: formulation, execution and monitoring. Each system is comprised of two sub-processes. Formulation includes planning the goals and ideas (equivalent to Levelt's conceptualising) and translating the ideas into written messages (corresponding to Levelt's formulating). Although translating is the stage where the output of planning is converted into linguistic forms, the output of planning already contains partially translated linguistic units with sketchy syntactic structures. Execution (similar to Levelt's articulating) consists of programming and executing. This is the stage in which complex muscle movements (i.e. handwriting and typing) are activated for text production. Reading and editing are sub-processes of monitoring.

Reading entails the processes of word recognition, sentences comprehension, coherence establishing and global discourse structure building. Editing takes place when mismatches between the writer's intention and the output are detected and diagnosed. The three basic systems interact and can be activated simultaneously as long as the demands placed on cognitive mechanism can be handled against working memory limitations.

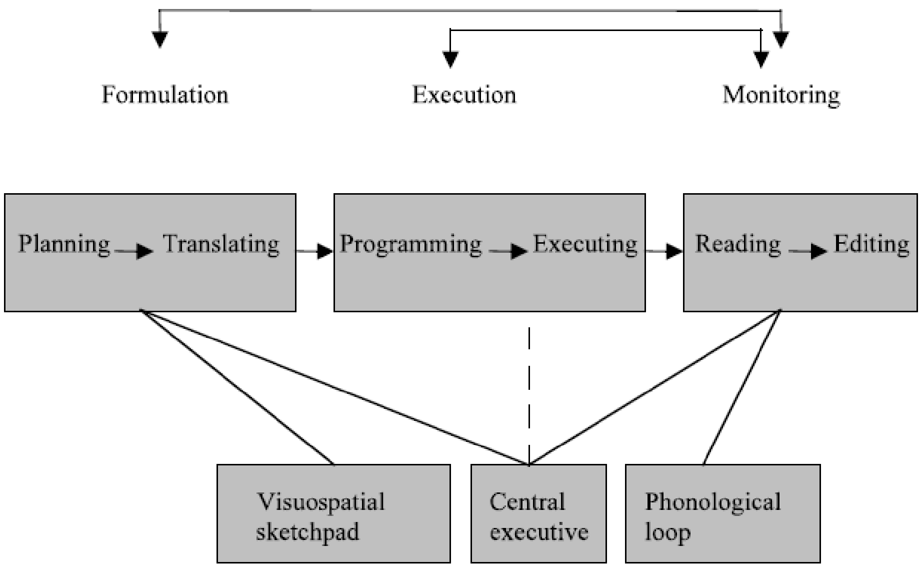

Figure 2.4 Kellogg's model (1996)

Additionally, Kellogg relates the process of writing to the multi-component model of working memory proposed by Baddeley (1986, 1992) and Baddeley and Hitch (1974). According to Baddeley's model[1], working memory contains three components, the central executive and two slave systems, the phonological loop and the visuospatial sketchpad. The central executive is an attentional control system responsible for coordinating information from the slave systems. The phonological

[1] In the latest version of working memory, a fourth component, the episodic buffer, was added (Baddeley, 2010). The episodic buffer provides a temporary information store in which other components of working memory can interact. It also serves as an interface between working memory components and long-term memory.

loop and the visuospatial sketchpad are two short-term storage systems for verbal-acoustic materials and visual materials respectively (see also in Baddeley, 2010).

In his model of writing, Kellogg specifies the demands that formulation, execution and monitoring place on working memory. Formulation places the highest burden on working memory as it uses all three components. The sub-process of planning makes use of both the visuospatial sketchpad and the central executive. Visual components can be involved in retrieving ideas, particularly those related to concrete items, and spatial components can be engaged when writers are organising ideas. Planning also loads the central executive because generating ideas, debating the proper tone for the intended audience, and establishing an appropriate structure for the text all invoke high levels of mental activities. Translating demands the phonological loop. Phonological components can be activated when writers are converting ideas into linguistic forms using the inner speech. If the writer is having trouble finding the proper words or expressions, translating can also place relatively high demands on the central executive.

Execution requires the least demands on working memory since programming demands little on the central executive, and executing does not make use of working memory components. However, this assumption does not apply to young children who have not yet achieved full automaticity in handwriting or typing (Bourdin & Fayol, 1994, 1996) or writers who are learning to write in an additional language with a different writing system from their L1 (Ellis & Yuan, 2004).

As for monitoring, reading requires resources of both the phonological loop and the central executive. Editing draws only on the central executive, but it places heavier demands than reading on working memory. It is possibly because ‘there are so many ways to make mistakes’ (Kellogg, 1999, p. 47). Therefore, editing can occur at various levels, ranging from correcting a misspelt word to detecting a problem in the text structure. Another reason that makes reading less demanding than editing is that reading self-generated text ‘seems to be greased’ (Kellogg, 1996, p. 65) due to the preceding processes of planning and translating.

Empirical studies in both L1 and L2 writing have lent some support to the role of working memory components in text production. It was reported that writers who were faced with dual-task demands performed poorer in text production. Articulatory suppression, such as irrelevant speech or simultaneous articulation, which added demands on the phonological loop during composition, was found to more or less disrupt the fluency and/or the quality of writing (Chenoweth & Hayes, 2003; Levy & Marek, 1999; Madigan, Johnson, & Linton, 1994). Competing tasks on verbal working memory first impeded the speed of writing, and only impacted the quality of writing when larger secondary task demands in working memory were required (Ransdell et al., 2002). The findings implied that the verbal system of working memory, the phonological loop, was involved in translating ideas into sentences.

As for the role of visual working memory, it has been revealed that demands that the writing process places on visuospatial sketchpad are probably selective. Passerault and Dinet (2000) observed that a decrease in writing fluency occurred due to a concurrent visual task when writers were carrying out a descriptive writing task but not when writing an argumentation. Similarly, Kellogg, Olive, and Piolat (2007) found that the disruption in fluency resulted from tasks involving image-evoking words. Thus, the researchers concluded that visual working memory played a 'selective role' (p. 108) which implicated in unimpeded text generation when images and abstract propositions were activated by the conceptual representation.

As the model also emphasises the limited capacity of the central executive, some studies also looked into the effects of attentional trade-off in the writing process. Bourdin and Fayol (1994, 1996) observed a trade-off between formulation and execution systems in children. Brown et al. (1988) reported that capacity allocation in the writing process of adult writers was also asymmetric. Priority was given to formulation over execution and monitoring when the cognitive mechanism was overloaded. As for L2 writers, due to limited L2 proficiency, attention was often paid to translating ideas into written messages at the expense of planning and revising (Roca de Larios et al., 2008).

Strategies that may help to reduce cognitive overload have also been studied. Kellogg (1990, 1988) found that outlining reduced cognitive demands, enabling writers to devote more capacity of the central executive to translating, thereby enhancing text quality. Consistent with Kellogg's findings, Ellis and Yuan (2004) also suggested that planning could be helpful in improving fluency and complexity for L2 writers.

Some researchers have compared the writing behaviours and/or outcomes of writers carrying out writing tasks with different cognitive demands. Probably owing to the limitation of working memory capacity, increasing task complexity was found to result in poorer overall text quality (Ong & Zhang, 2013), more cognitive resources dedicated to metacognitive processes (Ong, 2014), a decline in the number of ideas generated (Glynn et al., 1982), and more pausing but less revision (Révész et al., 2017).

2.1.4 Galbraith's model of the writing process

Different from most classic models (e.g. Bereiter & Scardamalia, 1987; Flower & Hayes, 1981) in which writing is defined as a top-down, problem-sloving process, Galbraith's (1999, 2009a, 2009b) model describes writing as a discovery, which includes two types of processes: an explicit planning process and an implicit text production processes. As in other writing models, the explicit planning process involves the retrieval of knowledge from long-term memory. The process takes place in forms of either knowledge telling or knowledge transforming (Bereiter & Scardamalia, 1987). This process only results in a reorganisation and/or a selection of knowledge in working memory to satisfy the rhetorical goal of the writing task. However, the implicit text production process, also referred to as the knowledge-constituting process, involves the synthesis of knowledge, leading to the spontaneous emergence of ideas and the development of the understanding of the topic during the writing process. Figure 2.5 illustrates a brief description of the knowledge-constituting process (Galbraith, 1999).

The mechanism underpinning the knowledge-constituting process is the writer's disposition, described as a fixed set of mental units interconnected with each other either positively or negatively. The units are activated when the writer receives input constraints consisting of the topic and task specifications. The activation is passed through the units until the network reaches a stable pattern. Ideas emerge as a result of the pattern of activation. The idea is then passed to the linguistic network (arrow labelled A) and is formulated into a written utterance (boxes labelled B). After the utterance has been output, inhibitory activation is provided to the writer's disposition (arrows labelled C) as the result of feedback connections of the output represented in working memory. Therefore, a different pattern of activation emerges when the topic and task specifications are input again, leading to the formation of a new utterance (arrow labelled D) which serves as feedback for the further circles of text composition (arrows labelled E and F). The knowledge-constituting process implies that writing also involves an internal dialectic between 'the writer's disposition and the emerging text' (Galbraith, 2009b, p. 21) which is not restricted by explicit goal-directed planning.

A series of studies conducted by Galbraith and his colleagues (Galbraith, 1992, 1999; Galbraith et al., 2006) have offered empirical evidence to the dual-process model. Two types of participants, high self-monitored (HSM) and low self-monitored (LSM), were examined in these studies. The HSM writers were assumed more likely to tailor the writing to satisfy the rhetorical requirements, while the LSM writers were more likely to express ideas directed by dispositional goals. Galbraith (1992) found that LSM writers generated twice as many new ideas as the HSM writers during writing, although the HSM group produced more ideas when making notes. The number of new ideas produced was positively correlated with how much the writers felt they knew about the topic in the LSM group but not in the other group. Based on the findings, Galbraith concluded that there might be two different processes responsible for idea generation and text production, and the idea generation controlled from within was associated with the development of writer's personal understanding of the topic.

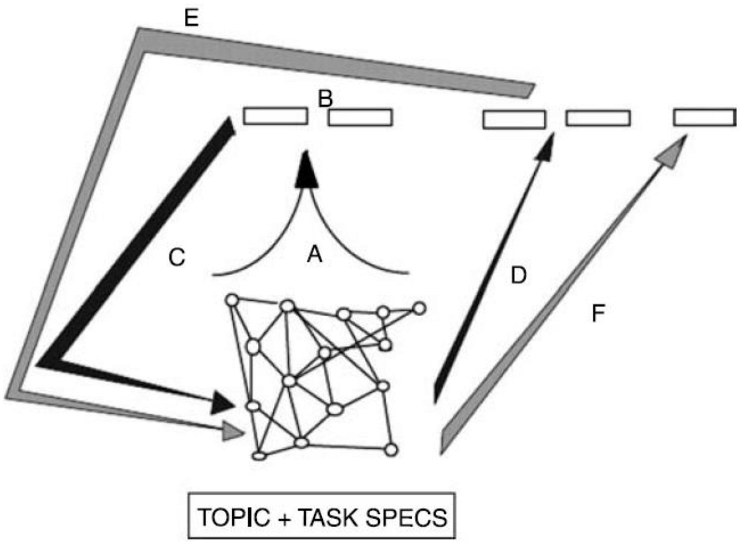

Figure 2.5 Knowledge-constituting process (Galbraith, 1999)

The findings were confirmed by an experiment (Galbraith, 1999) where the two types of writers were asked to write under three conditions: unplanned, synthetically planned, and outline planned. Galbraith (1999) also investigated changes in the organisation of thoughts. It was revealed that HSM group experienced more changes in organisation. There was an association between new ideas produced and changes in the organisation of the LSM writers under the outlined planning condition where explicit planning was dominant. However, under the unplanned and synthetically planned conditions in which less explicit planning was involved, new ideas were more associated with the increase in knowledge. The results once again confirmed the assumption of the dual-process in writing by further suggesting that explicit planning had a greater influence on the organisation of thoughts, while implicit text production played a more important role in developing understanding.

In a more recent study, apart from confirming the findings of previous research, Galbraith, Torrance, and Hallam (2006) also suggested that there was a significant reduction in the coherence of ideas in HSM group under planned conditions but not in other groups. The authors owed the difference to the two distinct processes of idea generating and suggested that dispositionally-produced ideas could be more conceptually coherent as they emerged during the course of text production and interrelated with the writer's understanding of the topic.

Although most empirical evidence supporting the dual-process model is from the field of L1 writing, Galbraith (2009) proposed two implications derived from the model for L2 writing. First, writers engaging in L2 writing activities are less likely to achieve the same level of topic understanding than when they write in L1. It is due to the fact that text production occurs in units of bursts (Chenoweth & Hayes, 2003), which play a constitutive role in the development of understanding in the dual-process model. The limited proficiency of L2 writers leads to shrinking in burst size, and thus understanding is reduced accordingly. Second, writing in a L2 can exert influences on the balance between explicit planning and implicit text production processes. Since

L2 writing is a more self-conscious activity than L1 writing, it seems that writers are more concerned about satisfying rhetorical requirements than articulating personal understanding. Therefore, priority is more likely to be given to explicit planning processes in L2 writing. Also, writers tend to become de-motivated if they encounter too many difficulties in expressing their personal understanding. This prediction suggests the interplay among cognitive, affective and social factors in the writing process.

2.1.5 Rijlaarsdam and Van den Bergh's model of the writing process

Although the writing models discussed in the previous sections have conceptualised writing as an interactive and recursive process, dynamicity of writing has been more explicitly addressed in Rijlaarsdam and Van den Bergh's (1996) theoretical framework (Figure 2.6). The main components of the model include an executive module, a management module, and a strategic domain. The executive module contains cognitive activities involved in the writing process. The management module is made up of the monitor and several domains of knowledge including linguistic, pragma-linguistic, textual and socio-cultural knowledge. The strategic domain stores crucial writing process information, namely cognitive strategies. The cognitive strategy contains parameter settings which can differ among individuals due to different learning experiences. Parameter settings are information about 'the probability of occurrence of a cognitive activity given the context of occurrence' (p. 109). As writing is characterised by a continuously changing task situation, probabilistic information about the occurrence of a cognitive activity and its relation to preceding and following cognitive activities is assumed to change accordingly. A meter, with values ranging from zero to one, is used to indicate the probability of occurrence of cognitive activities. By studying values on the meter of each cognitive activity in relation to the moment at which they take place, the dynamic character of the writing process can be visualised.

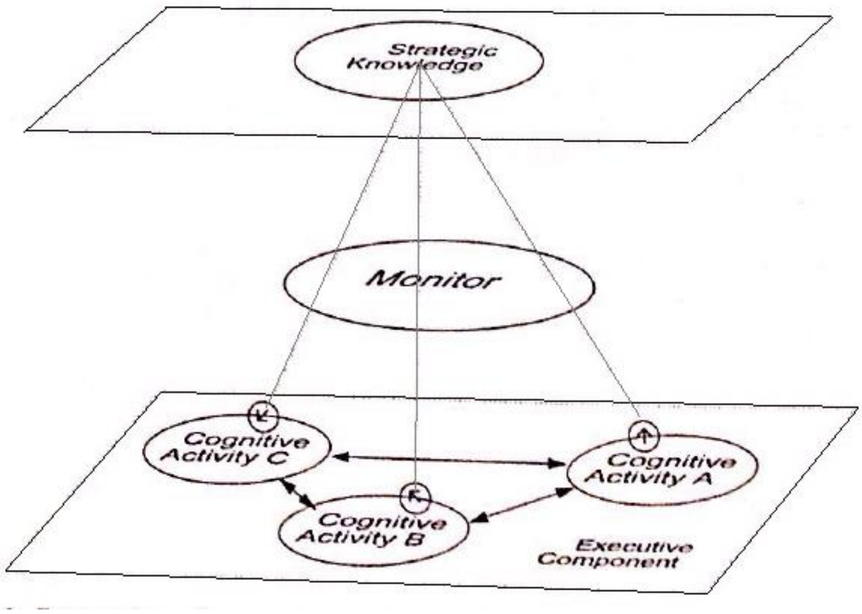

Figure 2.6 Rijlaarsdam and Van Den Bergh's model (1996)

Seven implications for the research on writing processes are inferred based on the model (Rijlaarsdam & Van Den Bergh, 1996), which can be roughly summarised in four categories. First, the distribution of the cognitive activity is not random throughout the writing process. Certain cognitive activities tend to have a higher probability of occurrence in certain stages of writing than in other stages. For example, 'reading the assignment' has a higher probability of occurrence at the beginning of the writing process, and it decreases as the text grows. In other words, this cognitive activity, most likely to be observed at the initial stages of writing, becomes increasingly unlikely towards the end (see also in Rijlaarsdam & Van Den Bergh, 2008).

Second, the same cognitive activity can play varied roles at different points of the writing process. That is, the function of a cognitive activity will depend on the moment at which it is activated. For instance, ‘read the assignment’ in the early stages of writing may be interpreted as an activity tuning the writer to the task. However, the same activity carried out towards the end is more likely to be considered as an evaluation of whether the writing has met the task requirements or gone off at a tangent.

Third, text quality relates to the moment at which a cognitive activity takes place rather than the proportion and frequency of the occurrence. Empirical evidence (Van Den Bergh & Rijlaarsdam, 1996) shows that ‘reading the assignment’ positively correlates with the text quality in the early phases of writing, while the correlation declined drastically and soon turns to negative as writers proceed into the latter phases. Therefore, it can be predicted that those who refer to the assignment more often at the beginning will possibly produce texts with higher quality. On the contrary, writers who consult the assignment more frequently in the latter phases are more likely to write poorer texts.

Lastly, the time-based feature of writing is constrained by inter- and intra-writer variances (see also in Rijlaarsdam & Van Den Bergh, 2008). The principal reason behind the argument lies in the fact that there exists no such thing as an average writer. Writers may be distinct from one another in terms of their default parameter settings, i.e. the ordinary way in which an individual carries out a writing task. In addition, the same writer may also vary in the way he/she adapts parameter settings in response to a changing task situation within one assignment, or as a result of differences among writing assignments.

The essence of the model is the 'attention paid to time as a key variable' because time is 'inherent in the concept of process' (Rijlaarsdam & Van Den Bergh, 1996, p. 112). The inclusion of time in the model helps capture perceptions of the relations among cognitive activities, different functions served by one cognitive activity, and the process-product relationships. Over the past twenty years, writing researchers have begun to realise that it is essential to examine from a temporal dimension how cognitive activities are distributed and interact throughout the writing process. They agree that how cognitive activities are orchestrated is 'a more sensitive and valid reflection of the quality of (the writing) process' (Tillema, 2012, p. 37) than the frequency and proportion of the occurrence of the activities. Empirical studies motivated by Rijlaarsdam and Van den Bergh's (1996) model of writing (Braaksma et al., 2004; Breetvelt et al., 1994; Manchón et al., 2009; Manchón & Roca de Larios, 2007; Roca de Larios et al., 2008; Tillema, 2012; Van Den Bergh & Rijlaarsdam, 1999; Van Weijen, 2008) have uniformly constituted supportive evidence to the notion of writing as a dynamic process, although the findings on the temporal distribution of writing behaviours and/or cognitive activities vary from one study to another.

2.1.6 Summary

To sum up, the models that have been reviewed above share some traits in common in spite of their unique focus on different aspects of the writing process. First, it is generally agreed that effective writing is not only a process of translating ideas into written messages, but also involves the efforts made to satisfy rhetorical requirements. Another common trait of these models concerns the fact that the writing process is non-linear and recursive. It consists of a complex set of cognitive activities which are carried out within the limited capacity of the writer's working memory. To avoid cognitive overload, writers tend to adopt different strategies throughout the process of text production.

It is worth noting that although initially developed to investigate L1 writing processes, these models also frequently serve as theoretical basis in L2 writing process research in either an explicit or implicit way. A main reason for adopting L1 writing models in L2 setting links to an intuitively sound notion that the processes and strategies in text production are transferable across languages (Krapels, 1990). It has been presumed and tested that once a writer has acquired a literacy operation or procedure (e.g. planning) in his/her mother tongue, the skill can be transferred when he/she is learning to write in another language (Roca de Larios, Murphy, & Marín, 2002; and see more in Cummins, 1980). In other words, the mental processes of L1 and L2 writing are assumed to be broadly similar as they are supposed to be supported by a common set of abilities. The only difference between the two processes lies in how written texts, or the products of the process, are 'linguistically formulated' (Roca de Larios et al., 2016, p. 270).

The current thesis jointly adopted Kellogg's (1996) cognitive model of writing and Rijlaarsdam and Van Den Bergh's (1996) temporal model of writing as the theoretical basis. Kellogg's model was selected for two reasons. First, it divides the writing process into several activities and offers detailed descriptions of each category. More importantly, compared to other writing models, Kellogg's models places a greater emphasis on linguistic encoding processes, which are expected to involve more conscious effort in L2 than L1 writing (Kormos, 2012). This feature makes Kellogg's model particularly suitable for the thesis given that the main goal of this study is to investigate the cognitive processes involved in L2 Chinese writing. Noteably, coverting ideas into written messages in Chinese using the Pinyin input method may be even more cognitively demanding, as the encoding process tends to involving establishing connections among mental images, sound and logographic written forms.

The thesis was additionally framed within Rijlaarsdam and Van Den Bergh's (1996) temporal model of writing on the ground that it provides a dynamic insight into analysing the writing process. Although the idea of a dynamic writing process has been either explicitly or implicitly discussed in other writing models, the temporal model seems to be the only one that attempts to visualise the dynamism of writing by including time, a key variable in the writing process, in the analysis. The model does not directly predict how exactly different activities relate to time during writing, but it does demonstrate the notion that a temporal analysis of the writing process is essential.

The model also lends itself well to empirical research, as reflected in a small number of empirical studies in the fields of L1 and L2 writing (e.g. Tillema, 2012; Van Weijen, 2008). Results from these studies indicate the dominance of certain cognitive activities in certain stages of writing, and that the contribution of writing processes to text quality is associated with the moment at which writing behaviours and/or cognitive activities occur.

2.2 Empirical studies

This section conducts a critical review of previous empirical studies related to the current one, including empirical research on the temporal dimension of writing processes, studies on the links between writing processes and text quality, and work on the role of genre and L2 proficiency in writing processes and products.

2.2.1 Writing processes and the role of time

Previous work on the role of time in writing processes has mostly been conducted in the field of L1 writing. For instance, Breetvelt et al. (1994) investigated the distribution of eleven cognitive activities during writing using think-alouds. Twenty ninth-graders performed two argumentative writing tasks. The think-aloud protocols were divided into three episodes, each containing an equal number of fragments of reported cognitive activities. The researchers found that the frequency of most cognitive activities changes during writing. The frequency of reading the assignment, self-instruction, goal-setting, generating ideas and making comments decreased from the first episode to the last, while that of formulating, rereading, evaluating and revising increased over time.

Kellogg (1988) conducted two experiments examining how different outline conditions affected the allocation of cognitive activities. Participants were asked to report their mental activities (i.e. planning, translation and revising) once every minute while writing. Each writing session was divided into three equal phases based on the total writing time. The results indicated that participants in the non-outline condition reported more planning and translation activities in phases 1 and 2, but more translating and revising activities in phase 3. Participants in the outline condition concentrated on translating in all three phases. The manipulation of outlines (rough versus polished, mental versus written) had little influence on the distribution of cognitive activities. Difference was only found in revising in phase 1 between the rough and polished outline conditions.

Tillema et al. (2011) investigated to what extent the distribution of cognitive activities could be predicted by a questionnaire that assessed participants' writing styles. Twenty secondary school students produced four argumentative essays while verbalising the thoughts they had while writing. The participants also completed a questionnaire measuring their writing style in terms of the degree of planning and revising. The verbal protocols elicited in the study were divided into five equal episodes, each including an equal number of cognitive activities. The researchers reported that writing style had an effect on the distribution of two cognitive activities, planning and reading the assignment. Participants with higher planning scores tended to engage in planning activities more often at the beginning of writing, but less often at the end, than participants with lower planning scores. Participants with higher revising scores were less likely to read the assignment in episode 1 than those with lower revising scores; however, they were more likely to read the assignment in episodes 2–5.

In the field of L2 writing, the temporal aspect of writing has been touched upon in some early work on planning and revising. Cumming (1989) distinguished two types of L2 writing experts, advanced planners and emergent planners, depending on how they allocated planning activities during the writing process. Advanced planners usually spend several minutes before writing thinking about the major elements and summarising them into numbered points. They only begin to produce texts when they are certain about what to write. In contrast, emergent planners plan and discover what they want to write as they are writing it on paper. Consequently, major decisions on the content and organisation are usually made at various points during writing, rather than at the beginning. In the same vein, Ellis and Yuan (2004) also differentiated the two principal types of planning activity by the moment at which it takes place. The two types of planning are pre-task planning, in the form of either strategic planning or rehearsal (Ellis, 2005), and online planning, which takes place during the writing process. In terms of revising, Zamel (1983) reported that experienced L2 writers were more likely to postpone linguistic concerns to the end of writing, whereas novice writers tended to be concerned with revisions from the very beginning so as to avoid lexical and syntactic errors. In contrast, Porte (1997) found that unskilled L2 writers tended to carry out most text-level revisions in the latter phases of writing, and Victori

(1999) observed that some skilled L2 writers also engaged in revising activities throughout the writing session.

Motivated by Rijlaarsdam and Van den Bergh's (1996) theoretical framework, growing attention has been paid to the role of time in L2 writing in recent years. A few L2 researchers have explored the distribution of writers' cognitive activities at different writing stages (e.g. beginning, middle, end) using verbal reports, such as think-alouds and stimulated recall. Manchón and her colleagues (Manchón et al., 2009; Manchón & Roca de Larios, 2007; Roca de Larios et al., 2001, 2006, 2008) carried out a series of studies investigating how cognitive activities were distributed throughout L2 writing processes. Twenty-one Spanish learners of L2 English were asked to write an argumentative essay in English on the topic of education while saying aloud the cognitive activities in which they engaged. The think-aloud protocols were divided into three equal periods based on the participants' total writing time. The researchers found that planning, formulation and revising were three major cognitive activities involved in text production, accounting for around 90 per cent of total writing time. Among these, formulation took up the largest amount of time during writing. Planning mainly occurred in period 1; formulation peaked in period 2; revising increased gradually from period 1 to period 3.

Van Weijen (2008) examined 20 Dutch undergraduates by asking them to think aloud while writing four argumentative essays in L2 English. She found that both process planning (i.e. self-instruction, goal-setting and structuring) and content planning (i.e. ideas generation) were very unlikely to occur during text production. On the few occasions where reference was made to planning, both types of planning occurred more frequently at the beginning of writing. Participants were found to carry out more formulation activities in the middle, while reading the assignment more often at the beginning and the end.

Tillema (2012) conducted a study with a similar design to Van Weijen's (2008) on teenage L2 English learners. Data were collected from 20 Dutch secondary school students via think-alouds and keystroke-logging. The transcripts of think-alouds were segmented, with each segment representing a cognitive activity. The segmented protocol was then divided into five episodes, each containing the same number of segments. The researcher reported that the occurrence of content planning (i.e. ideas generation, goal-setting and structuring) decreased throughout the writing process, while that of process planning (i.e. monitoring and metacomments) first decreased and then increased. Formulation occurred most frequently in the middle. Reading the assignment was most likely to occur in episode 1, and least likely to occur in episode 5, whereas evaluating one’s own text exhibited an opposite trend. The occurrences of revising and reading one’s own text were equal but infrequent throughout the writing process.

Barkaoui (2015) investigated L2 writers' cognitive activities when performing an independent and an integrated writing task. The participants were 22 users of L2 English. Stimulated recall was employed to elicit participants' cognitive activities when completing the writing tasks. Each writing session was segmented into three equal phases based on the total writing time, and data collected from each stimulated recall protocol were divided into three parts accordingly. Results showed that, for both tasks, participants reported more interaction with task in phase 1 and more evaluating and revising in phases 2 and 3. In terms of independent tasks, more stimulated recall comments referred to planning activities (i.e. planning, organising and global planning) in phase 1 and time-checking in phase 3. In terms of integrated tasks, participants tended to interact with sources more often in phase 1, carry out more global planning in phase 2, and engage in the activity of detecting writing difficulties, using writing strategies, and evaluating and revising more frequently in phases 2 and 3.

Khuder and Harwood (2015) explored the cognitive activities L2 writers engaged in when producing texts in test and non-test conditions. Ten English L2 learners were asked to write an argumentative essay in each of two conditions, test and non-test. Their writing processes were recorded, and stimulated recall interviews were utilised to tap into participants' cognitive activities. The writing process of each performance was divided into three periods, each covering one third of the total writing time. The researchers found similar distributions of planning and evaluating between the two conditions: planning decreased from period 1 to period 3, whereas evaluation decreased slightly in period 2 but increased considerably in period 3. However, differences were observed in terms of revising and formulation. More surface revisions were made in period 3 than in periods 1 and 2 in the test condition, while in the non-test condition, most surface revisions occurred in periods 1 and 2. The majority of meaningful revisions occurred in period 3 in the test condition, whereas in the non-test condition, most meaningful revisions were carried out in periods 2 and 3. Participants referred to formulation activities most often in period 3 in the test condition, while in the non-test condition, most formulation activities were found in period 2.

Other studies have focused on how writing behaviours, such as pausing and revision, were distributed throughout L2 writing processes. In terms of pausing behaviours, Xu and Qi (2017) explored the distribution of pausing behaviours across five intervals of writing. Participants were 30 less skilled and 29 more skilled writers of L2 English, grouped by their writing score in an English placement test upon entering university. An argumentative writing task was used to elicit participants' performances in L2 writing. Their pausing behaviours were captured by *Inputlog* 6.0. Each writing session was divided into five intervals, each with an equal length of writing time. Statistical analyses revealed that less skilled L2 writers paused more often in interval 4 than in intervals 1, 2 and 5. They also paused longer in interval 2 than in intervals 1, 3 and 4. More skilled writers paused less frequently in interval 1 than in intervals 2–5, but paused longer in interval 1 than in intervals 2–4. More skilled writers, in addition, pause more often in interval 2 than in interval 4.

Barkaoui (2019) looked at the pausing behaviours of 68 learners of L2 English who performed two writing tasks, one independent and one integrated, on a computer. Their writing behaviours were logged. The writing sessions were segmented into three equal intervals based on the total writing time. Similar to what Xu and Qi (2017) observed in more skilled L2 writers, Barkaoui found that participants paused less often but for longer in interval 1 than in intervals 2 and 3.

In terms of revision behaviours, Barkaoui (2016) examined the revision behaviours of 54 L2 English learners who were asked to write an argumentative essay and a summary. Data on revision behaviours were collected via a keystroke logging program. Writing sessions were divided into three phases, as in Barkaoui (2019). The researcher found that the majority of revisions occurred in phase 2 rather than phase 3. When revision behaviours were sub-divided according to the context (i.e. the location of occurrence), the findings were slightly different from the overall picture. It was reported that contextual revisions (i.e. those made to the text written prior to the point of inscription) occurred more frequently in phase 3, whereas pre-contextual revisions (i.e. those undertaken at the point of inscription) were more likely to take place in phase 2.

Taken together, the studies described above reveal that, similar to L1 writing, L2 writing is a dynamic process, given that the occurrence of most cognitive activities and writing behaviours is not constant across different writing stages. Specifically, researchers investigating cognitive activities during L2 writing using think-alouds tend to agree that planning dominates the initial stages of writing. This finding is further supported by keystroke-logging studies, which observe fewer but longer pauses at the beginning of writing. Formulation was found to be the dominant activity in the middle of writing stages in most think-aloud studies. However, this does not fully align with studies exploring pausing behaviours using keystroke logging, as difference is not always detected between the middle and the end (e.g. Barkaoui, 2019;

Xu & Qi, 2017). This minor inconsistency may be explained by contextual factors. For example, in Khuder and Harwood (2015), L2 writers engaged more in formulation activities in the final stage in the test condition than in the non-test condition, which the researchers interpreted as 'an attempt to spend more time on the task' (p. 266).

No real consensus has been reached on the temporal distribution of revision activities. This lack of uniformity may be explained by several methodological issues. First, different data collection methods can lead to different distribution patterns in revision. Keystroke logging, normally employed by researchers investigating writing behaviours, can provide more accurate and comprehensive information about L2 writers' revision activities than verbal reporting. Second, the different coding schemes used for categorising revisions may be another possible reason for inconsistencies. Barkaoui's (2016) and Khuder and Harwood's (2015) coding schemes are much more fine-grained than those used by other researchers when analysing revisions. Thirdly, discrepancies in the findings can also be attributed to the number of tasks used to elicit the processes in which L2 writers engaged. Among the studies described above, the majority adopted a single-task design, apart from Van Weijen (2008) and Tillema (2012). In other words, in most studies, participants were only asked to complete one writing task in a certain condition. A design with a single task carries a potential risk, as it may fail to disentangle the effect of the task. As a result, the task effect may become a confounding variable in the analysis, thus yielding mixed results for revision. In addition, the lack of consistency may be ascribed to the way in which the variable of time is operationalised. In studies other than Tillema (2012), time was measured in minutes and seconds, while in Tillema (2012) the time variable was constructed in terms of segments, where the start of a new segment was represented as the start of a cognitive activity. These operational differences make the weight of revision in relation to other writing activities different across studies, which might contribute to inconsistent findings. Lastly, diversity in the demographic backgrounds of participants (e.g. age and educational background) may be an additional reason for differences in terms of how revision was distributed during writing.

2.2.2 Writing processes and text quality

Process-product relations have received considerable attention in L1 writing research. One of the first studies was conducted by Breetvelt et al. (1994), who explored the link between text quality and cognitive activities elicited by think-alouds. Twenty ninth-grade students performed two argumentative writing tasks in L1. Think-aloud protocols were divided into units, with each unit referring to one cognitive activity. Text quality was assessed by three raters on four aspects (i.e. goal orientation, structure and organisation, audience awareness, and language usage or style). In order to determine whether the relation between cognitive activity and text quality changes across the process of writing, think-aloud protocols were segmented into three episodes, each containing one-third of the total number of units in the protocol. In episode 1, text quality was positively correlated to reading the assignment and evaluation, while negatively correlated to goal-setting, structuring and revising text. In episode 2, more goal-setting and structuring improved text quality, whereas more reading of the assignment, making comments, pausing, evaluating and revising the text had a negative impact on text quality. In episode 3, self-instruction, goal-setting, generating ideas, writing and rereading were found to have a positive influence on text quality.

More recent studies have employed less obtrusive methods, such as keystroke logging, to investigate the relationship between visible behaviours in which L1 writers engage and the quality of writing. For instance, Torkildsen et al. (2016) examined the relationship between writing behaviours recorded by a keystroke logging program and text quality. Forty-two L1 Norwegian third-graders completed a picture-elicited narrative writing task in a computer lab. Text quality was assessed in terms of overall quality, essay length, percentage of spelling errors, and percentage of capitalisation and punctuation errors. Results showed that participants with greater transcription fluency and who made more online revisions (i.e. changes made to the rightmost word in the text) produced longer narrative essays of higher overall quality. The researchers also observed a negative correlation between transcription fluency and percentage of spelling errors.

In L2 writing, much less is known about the extent to which writing processes predict text quality despite some initial attempts. A small number of studies have investigated how writing behaviours, such as fluency, pausing and revision, are related to the quality of writing. In Spelman Miller et al. (2008), 17 Swedish high school students learning L2 English were asked to write one descriptive essay each year over the three-year period of the research. Participants' writing behaviours were recorded by a logging tool. Fluency was measured by several indices, including number of characters typed per minute, burst (number of characters produced between pauses and/or revisions) and fluency during burst (time on text production between pauses and/or revisions). Pauses were defined as interruptions to keyboard and mouse activity of over two seconds, while deletions and insertions were grouped as revision. Results revealed that burst and fluency during burst emerged as strong predictors of text quality, while no relationship was found between pausing or revision behaviours and text quality.

Révész et al. (2017) looked into how the relation between writing behaviours and text quality was mediated by task complexity. The participants were 73 advanced learners of L2 English. They were randomly assigned into two groups and wrote one argumentative essay under either a simple or a complex task condition. Data on writing behaviours were captured by a keystroke logger. Text quality was assessed with a battery of linguistic complexity measures. The researchers observed that, for the simple task version group, participants who paused for a shorter time between clauses tended to produce texts with greater syntactic structure diversity. For the complex task version group, participants who paused for a shorter time between sentences were more likely to use less frequent words in the text, and those who revised less at and above the clause level tended to produce essays with greater lexical sophistication.

A few studies have explored the link between a single writing behaviour (i.e.

revision) and L2 text quality. In Stevenson et al. (2006), 22 young learners of L2 English wrote two argumentative essays in English. The participants' revision behaviours were captured by both think-aloud protocols and a keystroke logging program. Text quality was measured by separate holistic rating scales in terms of content and language. It was assumed that word-level revision would negatively correlate with text quality, since more attention to lower level writing processes could inhibit the allocation of attentional resources to higher level cognitive operations in writing. However, the findings did not confirm the prediction. No relationship was found between the frequencies of any types of revision and text quality.

Stevenson et al.'s (2006) findings were confirmed by Chambers (2011), who examined the relation between revision and text quality of adult L2 English learners. In this study, ten test-takers of a Business English Certificate Vantage Writing Test wrote a business report using a word processor. Data on revision were collected by a software program that took snapshots of the screen every 30 to 40 seconds while the participants were writing. Revision was identified by comparing a snapshot to a preceding one. No relationship was found between revisions and the quality of the business report.

There appears to be no clear pattern as to how text quality relates to writing behaviours from the empirical studies described above. Nevertheless, these studies tend to be uniform in suggesting that revision is a weak predictor of L2 text quality. This does, to some extent, run counter to studies that compared the revision practices of skilled and unskilled L2 writers. In these studies, noticeable differences were reported in terms of how skilled and unskilled L2 writers revised their texts. For example, skilled L2 writers revised more often at a global level (e.g. Hirose & Sasaki, 1994; Sasaki, 2004; Skibniewski, 1988) and tended to postpone revision towards the end of writing (e.g. Victori, 1999; Zamel, 1983). As skilled writers are supposed to produce better texts, differences in revision behaviours could, at least partially, predict variation in text quality. As a result, it is not uncommon for these researchers to propose pedagogical implications, such as encouraging unskilled writers to pay more attention to the global aspects of a text (e.g. content, organisation) by carrying out more revisions to larger textual units. However, keystroke-logging studies that have examined the relationship between revision and text quality do not provide enough supportive evidence. In other words, more revision to global aspects does not guarantee better texts, while frequent revision to local aspects (e.g. linguistic issues) does not necessarily result in a poorer text.

One possible reason for these mixed findings may be associated with the multifaceted nature of revision, which is both multi-dimensional (Lindgren & Sullivan, 2006; Stevenson et al., 2006) and temporal (Rijlaarsdam & Van Den Bergh, 1996, 2008). Previous research has rarely considered the relationship between revision behaviours and text quality from a temporal perspective. That is, revision behaviours in these studies are simply presented in numbers and percentages for the whole writing session. As the same activity can play different roles in different stages of writing (Rijlaarsdam & Van Den Bergh, 1996), it seems reasonable to assume that the relationship between revision and text quality may vary throughout the writing process. Therefore, text quality predicted by overall numbers and percentages of revision might not be so reliable. Although previous studies (Chambers, 2011; Spelman Miller et al., 2008; Stevenson et al., 2006) have observed no relationship between revision and text quality, it is still possible that revision might positively or negatively affect text quality at certain moments of the writing process. Clearly, more research is required to investigate the relation between L2 writing processes and text quality, in particular with the variable of time included.

2.2.3 Process-product relations and the role of time

The previous section has described empirical studies on process-product relations without considering the time points at which writing activities occur during the composition process. In this section, I turn to discuss L2 investigation of the timing of writing activities in relation to text quality, given that writing is a dynamic process, and the role of a writing activity is assumed to change across different stages of the text production process (Rijlaarsdam & Van Den Bergh, 1996).

The majority of such studies have focused on how text quality is affected by planning activities carried out at different moments of task execution (i.e. pre-task planning and online planning). Ellis and Yuan (2004) suggest that pre-task planning was hugely beneficial for fluency, while online planning led to greater accuracy in narrative writing tasks. These findings were replicated by Ghavamnia et al. (2013) and partially confirmed by Piri et al. (2015), in that pre-task planning significantly improved fluency in L2 narrative writing. However, Ong and Zhang (2010) did not obtain similar results for argumentative writing tasks. One hundred and eight pre-university Chinese learners of L2 English wrote an argumentative essay under one of four conditions: pre-task planning, extended pre-task planning, free writing and a control group. It was reported that, compared to the extended pre-task planning and pre-task planning groups, the free-writing group, which was allowed sufficient time for online planning, scored significantly higher on fluency and complexity, rather than accuracy. The conclusion that online planning improves complexity is supported by Abdi Tabari (2016) for L2 descriptive writing. Hsu (2012) also suggests there are advantages from online planning for complexity, accuracy and fluency (CAF), but no advantages for pre-task planning were found when L2 writers were involved in a synchronous computer-mediated communication context.

It is worth noticing that, although considering the temporal aspect of planning, these studies have a problem in operationalising online planning activities (Skehan et al., 2012). Although the researchers expected the participants to engage in online planning by allowing them unlimited time to write, there was no direct indicator, such as concurrent or retrospective verbal reports (Skehan & Foster, 2005), suggesting whether online planning indeed took place. In addition, there is also a problem associated with the pre-task planning group(s). According to Galbraith's (1999) dual-process model of writing, writers tend to produce new ideas when engaging in spontaneous text production even though they have planned and developed their ideas prior to formulation. New ideas are 'associated with the development of the writer's personal understanding of the topic' (Galbraith, 2009a, p. 17). There is also empirical evidence (Ong & Zhang, 2013) in the L2 setting showing that pre-task planners continued to plan while formulating, even though they were required to complete the writing task within a time limit. Therefore, it is arguable whether the improvement in the pre-task planners' text quality can be attributed to pre-task planning alone. The improvement might also be a result of continuous online planning or a combination of pre-task and online planning.

Another potential pitfall of the task-based studies mentioned above concerns the way in which fluency was measured. Without employing keystroke logging techniques, fluency in these studies was mainly determined based on the length of the final text. The sole use of product-based measures to capture writing fluency can be challenged on several grounds. First, writing fluency cannot be fully assessed with product-based measures only, as these measures do not take into account the real-time dimension of text production, such as pausing and revision (Abdel Latif, 2013). In addition, previous studies have shown that writing fluency is a multi-dimensional construct and cannot be validly measured from a single perspective (e.g. Leijten & Van Waes, 2013). For instance, writing fluency is also found to be related to production bursts, i.e. the number of bursts and the mean length of bursts (e.g. Abdel Latif, 2009; Chenoweth & Hayes, 2001). Given that the above studies incorporated neither aspect, the findings regarding the effects that the temporal features of planning had on writing fluency are questionable.

Two previously discussed studies, Tillema (2012) and Van Weijen (2008), explored the relationship between multiple cognitive activities and L2 text quality using verbal protocols. By rating text quality holistically, Van Weijen (2008) found that in L2 writing, process planning (i.e. self-instruction, goal-setting and structuring) negatively correlated with text quality at the beginning of argumentative writing, but the correlation became positive towards the end. Content planning (i.e. ideas generation) always demonstrated a positive correlation with text quality, but the correlation was much stronger at the beginning. Reading the assignment only positively correlated with text quality at the beginning. Formulation had a negative influence on text quality at the beginning, but the influence turned positive after the initial phase. Different findings were reported by Tillema (2012). As mentioned earlier, she divided the think-aloud protocols elicited in her study into five episodes, each including an equal number of segments of reported cognitive activities. She found only negative correlations between cognitive activities and text quality in certain episodes in L2 writing. Process planning negatively correlated with text quality in episodes 1, 3 and 5, and content planning in episode 4. There was a negative correlation between reading the assignment and text quality in episodes 1 and 2, and between formulation and text quality in episode 5. Evaluating one's own text negatively affected text quality in episodes 2, 4 and 5. No relationship was found between revising and reading one's own text and text quality.

It seems impossible to compare the findings of planning studies in task-based language teaching contexts to those of Van Weijen (2008) and Tillema (2012), given that the two lines of research assess text quality in different ways. In task-based planning studies, more attention has been allocated to areas of language use, such as

CAF (though see Kormos, 2012), whereas Van Weijen (2008) and Tillema (2012) employed holistic ratings that balanced linguistic and rhetorical concerns. Both assessment approaches carry some potential risks. Objective linguistic measures can explain a large part of text quality in terms of language use, but text quality is also dependent to rhetorical concerns, such as ideas and organisation, which are not captured by CAF measures (Deane, 2013; Weigle, 2013). Overall scores can be biased, especially when the rating scales are ill-designed and/or human raters are not well trained.

More recently, some researchers have looked into how writing behaviours, captured by keystroke logging or eye-tracking, predict L2 text quality at different moments during writing. In Xu and Qi's (2017) previously discussed study, the researchers additionally investigated the link between text quality and pausing behaviours in five writing intervals. Statistical analyses revealed that fewer but longer pauses in Interval 1 benefitted text quality.

Gánem-Gutiérrez and Gillmore (2018) examined the relationship between text quality and five writing activities, text construction, revising, pausing, rereading and using external resources. Twenty-two L2 English learners wrote an argumentative essay with their writing behaviours and eye movements captured by a screen recorder and an eye-tracker. Each writing session was divided into five equal-length periods based on each participant's total writing time. Text quality was assessed via holistic ratings. No relationship was found between text quality and the duration and frequency of any writing activities across the five writing periods.

In Xu (2018), 57 Chinese undergraduates wrote an argumentative essay in L2 English. Their revision behaviours were captured by a keystroke logging program. Revisions were first divided in terms of temporo-spatial location into immediate revisions (i.e. those made at the end of the current text), distant revisions (i.e. those made to already-written text followed by further text production) and end revisions (i.e. those made to already-written text at the systematic reviewing period, after the main text has been completed). Immediate revisions were then further distinguished according to frequency and scope (i.e. the number of letters deleted or produced), whereas distance and end revisions were classified by frequency, scope, distance (i.e. the number of letters between the revision and production points) and duration (i.e. the time in seconds that the revision event lasted). Xu's findings were not in line with Gánem-Gutiérrez and Gillmore (2018) and the empirical studies described in Section 2.2.2 (Chambers, 2011; Révész et al., 2017; Spelman Miller et al., 2008; Stevenson et al., 2006). Xu reported that fewer revisions to already-written text benefited text quality in the text production period, while revisions to larger textual units in the systematic reviewing period led to better texts.

Taken together, it seems that no conclusion can be drawn from the existing empirical studies as to how writing processes predict L2 text quality. Clearly, further investigation is needed of this issue.

2.2.4 Genre and writing

The previous sections have reviewed the empirical studies investigating the temporal dimension of writing processes and process-product relations in L2 writing. This section reports the empirical findings on the role of a task-related factor (i.e. genre) in writing processes, and the relationship between writing processes and text quality. Due to the limited L2 investigation of these issues, this section additionally includes studies in the field of L1 writing.

Genres are certain types of texts constructed by writers in order to achieve specific social purposes (Hyland, 2003). The two most extensively explored genres in writing studies are narrative and expository pieces (Ortega, 2015). Narrative writing involves characters, their actions and motivations, and it expresses the unfolding of events in a temporal framework (Ravid, 2005). Typical narrative essays include recounting, historical events, (auto)biographies, tales and fiction (Grabe, 2002). Expository writing, on the other hand, focuses on ideas, concepts, claims and arguments (Katzenberger, 2005). In a broad sense, expository essays cover a wide range of non-narrative text types, such as persuasive, compare and contrast, and argumentative pieces (Jeong, 2017). Given the specific discourse purpose that the text of a certain genre serves, it is assumed that the language features and processes involved in text production may vary across genres.

A great number of L1 and L2 writing studies have confirmed that non-narrative essays are in general more linguistically complex than narrative ones (see Beers & Nagy, 2009, 2011; Biber & Conrad, 2009 for L1 writing; see Lu, 2011; Way et al., 2000; Yoon & Polio, 2016 for L2 writing). However, much less is known about whether or not writers produce essays of different genres in a similar way. Most studies that have looked at this issue to date were conducted in the field of L1 writing, and the findings are both inconclusive and inconsistent. In a small-scale exploratory study, Matsuhashi (1981) observed traces of the effect of genre on the writing process. The participants were four high school students, who were skilled writers in their L1. Each participant wrote six essays in three discourse types, narrative reporting, generalising and persuading, over a period of approximately three weeks. Writing processes were analysed via pause length, which was assumed to offer cues to planning activities in writing. The researcher found that participants paused longer before T-units in generalising and persuading than in narrative reporting, indicating a greater demand on planning in producing generalising and persuading pieces.

Similarly, Kellogg (2001) found that genre influenced the cognitive processes in L1 writing. Forty-eight university students produced two timed essays (i.e. narrative, persuasive or descriptive), one by hand and one on a computer. While performing one of the writing tasks, participants were asked to identify the types of cognitive activities in which they engaged (i.e. planning, translating and reviewing) as soon as they heard a computer-generated probe that sounded every 30 seconds. Planning was found to be highest in percentage terms in persuasive writing (26%) and lowest in descriptive writing (21%). Descriptive essays elicited the most translating (55%), and narratives the least (48%). The percentage of reviewing was similar across three genres.

Genre effect was additionally reported in Medimorec and Risko (2017), who compared L1 English writers' pausing behaviours when composing argumentative and narrative essays. Fifty undergraduate students performed a narrative writing task, and 51 performed an argumentative task. Their pausing behaviours were recorded by a keystroke logger and coded in terms of three length intervals (300–999 milliseconds, 1000–1999 milliseconds and 1999+ milliseconds) and at three text boundaries (word, sentence and paragraph). Argumentative writing featured more pauses at word boundaries at all three length intervals and slightly more between-sentence pauses in the 300–999 interval than narrative writing.

Contrary to Matsuhashi (1981), Kellogg (2001) and Medimorec and Risko (2017), almost no genre effect was observed in Beauvais et al. (2011). Twenty-four psychology students completed a narrative and argumentative writing task while thinking aloud. The whole writing session was recorded by a video camera. The writing process was coded in terms of pre-task pausing, fluency (words per minute), planning, translating, reading and revising. The latter four activities were calculated in terms of the percentage of time spent on each activity, the number of episodes by activity per minute and mean length of episodes by activity. The researchers found that participants' writing processes were largely similar when composing narrative and argumentative essays. Participants spent significantly more time on translating than planning and reading, and the least on revising. Translating episodes were longer than planning episodes, followed by reading and revising episodes. In terms of number of episodes per minute, both genres exhibited more episodes of translating and reading than planning and revising. The only difference was that narrative writing activated more episodes of planning than revising per minute, while the figures were similar for argumentative writing. The results were, to some extent, confirmed by Beauvais et al. (2014), who compared the temporal management of writing processes between procedural and expository descriptive writing. Two hundred and fifty-two fifth-, seventh- and ninth-grade students composed either a procedural or an expository descriptive text on a tablet with 11 ideas provided. Temporal management of the writing process was operationalised in terms of the percentage of time spent on pre-task planning, pausing and translating. No difference was found on the three measures between genres, indicating similar temporal management of pre-task planning, pausing and translating in procedural and expository descriptive writing.

In addition, Beauvais et al. (2011) further investigated how process-product relations were influenced by genre. Results showed that in narrative writing, shorter episodes of planning and translating led to better texts; the length of pre-task pauses and fluency were not strong predictors of text quality. In contrast, better argumentative essays were associated with a greater amount of time spent on planning and translating. Participants also benefited from longer pre-task pauses in argumentative writing. The fluency indicator negatively correlated with the quality of argumentative texts, suggesting that a lower production rate might be beneficial for argumentative writing.

One study in the field of L2 writing (Thorson, 2000) examined whether or not writers exhibited different online revision behaviours when writing in different genres. Each one of 18 participants wrote a newspaper article and a letter to an Austrian pen pal in both L1 English and L2 German on a computer. Their revision behaviours were captured by a keystroke-logger. Revisions were distinguished based on the distance between the cursor and the revision as intermediate (i.e. the distance being zero) and distant (i.e. the distance being above zero). The researcher found no difference between genres in L2 writing in terms of the number of total revisions standardised by total number of characters produced and the ratio of distance revisions to total revisions. However, in L1 writing, participants revised more frequently when producing the article than the letter. Clearly, more studies are needed to confirm the findings.

2.2.5 L2 proficiency and writing

The previous section discussed the effect of genre, a task-related factor, on writing processes and process-production relations. In this section, I turn to review the empirical research on how L2 proficiency, a writer-related factor, is associated with the cognitive processes in which L2 writers engage.

The complex and demanding task of text composition requires writers to have strategic knowledge to satisfy rhetorical goals, metacognitive knowledge to control, monitor and evaluate their writing processes, and linguistic knowledge to convey the intended meaning in written form (Whalen & Ménard, 1995). While strategic and metacognitive knowledge may be transferable between L1 and L2, linguistic knowledge (i.e. L2 proficiency) is language-specific, and thus plays a particularly important role in L2 writing performance.

It is generally agreed that L2 proficiency is one of the major contributors to difference in L2 text quality. For example, Hirose and Sasaki (1994) and Sasaki and Hirose (1996) found that L2 proficiency accounted for the largest proportion of variance, 57% in Hirose and Sasaki (1994) and 52% in Sasaki and Hirose (1996), in L2 text quality, which was operationalised as the sum of the scores giving by two independent raters using a holistic rating scale. In Schoonen et al. (2003), L2 proficiency was found to be the second strongest predictor (47%) of overall writing performance (i.e. holistic ratings by six raters), following L1 writing proficiency (49%). The researchers additionally found L2 proficiency contributed more to L2 text quality than did L1 proficiency to L1 text quality, implying that language proficiency plays a more important role in L2 than L1 writing. Barkaoui (2014) further discovered that L2 proficiency explains the difference in the quality of L2 writing, regardless of task type and writing mode. Fifty-two more proficient and forty-five less proficient participants completed two independent writing tasks (one paper-based, one computer-based) and a computer-based integrated task. Results showed that higher-proficiency participants received higher holistic scores than lower-proficiency participants on all three tasks.

Although higher L2 proficiency is often associated with higher text quality, more and less proficient L2 writers may also differ in the ways in which they develop a piece of writing. Theoretically, it is assumed that a writer’s level of L2 proficiency may affect their distribution of attentional resources to higher-level (e.g. planning, monitoring) and lower-level (e.g. linguistic encoding) cognitive activities during writing due to limited working memory capacity (Kellogg, 1996). However, only a few empirical studies have examined the role of L2 proficiency in L2 writing processes. Cumming (1989) compared the writing processes of 23 university students at two levels of L2 proficiency, which was distinguished via a speaking test. Each participant wrote three essays in L2 English while thinking aloud. To analyse the writing process, think-aloud protocols were classified in terms of the aspects to which participants referred in decision-making and the types of problem-solving strategies that participants adopted while writing. Results indicated no relationship between L2 proficiency and the writing process, as the focus on decision-making and problem-solving behaviours during writing differed little between participants of high and low L2 proficiency.

In contrast, Chenoweth and Hayes (2001) did find an effect of L2 proficiency on writing processes. Thirteen undergraduate students taking L2 French or L2 German courses in either their third (low proficiency) or fifth semester (high proficiency) participated in the study. They were asked to perform a descriptive writing task in L2 while verbalising thoughts. Results showed that participants with high L2 proficiency wrote more words per minute than those with low proficiency. They also produced longer bursts between pauses and between revisions, and made fewer revisions than low proficiency participants.

Like Chenoweth and Hayes (2001), Tiryakioglu et al. (2019) also observed differences in writing processes between L2 writers with different levels of proficiency. The participants were eight Turkish high school students. They were divided into high- and low-proficiency groups (N = 4 in each group) based on their performance on an Oxford Placement Grammar Test. They were asked to compose an argumentative essay in L2 English under a think-aloud condition. Their writing behaviours were captured by keystroke logging software. Results revealed that high proficiency was associated with greater fluency (i.e. more words per minute), fewer pauses and less frequent revisions. In addition, students with high proficiency spent more time formulating, while low-proficiency students engaged more frequently in reading the task and searching for lexical items.

The lack of consistency in the findings of these studies might be as a result of the crude way in which L2 writing processes were analysed. As mentioned earlier, L2 writers may engage in certain writing activities to a different extent at different moments of writing. Thus, even though more and less proficient L2 writers were found to demonstrate an equal number/ percentage for a certain writing activity for the whole writing period, it does not necessarily indicate they were involved in similar processes of text production. In other words, they can still differ in terms of the time in which they engage in certain writing activities. For the same reason, the differences observed in Chenoweth and Hayes (2001) and Tiryakioglu et al. (2019) may not be fully valid either. It is possible that the processes in which high- and low-proficiency writers engage may only be different for a given stage, rather than the entire period of writing. Therefore, it is essential to also consider the role of time when investigating the link between proficiency and L2 writing processes.

Several studies mentioned in Sections 2.2.1 and 2.2.2 have sought to address this issue. Manchón and her colleagues (Manchón et al., 2009; Roca de Larios et al., 2008) found a relationship between L2 proficiency, operationalised as participants' level of education, and writing processes. For the whole writing session, with growth in L2 proficiency, there was a gradual increase in the time spent on planning and revising, and a parallel decrease in formulation. When writing was divided into three periods, L2 proficiency was found to mediate the temporal distribution of formulation and metacomments. Pre-intermediate and intermediate learners devoted the most time to formulation in period 1, and the time allocated to formulation gradually decreased in periods 2 and 3. Advanced learners concentrated on formulation in period 2. Metacomments mainly occurred in period 1 for pre-intermediate learners, but mainly in period 2 for intermediate and advanced learners. Similar distribution patterns were observed for planning, reading prompts, task interpretation and revising across proficiency levels. The first three activities mainly took place in period 1, whereas revising mainly occurred in period 3.

Tillema (2012) also found that L2 proficiency could, to some extent, modulate the temporal distribution of certain cognitive activities. L2 proficiency in this study was measured by a vocabulary test. Results showed that participants with higher L2 proficiency did more process planning (i.e. monitoring and giving metacomments) at the beginning of writing than average, while participants with lower proficiency did less than average. However, L2 proficiency did not seem to modulate the temporal distribution of content planning, formulation, reading and evaluating one’s own text, reading the assignment, or revising.

In Gánem-Gutiérrez and Gillmore's (2018) previously discussed study, participants with lower L2 proficiency (as measured by a C-test) were found to engage more in using external resources and less in text construction in terms of both frequency and duration. More proficient participants tended to revise more frequently than less proficient ones, but not more in terms of duration. When the role of time was considered, the researchers found no relationship between L2 proficiency and the temporal distribution of text construction, revising, pausing, rereading or using external recourses.

As for writing behaviours, two studies conducted by Barkaoui (2016, 2019) investigated the distribution patterns of pausing and revision behaviours during writing between L2 English writers with high and low proficiency. L2 proficiency in Barkaoui (2016) was operationalised as whether the participants were pre- or post-admission students of the university, while in Barkaoui (2019), the participants were divided into two proficiency groups based on whether or not they had achieved 83 or above on the Test of English as a Foreign Language iBT, or Band 6.5 or above on the International English Language Testing System. It was observed that, in terms of pausing behaviours, participants with low L2 proficiency were less fluent, measured by the number of characters produced per minute, and they paused more frequently than those with high proficiency for the whole writing session. When the timing of pauses was taken into account, the two groups differed only in interval 1 in terms of pause frequency and duration; no differences were observed in intervals 2 and 3. In terms of revision behaviours, the researcher found that, for the whole writing session, participants in the high proficiency group revised less frequently than those in the low proficiency group. However, similar distribution patterns for revisions across the three intervals were found for the two proficiency groups. Both groups made more revisions overall in interval 2 than in intervals 1 and 3; regardless of the level of proficiency, more pre-contextual revisions (i.e. those made at the end of the current text) occurred in interval 2, and more contextual revisions (i.e. those made to already-written text) occurred in interval 3.

To sum up, L2 proficiency explains a relatively large proportion of variance in L2 writing quality. In terms of the relationship of L2 proficiency to L2 writing processes, it seems that writers with different levels of L2 proficiency do indeed produce texts in different ways when writing processes are analysed as a whole. However, studies have yielded mixed findings with regard to what the differences are. For example, L2 writers with higher proficiency were found to engage more in revising in Roca de Larios et al. (2008) than those with lower proficiency, while conflicting conclusions were drawn in Barkaoui (2016), Chenoweth and Hayes (2001) and Tiryakioglu et al. (2019). Gánem-Gutiérrez and Gillmore (2018) also found that higher-proficiency L2 writers made more revisions, but only in terms of frequency rather than duration.

The picture is even less clear when comparing the findings of the studies that looked into how L2 proficiency related to the distribution of writing activities across various writing stages. First, some studies (Barkaoui, 2019; Roca de Larios et al., 2008; Tillema, 2012) confirmed the relationship of L2 proficiency to the temporal distribution of writing activities, while others (Barkaoui, 2016; Gánem-Gutiérrez &

Gilmore, 2018) found that L2 proficiency played a limited role. Second, even among those studies where significant relationships were observed, no real consensus has been reached on how L2 proficiency relates to what activity and in which stage of writing.

Possible explanations of the discrepant findings may be related to the way in which L2 proficiency was operationalised and writing processes were elicited. For example, L2 proficiency was established by participants' level of education in Larios et al. (2008), while other studies adopted more standard measures, such as proficiency tests, to determine L2 proficiency. In terms of data collection methods, Larios et al. (2008) and Tillema (2012) adopted think-aloud procedures, Barkaoui (2016, 2019) and Tiryakioglu et al. (2019) employed keystroke logging, while Gánem-Gutiérrez and Gillmore (2018) utilised eye-tracking techniques. Clearly, the role of L2 proficiency in writing processes and whether it modulates process-product relations in L2 writing deserves further exploration.

2.2.5 Summary

To sum up, despite a substantial body of literature on L2 writing processes, it was not until recently that L2 researchers began to move away from conceptualising writing as only a dynamic process and to analyse the writing process from a dynamic perspective by taking into account the variable of time during writing. Clearly, more studies are called for in the future to investigate how L2 writing behaviours and the cognitive activities associated with them are distributed across different stages of writing. Another under-explored area in L2 writing is how the cognitive processes involved in writing predict text quality. More studies are needed to examine these links by taking into account the temporal dimension, as text quality does not only depend on how frequently a writing activity occurs in the whole session of text production, but also on the timing of the activity in the writing process. Future research on these links would also benefit from analysing writing behaviours from a multi-dimensional perspective. For instance, as shown in the studies discussed above (i.e. Barkaoui, 2016; Khuder & Harwood, 2015; Stevenson et al., 2006; Thorson, 2000; Xu, 2018), revision behaviours during writing vary in quality (e.g. pre-contextual and contextual revisions differ in terms of the context of occurrence); thus, each type of revision may predict text quality in a different way. In addition, it is unclear the extent to which genre influences L2 writing processes, and how more and less proficient L2 writers differ in the process of text production; further studies are warranted to look into these issues. Lastly, attention should also be paid to how genre and L2 proficiency modulate the relationship between writing processes and L2 text quality.

2.3 Methods in writing process research

This section focuses on methods of data elicitation in process-oriented writing research. Verbal reports, including think-aloud and stimulated recall, are first discussed, followed by an overview of a more recent approach, keystroke-logging. Although a small number of studies have used eye-tracking to provide insights into reading activities during writing (e.g. Gánem-Gutiérrez & Gilmore, 2018; Révész et al., 2019; Torrance et al., 2016), no discussion of eye-tracking techniques is included in this section as the current study does not employ this technique for data collection. I will then close the section by emphasising the need for a joint application of methods to overcome the potential shortcomings of utilising a single approach (Révész & Michel, 2019).

2.3.1 Verbal reports

Think aloud, or concurrent verbal reporting, is one of the most widely used methods for collecting writing process data (Hyland, 2016; Polio, 2012). It involves participants reporting their thoughts concurrently while carrying out a writing task. This method has been used in studies in a number of areas, such as constructing models of writing (e.g. Flower & Hayes, 1981), comparing novice and expert writers (e.g. Scardamalia & Bereiter, 1987), exploring writing processes and sub-processes (e.g. Manchón et al., 2009), examining process-product relations (e.g. Van Den Bergh & Rijlaarsdam, 2001) and investigating the use of L1 in L2 writing (e.g. W. Wang & Wen, 2002).

Despite the frequent use of think-aloud protocols, there has been criticism of the validity of the method. A major concern has to do with the issue of reactivity. That is, verbalising thoughts simultaneously with task performance might constitute an additional task, thereby putting an extra cognitive load on processing mechanisms (Jourdenais, 2001; Valfredini, 2015). This resulting cognitive overload might, in turn, alter the cognitive process involved in task completion. Consequently, it has been suggested that concurrent verbal reports cannot provide a valid reflection of thoughts.

Several studies have investigated the issue of reactivity in L1 or L2 writing, but the results are mixed. Ransdell (1995) found think-alouds only affected the production rate in L1 writing. No difference was observed between think-aloud and silent conditions regarding syntactic complexity, text length and the number of clauses produced. Similarly, Levy and Ransdell (1995) reported a negligible impact of think -louds, as the participants produced similar numbers of words while talking aloud simultaneously and when they wrote in silence. Also, the participants were asked to respond to a computer-generated beep while writing by depressing a footswitch as quickly as possible whenever they heard the tone. The researchers found little difference in reaction time during planning and text generating between think-aloud and silent conditions, assuming that the think-aloud procedure did not change the cognitive process of writing.

In contrast, Janssen, Van Waes, and Van Den Bergh (1996) suggest that thinking aloud might inhibit the process of writing. Their study consisted of two experiments. In the first experiment, 20 students carried out a writing task requiring knowledge transformation (Scardamalia & Bereiter, 1987) under both think-aloud and silent conditions. Data of real-time writing behaviours (i.e. pauses at different text levels) indicating different levels of planning were captured by a computer program. In the second experiment, a similar design was adopted, but this time a knowledge-telling task (Scardamalia & Bereiter, 1987) was completed by 28 students. The writers in both experiments paused longer under the think-aloud condition, thus indicating reactivity. A higher positive correlation between pauses at different text levels and a greater amount of variance in pausing behaviours within and between individuals were identified when the writers were thinking aloud while also involved in a knowledge-transforming writing task. The researchers concluded that the think-aloud procedure was reactive because it altered the cognitive processes of writing, as revealed by comparing the pausing behaviours between think-aloud and non-think-aloud conditions. They further suggested that reactivity might increase in more complex tasks.

In an L2/bilingual setting, Yanguas and Lado (2012) found that engaging in think-alouds positively influenced accuracy and lexical complexity, but it had no effect on fluency. Yang, Hu, and Zhang (2014) reported incompatible results. Three groups of L2 English learners wrote an argumentative essay under non-metacognitive think-aloud, metacognitive think-aloud and silent conditions, respectively. The texts produced were rated in terms of complexity, accuracy, fluency and overall quality. Reactivity was not found for most measures, but both types of verbalisation impaired fluency and syntactic complexity. Metacognitive verbalisation also reduced the speed of production. Yang et al. thus concluded that the overall effect of thinking aloud was not strong, although it could have a detrimental impact on writing performance in some aspects. This conclusion was partially replicated by Yang, Zhang, and Parr (2019). Eighty-five university students in China were asked to perform a narrative writing task either silently or while verbalising their thoughts. The researchers observed that thinking aloud only inhibited two measures, i.e. fluency and lexical diversity. However, it is worth noticing that, when asking the participants to reflect on thinking aloud, the majority suggested that thinking aloud negatively affected their writing in one way or another. For example, the participants mentioned that thinking aloud was detrimental to translation processes, as verbalising thoughts occupied time and energy that would otherwise have been devoted to searching for more sophisticated lexical items.

To sum up, four studies (Levy & Ransdell, 1995; Ransdell, 1995; Yang et al., 2014, 2019) have found that think-aloud protocols have little impact on writing, whereas two others (Janssen et al., 1996; Yanguas & Lado, 2012) have argued that it led to reactivity. These conflicting results appear to indicate that the effect of the method may differ depending on a number of factors, such as type of verbalisation (metacognitive versus non-metacognitive), language of the writing task (L1 versus L2 writing tasks) and genre (e.g. letter-writing, explanation-writing, argumentative writing and narrative writing). Notably, except for Levy and Ransdell (1995) and Janssen et al. (1996), other studies only adopted product-based measures to examine reactivity. In other words, the studies did not directly address the question of whether or not think-alouds change the course and structure of the writing process. Therefore, even though most studies have concluded that the think-aloud procedure is largely non-reactive, it is questionable whether findings based on product-based measures alone can be taken as evidence for a lack of reactivity.

Other potential limitations include think-alouds not being employed with all participants (Tirkkonen-Condit, 2006). Some people may not be able to think aloud naturally while performing a task. For example, in order to avoid silence, they may simply report what they just did or what they are going to do next. In addition, thinking aloud may be embarrassing for some people, which can also pose a challenge in terms of the validity of data.

Another method for tapping into writers' cognitive processes is stimulated recall, also known as retrospective verbal reporting. It differs from think-alouds in that participants are asked to describe their thoughts during task performance after the task has been completed (Gass & Mackey, 2016). This recall is typically prompted by a stimulus, such as the text participants have produced or a video showing themselves writing. Many studies have employed stimulated recall to examine topics concerning the processes involved in producing a text, such as individual variations in writing processes (e.g. Bosher, 1998), strategies used during writing (e.g. De Silva & Graham, 2015), language issues and revision practices in collaborative writing (e.g. Brooks & Swain, 2009), how reader awareness is addressed in writing (e.g. Lindgren et al., 2011) and the effects of task complexity on writing processes (e.g. Révész et al., 2017).

Similar to the think-aloud method, there is also a reactivity concern for stimulated recall. It is possible that during recall, participants’ attention is drawn to issues which they may not have been aware of prior to the protocol. As a result, the recall may influence participants’ performance on subsequent tasks. However, Egi’s study (2008), one of the few empirical studies to address the issue, provided counter-evidence to this assumption. Three groups of learners of L2 Japanese took a pre-test on two Japanese structures and participated in two oral communicative activities on two consecutive days a week later. One day after the activities, one group of participants did a stimulated recall and reported their thoughts while watching a video of themselves taking part in the activities. The second group also watched the video but did not report their thoughts. The third group neither watched the video nor reported their thoughts. A post-test was administered after the recall or viewing session, and the participants additionally completed a delayed post-test a week later. By comparing the scores on the use of the target items on the pre-test, immediate post-test and delayed post-test, the researcher found no significant differences between the three groups, indicating that stimulated recall was non-reactive.

Another potential threat to the validity of stimulated recall has to do with veridicality, i.e. participants may report their cognitive processes during task completion inaccurately and/or incompletely during the protocol. Indeed, owing to memory decay, participants are less likely to accurately and fully recall the process in which they engaged during performance as the time increases between the event to be reported and the reporting itself. The stimulated recall method can only be justified when it is carried out shortly after the task is completed, when the cues which allow the effective retrieval of thoughts are still in short-term memory (Ericsson & Simon, 1993). Only one empirical study to date has explored the veridicality issue in the context of writing. Abdel Latif (2019) considered this issue by comparing stimulated recall with think-alouds. Thirty undergraduate students of L1 Arabic completed an argumentative writing task in L2 English while verbalising their thoughts about their writing processes. Immediately after the writing task, each participant participated in a retrospective interview session with the text he produced as the stimulus. The verbal protocols were analysed in terms of pre-writing stage, while-writing planning, L1 use, composing problems and problem-solving, and text reviewing and changing. The stimulated recall method was found to elicit fewer and less detailed data than think-alouds about writers' writing processes. The researcher also found that the two methods may reveal data about writing processes from different aspects. Stimulated recall was more useful for tapping into why writers employed certain writing strategies, while thinking aloud provided more information about how the strategies were used. It seems that the stimulated recall procedure may pose a greater threat to validity due to issues with accuracy and completeness. However, it is questionable whether a stimulated recall interview did indeed generate less rich data than a think-aloud procedure due to a potential pitfall of the study. The retrospective interview on writing processes was prompted by the draft rather than a recording of the writer's actual text-production process. Written texts may be less effective in eliciting writing process data as they do not normally entail traces of how a text was produced, such as pauses and revisions. In addition, as the participants took part in both think-aloud and stimulated recall procedures, it is likely that concurrently verbalising thoughts during writing may have altered the participants' writing processes, consequently affecting their behaviours in the follow-up stimulated recall interview.

In sum, verbal reporting is and is likely to continue to be a common method for eliciting data on writing processes. However, given the potential dangers of verbal reports to validity, both think-aloud and stimulated recall must be carried out with care, and the reported data have to be interpreted with caution.

2.3.2 Keystroke logging techniques

Given the potential drawbacks of verbal reports, some researchers favour observation as an alternative, albeit indirect, method for studying writing processes (Spelman Miller, 2005). In the early years of research, direct observation and video cameras were used for recording online writing behaviours (e.g. Matsuhashi, 1981; Zamel, 1983). However, the physical presence of a researcher or video camera could make the participants anxious and, consequently, alter their writing behaviours.

With the advent of computer technology, keystroke logging programs have been developed, making it possible to observe real-time writing processes with less interference in participants' behaviours (Leijten & Van Waes, 2006). Keystroke logging refers to a computer recording technique that involves 'the use of a resident software program that appears to the writer as a word processor with normal text editing functions' (Spelman Miller, 2005, p. 300). The rationale for this technique lies in the concept that visible behaviours during writing, such as pauses and revisions, 'reveal traces of the underlying cognitive processes' (Leijten & Van Waes, 2013, p. 360).

Early attempts to record writing behaviours by keystroke loggers began in the 1980s. Bridwell, Sirc, and Brooke (1985) were among the first researchers to use a logging tool, *Recording Wordstar*, for capturing novice writers' revision behaviours.

Other early studies include Bridwell-Bowles, Johnson, and Brehe (1987), Flinn (1987a, 1987b) and Lutz (1987). In subsequent decades, keystroke logging software has started to gain in popularity to explore various writing-related topics, such as the cognitive activities in which different types of writers engage (e.g. Barkaoui, 2016; Tillema, 2012; Van Waes et al., 2014), the use of additional materials during writing (e.g. Leijten et al., 2014; Leijten, Van Waes, et al., 2019), the effect of certain writing strategies (e.g. De Smet et al., 2014) and the processes involved in translation (e.g. Schrijver et al., 2016). Some of the most frequently used keystroke loggers today include *JEdit*, *ScriptLog*, *Inputlog* and *Translog*.

The increased use of keystroke logging programs offers a great number of advantages in investigating writing processes. The most apparent advantage is that the logging method is largely non-obtrusive to writers (Van Waes et al., 2009), thus overcoming the reactivity issue, a potential shortcoming of verbal reporting. With every keystroke and mouse movement registered, keystroke logging programs are able to generate a relatively complete and detailed report on writing behaviours and so are less likely to be criticised for a lack of veridicality. Keystroke logging also provides a clear picture of the dynamics of the writing process, as online writing behaviours are logged in relation to a time stamp (Leijten & Van Waes, 2006). In addition, although most research employing keystroke logging has mainly analysed process data with quantitative approaches, keystroke logging files are also useful for conducting more qualitative, linguistic analysis of the writing process (Van Waes et al., 2015). For example, Leijten et al. (2019) used *Inputlog* to investigate pauses and revisions in relation to textual context.

The keystroke logging methodology also has some practical advantages. First, the program runs in the background, ensuring the writing process is recorded in a comparatively naturalistic setting, rather than using a video camera (Van Waes & Leijten, 2006). In addition, writers are not required to have prior knowledge or receive special training before using the logging program, as the software is similar to a word processor on the computer.

Apart from being a powerful research tool, the logging software is found to have some value for L1 and L2 writing instruction. For instance, in Vandermeulen (2018), 10th graders in the Netherlands increased their performance in L1 writing after receiving process-oriented feedback by comparing their writing process to that of students who produced better texts. Lindgren and Sullivan (2003) reported that replaying the writing process allows L2 writers to reflect more on the extra-linguistic aspects of writing, as it stimulates writers to evaluate what they did and thought about during composition.

In spite of its numerous advantages, keystroke logging is not immune to criticism. One major problem concerns data interpretation (Abdel Latif, 2008). Logged keystrokes can be hard to interpret because they do not provide a direct reflection of cognitive activities during the writing process (Lindgren, 2005). Additionally, logging data alone reveal no information about reading behaviours during writing (Révész & Michel, 2019), such as what writers look at (e.g. recently produced words or sentences away from the point of inscription) and how they read (e.g. reading in a linear manner or moving back and forth) during pauses (Leijten & Van Waes, 2013).

2.3.3 Summary

To date, studies employing verbal reports (i.e. think-aloud and stimulated recall procedures) have generated a great amount of valuable information on writers' cognitive processes during writing. However, the validity of both approaches has been challenged given the potential threats of reactivity and veridicality. The increasing use of keystroke logging techniques in writing studies has helped to obtain more detailed data on real-time writing processes, but the exclusive use of this technique provides no insights into the cognitive activities that underlie writing behaviours, such as pausing and revision behaviours (Baaijen et al., 2012; Révész et al., 2019). Pauses with similar lengths or revisions made at the same linguistic unit level might link to different cognitive activities. Similarly, the same type of cognitive activity may occur when writers are pausing between different textual unit boundaries.

A possible solution to overcome the limitations of these methods is to combine keystroke logging techniques with verbal protocols. For example, Van Weijen (2008) studied the writing processes of 20 teenagers when writing essays in L1 Dutch and L2 English. In addition to asking the participants to report their thoughts during writing, the researcher also used *Inputlog* to record the participants' online writing processes. This combination of methods generated a richer data source for analysis, with thinking aloud shedding light on conscious cognitive activities during writing and keystroke logs providing accurate information on typing and revision behaviours. In

Révész et al. (2017), *Inputlog* and stimulated recall interview were used together to examine the effect of task complexity on the writing behaviours of advanced L2 writers of English. By triangulating keystroke logging data with stimulated recall comments, the researchers concluded that longer pauses between higher textual units tended to indicate that participants engaged in higher-order writing processes, such as planning content. The joint use of verbal protocols and keystroke-logging in both studies allowed for more valid conclusions to be drawn on both explicit and implicit writing processes than the use of a single a method would have revealed. Combining multiple data collection approaches could, indeed, be an auspicious practice to overcome the limitations associated with the use of a single method. Findings obtained by triangulating data from different sources would enable researchers to gain a more comprehensive understanding of writing processes.

2.4 Rationale and research questions of the thesis

In spite of the considerable amount of research on L2 writing processes in recent decades, many issues still need further investigation. To begin with, the role of time in L2 writing processes has not yet been fully explored. The findings of a few studies that have attempted to address this gap (Barkaoui, 2015, 2016, 2019; Roca de Larios et al., 2008; Tillema, 2012; Van Weijen, 2008; Xu & Qi, 2017) have painted a mixed picture regarding how cognitive activities and writing behaviours are distributed during writing. For example, Roca de Larios et al. (2008) found more revisions in the final stage of writing, Barkaoui (2016) observed that revisions mainly occurred in the middle stage, while Tillema (2012) reported that revisions were equally distributed across all stages. Even less is known about how the distribution patterns relate to other factors, such as genre and writers' L2 proficiency. In addition, there is a major limitation concerning the data collection methodology of these studies. The majority of them employed a single method, such as verbal reporting or keystroke logging, to elicit data on writing processes.

Second, there is a paucity of studies on process-product relations in L2 writing. Exploring this link is important because it can shed some light on how a good text is produced. Although most L2 writers are and will continue to be judged on their texts in many real-life cases, analysing texts alone may not adequately explain why some texts are of higher quality than others. Empirical evidence shows that writing processes can, at least partially, predict text quality (e.g. Rijlaarsdam et al., 2012). Therefore, research on product-process links may provide insights into how text quality benefits from certain types of writing activity and at which moments of writing. Understanding how a good/ poor L2 text is produced may inform language instructors of the difficulties that writers encounter and help them to identify strategies that may be worth teaching. In addition, the few existing studies on process-product relations in L2 writing have failed to yield consistent findings. This may be due to the different ways in which writing processes (frequency/ percentage counting versus temporal analysis) and products (holistic rating versus objective linguistic measures) have been analysed.

Third, the extant literature on L2 writing has a predominant focus on writing in alphabetic languages. So far, little research has looked at the cognitive processes in which non-alphabetic language writers engage. Unlike an alphabetic language, the orthographic system of a non-alphabetic language suggests no direct link between pronunciation and written form (Zhang, 2017), making transcription skills more crucial during text production (Wong, 2018). It is anticipated that, due to partial acquisition in sound-form mapping, L2 writers of a non-alphabetic language might pause after typing phonetic readings (e.g. Pinyin of Chinese, Rōmaji of Japanese) to search for desired logographic or syllabic symbols among homophones. For the same reason, they are expected to make frequent changes to phonetic readings before these are converted into logographic or syllabic symbols. Pauses and revisions of these types are most likely to occur in the middle stages of writing when L2 writers engage most in formulating the text. In addition, higher L2 proficiency is assumed to lead to greater automaticity in sound-form mapping, which is anticipated to manifest as fewer pauses being devoted to looking for symbols among homophones and fewer changes being made to phonetic readings. Previous studies on alphabetic language writing have provided limited information about these predictions, as transcribing a text in an alphabetic language is generally considered an automated process that requires few attentional resources for adult L2 writers. Thus, to build a fuller understanding of L2 writing, it is important to investigate the processes involved in non-alphabetic writing and the influence that a logographic writing system may have on process.

In light of the research gaps mentioned above, the primary goal of the thesis was to explore the writing behaviours and associated cognitive activities of L2 writers of

Chinese from a temporal perspective. It built and expanded on existing process-oriented research by further investigating online writing behaviours and the cognitive activities underlying them at five different stages of writing. The current thesis also aimed to explore the relationship between writing processes and text quality via a finer-grained analysis that took into account both temporal and multi-dimensional perspectives of writing processes. In addition, this thesis extends previous studies by examining how genre and L2 proficiency relate to the temporal distribution of writing behaviours and associated cognitive activities, and process-product relations. Notably, due to the limited research on non-alphabetic language writers, the current thesis also includes data collected from L1 writers of Chinese to determine whether the patterns observed are unique to L2 Chinese writing.

On the methodological front, the current thesis also presents some innovations. As discussed in the previous section, the use of a single research method for data elicitation inevitably has some drawbacks. To address this issue, the current thesis adopted a combination of data collection methods: writers' writing behaviours (i.e. speed fluency, pausing and revision) were recorded by a keystroke logging program, while the cognitive activities associated with pausing and revision behaviours were investigated through stimulated recall comments. This triangulation of data was expected to yield more valid and accurate results about L2 writing processes than relying on a single data source, such as think-aloud procedures.

To address these gaps, five research questions were formulated:

Research Question 1: To what extent do stages of writing affect cognitive processes in L2 Chinese writing, as reflected in (a) writing behaviours and (b) stimulated recall comments?

Research Question 2: To what extent does genre affect the cognitive processes in L2 Chinese writing, as reflected in (a) writing behaviours and (b) stimulated recall comments?

Research Question 3: To what extent are L2 proficiency and the behaviours in L2 Chinese writing related? To what extent do the stage and genre of writing mediate these relationships?

Research Question 4: To what extent do writing behaviours predict L2 Chinese text quality?

Research Question 5: Do these relationships differ in L1 Chinese writing?

Chapter 3 Methodology

This chapter 3 describes in detail the methodology employed to address the research questions. The overall research design is first presented. It is followed by a description of the participants and instruments, an overview of the recruitment and ethical procedures and an explanation of the procedures for data collection, coding and scoring. The last section of the chapter explains the statistical analyses utilised.

3.1 Research design

The current study employed a repeated-measures design to investigate non-alphabetic language writers' writing behaviours and the cognitive activities associated with them. Data were collected from 32 writers of L2 Chinese, and 32 L1 writers of Chinese provided baseline data. All participants took part in two sessions on two separate days. In each session, they were asked to perform two writing tasks, one argumentative and one narrative, on a computer, resulting in 256 writing performances in total. The order of the four tasks was counterbalanced across participants and sessions, as shown in Table 3.1. The participants' pausing and revision behaviours were recorded by keystroke logging software, *Translog 2.0* (Carl, 2012). All participants completed a background questionnaire (see Appendix A) with their demographic information in Session 1 after the writing tasks. In Session 2, all participants were invited to a stimulated recall interview immediately after writing. In the interviews, they were asked to report their thoughts when performing the last writing task. In other words, for each group, I collected 16 verbal reports on argumentative writing performances and 16 on narrative writing performances. L2 writers of Chinese additionally completed an L2 proficiency test after the stimulated recall in Session 2.

Table 3.1 *Task order across participants and sessions (first eight participants)*

Participant ID	Writing tasks			
	Session 1		Session 2	
1	Argumentative 1	Narrative 1	Argumentative 2	Narrative 2
2	Argumentative 1	Narrative 2	Argumentative 2	Narrative 1
3	Argumentative 2	Narrative 1	Argumentative 1	Narrative 2
4	Argumentative 2	Narrative 2	Argumentative 1	Narrative 1
6	Narrative 1	Argumentative 1	Narrative 2	Argumentative 2
7	Narrative 1	Argumentative 2	Narrative 2	Argumentative 1
8	Narrative 2	Argumentative 1	Narrative 1	Argumentative 2
9	Narrative 2	Argumentative 2	Narrative 1	Argumentative 1

3.2 Participants

According to G*Power (Faul et al., 2007), the current study required 36 participants in each group to identify medium-size relationships, given the repeated-measure design and the number of observations. However, only 16 participants were needed in order to cover all possible combinations of writing tasks with one argumentative and one narrative essay in each session. Therefore, the desired figure was reduced from 36 to 32 (two groups of 16 participants) to ensure that the writing tasks were fully counterbalanced.

Thirty-two adult L2 writers of Chinese studying or working in London participated in the study. To control for L1 orthography effects, only L1 users of an alphabetic language were recruited, including 27 English, 2 German, 1 French, 1 Italian and 1 Polish L1 users. The group consisted of 19 males and 13 females. They were aged between 19 and 41 years, with an average age of 26.62 years (*SD* = 6.52). They varied in their L2 Chinese proficiency, which was determined by a proficiency test (see Section 3.3.4). The participants had an average of 58.84 months (range 15 to 120 months) of previous Chinese study, with 12 participants starting to learn Chinese in high school and the rest at university. The participants reported that they had used Chinese for an average of 25.78 hours per week (range 0 to 67 hours) in the past three months, of which only 1.09 hours (range 0 to 6 hours) was devoted to writing (including typing and handwriting). Most participants (*N* = 24) interacted with L1 users of Chinese at least once per week, and the majority (*N* = 31) had experience of studying or staying in a Chinese-speaking country or region (range 0.5 to 96 months).

Baseline data were provided by 32 L1 writers of Chinese who were studying or working in London at the time of the study. They were 5 males and 27 females, and their ages ranged between 18 and 32 years, with an average of 24.72 years (*SD* = 3.54). The participants had on average been in the UK for 6.37 months (*SD* = 2.32). They had been using Chinese on a daily basis in the previous three months, with a reported estimation of 39.16 hours per week (range 11 to 100 hours). Of that, an average of 9.33 hours per week (range 1 to 30 hours) was spent writing in Chinese (including typing and handwriting). All participants were L2 users of English, and four participants had experience of other foreign languages, including Korean, French, Japanese and Czech. The reported average time that the participants spent using the foreign language(s) was 25.78 hours per week (range 4.5 to 69 hours), with 5.11 hours (range 0.5 to 20 hours) on writing. A *t* test was conducted to compare the hours spent writing in Chinese and foreign language(s), which confirmed that the participants wrote significantly more in Chinese (p = .01). This was to ensure that the L1 writers' written output in Chinese would not be considerably influenced by English conventions, although this batch of data was collected in an English-speaking country.

Both groups of participants were familiar and comfortable with producing essays using the Pinyin input method on a computer.

3.3 Instruments

3.3.1 Writing tasks

The participants were asked to complete two argumentative and two narrative writing tasks on a computer. Two prompts for each genre were used to elicit participants' writing performances so as to avoid a potential prompt effect. Each writing prompt contained two parts, information about the topic and task instructions. Given the variation in the L2 writers' Chinese proficiency, information about the topic was given in both Chinese and English, but task instructions in English only. The participants were given 30 minutes to complete each writing task. No word limit was set, given that the L2 writers varied greatly in terms of their Chinese proficiency. However, in order to elicit effective writing performances, the participants were required to produce a complete essay within the time limit. The participants were allowed no planning time and could not use dictionaries or refer to additional materials during writing. On average, L2 writers of Chinese spent 29 minutes (*SD* = 3.67) on task completion, and L1 writers 23 minutes (*SD* = 5.36). Table 3.2 lists one writing prompt from each genre. See Appendix B for a full version of the writing prompts.

Developing the writing prompts involved four steps. First, a pool of 20 writing prompts (10 for each genre) was created by searching through a variety of resources, such as standardised test preparation materials, text books and the prompts used in writing competitions. Second, each potential prompt was rated in terms of how difficult it was for L2 writers of Chinese and how engaging it was by five independent raters using a five-point Likert scale. The raters were L1 users of Chinese and had a bachelor's degree or above in Chinese linguistics or education. They were also all qualified teachers of L2 Chinese with over 500 hours of teaching experience. Third, two writing prompts, which were at a medium level of difficulty (2.95 out of 5) but highly engaging (4.05 out of 5), were chosen for each genre. The inter-rater reliability was good for argumentative writing prompts and acceptable for narrative writing prompts. For the argumentative writing prompts, was .85 for difficulty and .89 for how engaging the topics were, while for narrative writing prompts, the figures were .72 and .79. I hoped that such prompts would enhance the participants' degree of involvement in the writing tasks. Lastly, I translated the topic information of the selected prompts into both Chinese and English. The final version of the writing prompts was reviewed by a professional translator.

Table 3.2 *Sample writing prompts*

Argumentative writing prompt

Write about the following topic:

如今，几乎人人都有手机。一些人认为，手机给生活带来了便利。另一些人认为，我们每天花太多时间看手机，也因此带来了很多问题。讨论这两种观点，并对使用手机给出你的看法。

Almost everyone has a mobile phone nowadays. Some people think that a mobile phone brings convenience to our life. Others say we spend too much time on them, and there are many problems associated with them. Discuss both views and give your own opinion on the massive use of mobile phones.

Give reasons for your answers and include any relevant examples from your own knowledge or experience. You have 30 minutes to finish the task. There is no word limit for the essay, but you should write a complete essay.

Narrative writing prompt

Write about the following topic:

回忆童年的时候，我们经常会想起一些对我们产生重要影响的事。叙述一件发生在你童年时的重要事。请写出这件事的内容，这件事发生的时间、地点，你看见了什么、做了什么，还有哪些人参与，并谈谈这件事对你的影响。

When we reflect upon our childhood, we often come back to a few key events that had a major impact on us. Write a story about one of those defining events from your childhood. You may include what the event was, when and where it happened, what you saw and did, who else was involved, and explained how it impacted on you.

Be sure to develop your story by including relevant examples and specific details. You have 30 minutes to finish the task. There is no word limit for the essay, but you should write a complete essay.

3.3.2 Chinese input methods for computer users

The logographic nature of the Chinese writing system makes it impossible to have a one-to-one match between the characters and keys on a keyboard. Thus, multiple steps are needed for computer users to input Chinese characters using a keyboard. In general, there are two ways to type Chinese: the shaped-based method and the phonetic-based method.

For shape-based input methods, such as *Canjie* and *Wubizixing*, each key is assigned to at least one component of a character. Users of the methods have to press the keys with the components of the character they want to type in the right sequence in order to obtain the desired character. This requires the user to learn by heart how Chinese characters are divided into different components in a certain shape-based input method. It can take hours for users to familiarise themselves with the rules of character division for a shape-based input method; therefore, this method is not generally favoured by L2 writers of Chinese (Kang, 2011).

Phonetic-based (i.e. Pinyin and *Zhuyin fuhao*[2]) input methods, on the other hand, require the user to enter pronunciations, which are converted into Chinese characters. The Pinyin input method is the most typical phonetic-based input method among L2 writers of Chinese and in the People's Republic of China (Xiao et al., 2007). The input method requires the user to first enter the Pinyin, phonetic symbols for Chinese characters written in the Latin alphabet, of the character. Then, the user selects the

[2] *Zhuyin fuhao* is a phonetic system which encodes characters into symbols that look like simple Chinese characters rather than letters in the Roman alphabet. It is mainly used by Chinese speakers in Taiwan.

character from a list containing those characters with that pronunciation, by pressing either the space bar for the first option on the list, such as *贤* (virtuous) in Figure 3.1, or a numeric key, such as pressing 2 for the second option *限* (limit), 3 for the third option *西安* (Xi'an, name of a city in China), 4 for the fourth option *鲜* (fresh) or 5 for the fifth option *嫌* (resent) .

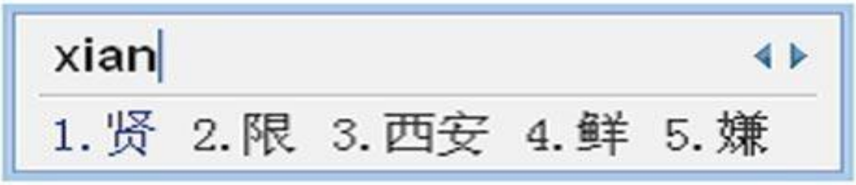

***Figure 3.1** A screenshot of typing using the Pinyin input method*

However, the user is not restricted to input Pinyin one character at a time. In other words, the input method allows the user to enter the Pinyin of more than one character (e.g. a multi-character word, phrase or clause) before selecting characters. Figure 3.2 provides an example of typing the Pinyin of multiple characters before conversion is required. The Pinyin of a four-character phrase (*绝妙创意*, an excellent idea) has been entered. Once the person typing presses the space bar, four characters will appear on the screen at the same time. Given the dual step transcription process when using the Pinyin input method, writers can pause as they finish typing the Pinyin of the character(s) and are searching for the desired character(s) among homophones. For the same reason, revisions can be made to both Pinyin and character(s) during text production.

Note: *绝妙创意* = an excellent idea, *绝妙* = excellent, *觉* = feel, *绝* = absolute, *崛* = rise

Figure 3.2 Entering the Pinyin of multiple characters

Given its popularity among L1 and L2 writers of Chinese, the Pinyin input method was used by the participants to produce essays on a computer in the current study. The *Sogou* Pinyin input method was chosen as it is the most widely used Pinyin input method on digital devices (Z. Wang et al., 2018).

3.3.3 Stimulated recall

Stimulated recall was used in the current study to collect data on the cognitive activities in which the participants engaged during writing. The rationale for making this decision was that stimulated recall would not interfere with task performance, as verbalisation was carried out after writing. Verbalising thoughts while writing, such as thinking aloud, could intrude into the writing process for some participants, especially for L2 writers (Barkaoui, 2015), if they were asked to report their thoughts in another language while writing in Chinese.

However, as mentioned in Chapter 2, stimulated recall also carries some potential risks. It can be reactive, as it might draw participants' attention to issues which they might not have been aware of prior to the interview, thus altering their writing processes on the tasks that follow. There is also a concern of veridicality over stimulated recall, as a verbal report might not be able to fully capture the participant's thoughts during writing owing to memory decay (Gass & Mackey, 2016). To mitigate issues with veridicality, for the current study, each participant participated in a stimulated recall interview immediately after they finished writing, when the cues which allowed effective retrieval of thoughts were still in short-term memory (Ericsson & Simon, 1993). The interviews were based on the second writing task they performed in Session 2, so as to avoid reactivity.

The stimulus used to prompt recall was a recording of the participants' keystrokes, made by *Translog 2.0,* when they were performing the last writing task. The purpose of the interview was first explained to the participants in everyday language. Next, oral instructions, adapted from Barkaoui (2015) and Gass and Mackey (2016), were given on how to perform the stimulated recall task. The participants were told that they were going to watch a video on how they wrote the last essay. While watching, I paused the video from time to time and asked them to say aloud what they were thinking while writing the essay. They were informed not to report what they were thinking now (i.e. at the time of the interview). They were also encouraged to pause the recording whenever they wished to share what they were thinking during writing.

After the instructions were given, a participant watched his/her text appear on the screen, and I stopped the recording whenever a pause or a revision occurred and asked the participant to report the thoughts s/he had when completing the writing task. The questions used to elicit the participants' thoughts were open-ended, such as, 'What (else) were you thinking at that moment/ that point/ right then?' No concrete responses were made to the participants' responses, apart from 'I see, good, ok, hmm', i.e. ‘verbal nods’. However, if a response was considered too brief, I asked the participant to elaborate on their answer. If the participant started to talk about his/her current thoughts, I steered the exchange back to the time of writing by asking, 'Were you thinking that when you were writing the essay?' If the participant was not able to recall their thoughts while writing, the response was accepted, and no other question was asked concerning that pause or revision. The participants could replay certain segments of the recording if they wished. A video camera was used to record the stimulated recall interviews to capture both the participants' verbal comments and spatial movements, such as pointing to the screen. The interviews were conducted in English for L2 writers of Chinese, given that they were L1 users or advanced L2 users of English, and in Mandarin Chinese for L1 writers of Chinese. However, the participants were allowed to switch between English and Chinese whenever they felt like it. For example, L2 writers of Chinese reported their thoughts partially in Chinese where the responses addressed the use of language in the essay.

3.3.4 L2 proficiency test

A cloze test was used to measure participants’ proficiency in L2 Chinese. The test consisted of nine short passages from the reading section of the Test of Chinese as a Foreign Language (TOCFL), administered by the Taiwan Ministry of Education.

Among the nine passages, three were from TOCFL Band A, equivalent to Level A in the Common European Framework of Reference (CEFR) for Languages, three from Band B, equivalent to CEFR Level B, and three from Band C, equivalent to CEFR Level C. They were ordered from easy to difficult. Each passage contained five blanks, and there were three or four possible options for each blank. The internal consistency of the test was good (Cronbach's alpha = .89). L2 writers of Chinese were asked to choose the best answer for each blank. The time allowed for the test was 30 minutes. Table 3.3 provides two sample passages (the easiest and the most difficult ones) from the cloze test. See Appendix C for a full version of the proficiency test.

The rationale for using a cloze test was that cloze test scores have been shown to positively correlate with those of standardised proficiency tests (e.g. Aitken, 1977; Bachman, 1986; Fotos, 1991; Hanania & Shikhani, 1986), although some researchers (e.g. Alderson, 1979; Gaillard & Tremblay, 2016) argue that a cloze test taps more into linguistic skills than comprehension. There was also a practical advantage in employing a cloze test, as it was less time-consuming (Gellert & Elbro, 2013). This was particularly important for this study, as the test was administered after the writing and stimulated recall sessions, which had already taken two hours and a half.

Table 3.3 *Sample passages in the cloze test*

Passage 1:

九月五日是小红的 1.____。大家都 2.____她庆祝。她 3.____到很多礼物。所以，她今天非常 4.____。她希望明年能 5.____德国去玩。

1. A. 生活　B. 生日　C. 星期日　　　Answer:__________

2. A. 帮　B. 让　C. 对　　Answer:__________

3. A. 收　B. 寄　C. 借　　Answer:__________

4. A. 热闹　B. 舒服　C. 高兴　　Answer:__________

5. A. 到　B. 去　C. 来　　Answer:__________

Passage 9:

每每回顾生命历程，总能发现某些当年所 1.____的意外和经历的曲折，后来好像都转为一种能量和养分。若 2.____这些意外与曲折，我似乎就不会在人生的路上与难得的人、事相遇；而这些人、那些事在时间渐渐 3.____一切后，只留下了由欢笑与泪水交织而成的一股暖意。曾经的悔恨和不满仿佛都已 4.____。我经常拿这些人与事和朋友分享，他们鼓励我将之化为文字，5.____自己带着回忆走进棺材，5.____集结成册，当我有一天什么都不记得时，至少还有人帮我记得这些人、那些事。

1. A. 盲从　B. 遵循　C. 叫好　D. 遭遇　　Answer:__________

2. A. 未　B. 甭　C. 勿　D. 非　　Answer:__________

3. A. 冲淡　B. 裁剪　C. 消耗　D. 剥削　　Answer:__________

4. A. 香消玉殒　B. 云消雾散　C. 卷土重来　D. 魂飞魄散

Answer:__________

5. A. 为了…因此…　B. 就算…也能…　C. 总得…进而…　D. 与其…不如…

Answer:__________

3.4 Ethical issues and participant recruitment

To adhere to the research ethics guidelines of the Institute of Education, University College London (UCL), an ethics application to conduct the research was submitted to and approved by the department prior to data collection.

As regards participant recruitment, an invitation containing the overall information of the study was circulated via various channels. To recruit L2 writers of Chinese, I approached the administrators of Chinese language programmes at several universities in London to disseminate the research invitation, together with posting the recruitment information at language events, language courses and on social media platforms. To recruit L1 writers of Chinese, the research invitation was dispersed across different UCL departments with the assistance of departmental administrative staff members. The research information was, in addition, advertised on social media platforms. Snowball sampling was also employed for participant recruitment. That is, I encouraged potential participants to introduce to me other willing participants who met the criteria through their personal networks.

Interested potential participants were given an information sheet about the study (see Appendix D). They were given sufficient time to read it and allowed to ask any further questions about the study before they decided whether or not to consent to participate. Those who agreed to participate were then asked to sign a consent form (see Appendix E), which provided information about their rights and responsibilities. Importantly, the participants were informed that they were free to withdraw from the study at any time without giving a reason and without it affecting any benefits that they were entitled to. Any data provided by the withdrawn participants were destroyed. Data collected from the participants was stored in line with UCL's Information

Security Management Policy. Written texts and interview recordings were kept on a password-protected laptop and on portable hard drives. The proficiency tests and background questionnaires were stored in a locked cabinet.

3.5 Data collection

Prior to collecting data, the procedure of the experiment was tested with two L1 writers of Chinese. Two issues arose from piloting. First, both writers reported that it was extremely difficult to scroll in the writing window of *Translog 2.0* if they wanted to produce a longer text. Secondly, the *Sogou* Pinyin input method was designed to memorise the user's input pattern. As a result, when writing with the input method shaped by the typing habit of previous users, the current user might have to pause longer to select the desired Chinese character(s) or delete more often because the character(s) needed was/were not always the first item(s) in the list. To alleviate these limitations, the font size of the writing window was reduced so as to accommodate up to 1,200 Chinese characters without scrolling. In addition, the *Sogou* Pinyin input method was reset every time before a writing task to clear the input pattern in its memory.

The participants attended two individual sessions on two separate days in a quiet room at the UCL Institute of Education. In the first session, they wrote one argumentative and one narrative essay on a computer. Their writing behaviours were recorded by keystroke logging software, *Translog 2.0*. Although initially developed for translation studies, *Translog* was used in the current study to record participants' writing processes as it was the only keystroke logging software to date that could record non-alphabetic languages (i.e. Chinese and Japanese). The participants completed a background questionnaire after writing. In the second session, the participants first wrote another argumentative and another narrative essay on a computer with their writing behaviours recorded using *Translog 2.0*. Immediately after writing, they took part in a stimulated recall interview, during which they were asked to describe their thoughts while writing, prompted by the replay of a keystroke logging recording based on the last writing task they performed. L2 writers of Chinese additionally completed a proficiency test after the interview. There were short breaks between activities (see Figure 3.3 for a summary of the data collection procedures). The whole data collection session took around four hours for L2 writers of Chinese and three hours for L1 writers. Participants received a small financial reward for their participation. Two L2 writers of Chinese additionally took up my offer to engage in three hours of free Chinese tutoring, which I offered as a supplement to the financial reward.

Session 1	Session 2
Writing Task 1 (30 minutes)	Writing Task 1 (30 minutes)
Break (5 minutes)	Break (5 minutes)
Writing Task 2 (30 minutes)	Writing Task 2 (30 minutes)
Break (5 minutes)	Break (5 minutes)
Background questionnaire (10 minutes)	Stimulated recall (20 to 60 minutes)

Break (5 minutes)

L2 proficiency test (L2 writers only, 30 minutes)

Figure 3.3 Data collection procedures

3.6 Data analysis

3.6.1 Division of writing stages

Writing behaviours and associated cognitive activities in this study were analysed not only for the whole writing session, but also in relation to the variable of time so as to visualise the dynamism of the writing process (Van Den Bergh & Rijlaarsdam, 1996). In order to operationalise the time variable while writing, each writing session was segmented into equal stages following previous studies (e.g. Barkaoui, 2016; Gánem-Gutiérrez & Gilmore, 2018; Roca de Larios et al., 2008; Tillema, 2012; Van Waes & Leijten, 2015). It should be noted, however, that existing studies differ in two respects when dividing the writing session. First, some researchers split the writing session into equal length stages in terms of the total number of cognitive activities that a participant reports (e.g. Breetvelt et al., 1994; Tillema, 2012), while others segment the writing session based on the total writing time a participant spends on the writing task (e.g. Barkaoui, 2016; Roca de Larios et al., 2008). In the current thesis, I chose the second option, i.e. dividing the writing session based on time, as I did not ask the participants to report their cognitive activities during writing for all four writing tasks. Secondly, existing studies are also different in terms of how many stages the writing session was divided into (ranging from 3 to 15). I adopted the practice of Gánem-Gutiérrez and Gilmore (2018) and Tillema (2012) in this thesis by dividing each writing session into five equal stages. There were two reasons behind this decision. To begin with, it is generally agreed that the initial, middle and final stages of writing exhibit their own characteristics because they are dominated by different writing activities (Ong, 2014). However, it might not be enough to segment the writing session into three stages in order to capture the changes across the writing time, as most writing sessions in the study lasted for over 25 minutes. Therefore, the number was increased from three to five in the hope of visualising the dynamic writing process in a more informative way. In addition, it is conceded that, in theory, more details of the writing process could be captured if more stages were included in the analysis. However, in reality, it was not practical for me to segment each writing session into more than five stages due to the limited time and human resources.

Taken together, each writing session in this study was divided into five equal stages based on the total time the participant spent completing each writing task. For example, if a participant spent 25 minutes writing an essay, the writing session was divided into five five-minute long stages. Next, each keystroke logging file and stimulated recall protocol was segmented into five parts, with each part corresponding to one of the five stages in the writing session.

3.6.2 Writing behaviours

Writing behaviours were analysed in terms of speed fluency, pausing and revision. I additionally analysed revision behaviours, although no extra time was given to revise the texts in the present study. The main reason for including revision behaviours for data analysis was that as shown in the data, participants in this study revised quite often (6.18 times per minute and 3.86 times per minute for L1 and L2 writers, respectively) during task performance. Also, the majority of the revisions (90% and 85% of the total made by L1 and L2 writers, respectively) were carried out at the end of the current text and to linguistic episodes that were just produced by the participants. This seems to indicate that the time limit for text production does not necessarily prevent participants frequently revising their texts along drafting, although it tends to reduce the chance for extensive revision in the systematic reviewing period.

Speed fluency in this study was operationalised from a process perspective, although traditional measures of fluency often relate to the length of the final text or its derivatives that control the total writing time (Van Waes & Leijten, 2015), or the number and length of a certain kind of unit, such as a T-unit (Wolfe-Quintero et al., 1998). The rationale for employing process-based rather than product-based fluency measures was that measures based on the final product cannot adequately assess fluency, because text length may be affected by a number of factors, such as how familiar the topic is to the writer, writers' attitudes towards the task, and the amount of time they choose to allot to planning, monitoring and editing (Abdel Latif, 2009; Mazgutova, 2015). Moreover, measures involving the number and length of a unit have been found wanting as they are tapping into an aspect of syntactic complexity rather than fluency (Bulté & Housen, 2012; Jiang, 2013; Norris & Ortega, 2009). In addition, process-based fluency measures provide information that cannot be obtained by analysing writing products. For example, a writer who pauses or revises less is considered more fluent. This cannot normally be revealed in the final text, but it can be captured by employing process-based fluency measures.

In this study, two measures were used to evaluate speed fluency: production rate, computed by dividing the number of characters (including revised characters) by active writing time (i.e. writing time in minutes excluding time on pauses), and mean length of P-bursts (i.e. the number of characters between pauses). Characters, rather than words, were used as the basic unit, given that counting characters involves less controversy than counting words in Chinese, and the numbers of characters and words in a written text are closely correlated (Jiang, 2013). The number of characters (including revised characters) in each text was obtained via Corpus Word Frequency App.[3]

Pause in this study was defined as any inactivity during writing for over two seconds, following the conventions of previous studies on adult writing (e.g. Révész et al., 2019; Xu & Qi, 2017). In other words, only pauses exceeding two seconds were considered meaningful and included in the analysis. According to Chenu et al. (2014) and Wengelin (2006), a two-second pause threshold allows excluding those pauses that are not associated with cognitive activities during writing but caused by motor execution, such as moving from one key to another, even for the slowest typist. Recently, a few studies have used shorter pause thresholds, such as 200, 250 or 500

---

[3] Corpus Word Frequency App is a character and word frequency counting tool for Chinese texts developed by the Institute of Applied Linguistics, Ministry of Education, People's Republic of China. For more information about the app, see: www.cncorpus.org/CpsTongji.aspx.

milliseconds (e.g. Aldridge et al., 2018; Van Waes & Leijten, 2015). Compared to longer pause thresholds, shorter thresholds may be able to provide more insights into lower-order processes that are mainly connected to linguistic issues with spelling and lexical items. Therefore, they may better reflect the nature of L2 writing in which relatively more cognitive resources are required for linguistic encoding. Although acknowledging the benefits of using shorter pause thresholds, I employed the threshold of two seconds due to the heavy workload involved in manually coding the pauses captured by *Translog 2.0* against the limited time and human resources allowed for the study.

Pausing behaviours were operationalised in terms of pause frequency and pause duration. The linear view tool in the *Translog 2.0 Supervisor* program was used to identify pauses for each task, and the pause threshold was set to two seconds. See Figure 3.4 for an excerpt from a *Translog* linear view output file. First, all pauses in the output file (e.g. those in square brackets in Figure 3.4) were manually extracted and ordered in accordance with the time sequence in which they occurred in an excel file. Next, each pause was coded in terms of its text location. I started with the categories commonly used in alphabetic writing studies (e.g. Barkaoui, 2019; Révész et al., 2019), i.e. whether a pause occurred within a word, between words, between clauses or between sentences. Each pause was then manually assigned to one of these categories (See Figure 3.5 for an extract from a coding file). Two additional categories emerged when coding the pauses. First, due to the dual-step transcription process of Chinese, from time to time, writers paused when they were selecting the Chinese characters from the list after finishing typing the Pinyin. Such pauses were coded as a pause between Pinyin and character(s). Second, pauses could also appear between two contextual revisions (i.e. revisions made to already-written text), usually when writers were rereading their text. It seems impossible to tell at which textual unit level these pauses occurred; thus, they were coded as an individual category, pauses between contextual revisions. Pause frequency by location was then calculated for each participant and standardised by the time (in minutes) of the whole writing session or stage. Pause duration was expressed as median length of pause.

我[D]•的••••[O][N][G]••◀◀••◀[T][O][N][G][•02.719]◀◀[•02.219]◀◀
[•03.516][T][O][N][G]••[N][N][A]潼南••◀◀••[T][O][N][G]••[N][N][A]•••◀◀◀••••••[N]•[I][A][N]••
童年••••••[Z][U][I]••最•••[H]••◀••[K]••[U][I][A]魁岸•••◀◀••••[K]•[U][A][I]•[L]••快乐••[D]•的
••[S][H][I]•••[G][U][A][N][G]••时光••[S][H][I]是••••••[Z]•[A][I]在••[X][I][A][O]•[X][U]••••◀•
◀•••••[X][U][E]•小学••，[•18.015][E][R]•而[•05.156][X][I][A][O]••[X][U]小学••[H][Z][O][N][G]
中•••••••◀◀◀••◀•••[Y]•[I][N][C]••◀••[W][E][I]因为••[X][I][A][O]•[U][X]•小学•••••[D]的
••[X][U][E]•[Y]•••[E]••••••学业•••[Q][I][N][G]••[S][O][N][G][•06.281]轻松•••••，
••••••••[U][Z][O]•••••◀••••◀••••••◀◀◀◀◀◀◀◀◀◀◀◀[•13.516][Y]•[I][N]•[W][I][E]因为••[X][U][E]•[Y][E]
学业•••••[B][I]•[J][I][O][A]比较••[J]••◀•[Q][I][N][G]•[S][O][N][G]轻松，
[•02.625][M][A][M][A]•妈妈•[J][I][U]就••••••[J][I][N][G]••[H][C][A][N][G]••经常•[D][A][I]•
带••[W][O]我••[Q][U]••去[•02.032][Q][U]•[A][N]••••[G][U][O]全国•••[G]•[E]•••[D][I]各地
••[O][U]••••••◀◀•••••[Y][O][U]••[W]•[N][A]••游玩[•06.282]，[•02.625][W][O]我

Note: The number in square brackets refers to pause duration in seconds. Each • stands for a pause for 200 ms. The symbol ◀ stands for the action of pressing the backspace/delete key during writing.

Figure 3.4 *An excerpt from a* Translog *linear view output file (two-second threshold)*

Table 3.4 *An extract from a coding file for pauses*

Stage	*Pause*	*Location*	*Length*
1	网上订票，**[•05.656]**[K][A][N][X][I][N][W][E][N]看新闻	Bw	5.656
1	看新闻，**[•16.078]**[D][U][A][N][X][I][N]短信	Bw	16.078
1	短信，**[•02.578]**[W][A][N][G][S][A][H][A][N]•[G]网撒航	Bw	2.578
1	网上**[•18.484]**[Z][H]••[N]•◀••••[A]•[O]•找	Bw	18.484
1	朋友••，**[•03.704]** ◀••。	Bc	3.704
1	。**[•14.015]**[W]••[O]	Bs	14.015
1	[W][O]**[•06.671]**我	Bpc	6.671
1	给人看**[•02.328]**[W][E][N][J][U][A][N]	Bw	2.328
1	[W][E][N][J][U][A][N]**[•06.969]** ◀◀◀◀	Bpc	6.969
1	[Q][U][A][N]◀◀◀◀◀◀◀**[•03.375]**[T][U]•[P][I][A][N]•图片	Bw	3.375

1	[S][H][O][U][J][I]**[•05.094]**手机	Bpc	5.094
1	数据。[Return]**[•09.984]**[K][A][O]	Bs	9.984

Note: Bw = between words, Bpc = Between Pinyin and character(s), Bc = Between clauses, Bs = Between sentences

Revision behaviours were analysed based on the multi-dimensional taxonomies put forward by Lindgren and Sullivan (2006) and Stevenson et al. (2006). Although a number of taxonomies have been developed to model revision behaviours (e.g. Faigley & Witte, 1981; Matsuhashi, 1987; Thorson, 2000), multi-dimensional taxonomies seem to be the most comprehensive frameworks to categorise revision behaviours in terms of multiple dimensions. In their frameworks, Lindgren and Sullivan (2006) and Stevenson et al. (2006) first distinguished revisions as internal and external. Internal revisions occur within the writer's mind prior to transcription and are thus not visible. They can entail changes made to non-linguistic mental representations (i.e. pre-linguistic revisions), such as revisions to goals, plans and ideas, as well as changes made to linguistic episodes formulated in the writer's mind (i.e. pre-textual revisions). In contrast, external revisions are visible changes made to the text. Depending on the context (i.e. time of the occurrence), external revisions can be sub-divided into pre-contextual revisions, which are made at the point of inscription, and contextual revisions, which are made to already-written text. Both pre-contextual and contextual revisions may have different orientations; the changes can relate to content, language or typography. The two types of revisions can also involve different levels of linguistic units (i.e. linguistic domain), such as revisions within a word, to a word, to a clause or to a paragraph. External revisions can be additionally classified in terms of action, that is, whether they include additions, deletions, substitutions, permutations, distributions and consolidations. This study only examined external revisions that could be identified in the keystroke logging files.

The scheme used to code external revisions was developed based on the multi-dimensional taxonomies proposed by Lindgren and Sullivan (2006) and Stevenson et al. (2006). After all revisions were manually identified from the keystroke logging files, each revision was first distinguished in terms of its linguistic domain, whether it occurred below the character level, at the character level, below the word level, at the word level, below the clause level or at the clause level or above. However, below-character and character-level revisions were later merged with below-word revisions to form one category, due to the fact that a character and a word can sometimes overlap in Chinese. Second, revisions were classified according to context, whether they involved pre-contextual or contextual revisions. Next, a new dimension, level of transcription, was added to the taxonomies, given the two steps involved in typing Chinese using the Pinyin input method. The new dimension distinguished a revision in terms of whether it entailed a change to Pinyin or Chinese characters. Finally, the frequency count by revision category for each participant was calculated and standardised by the time (in minutes) of the whole writing session or stage. The rationale for using time rather than the number of words in the final text (e.g. Barkaoui, 2016; Stevenson et al., 2006) for standardising is that although the same time (30 minutes) was allowed for writing, the actual time that each participant took to finish each task varied. In this case, a time-based standardisation procedure is likely to be more effective than a word-based one to avoid effects arising from difference in the time used for task completion.

In this study, revisions were not coded in terms of action, as the boundary between some actions (e.g. additions and substitutions, permutations and consolidations) could be blurred. The orientation of revisions was not included in the coding scheme for revisions identified from the keystroke logging files either, as the coding was very likely to be invalid without evidence from verbal reports (cf. Stevenson et al., 2006). However, the orientation of revisions was considered when analysing the stimulated recall protocols for cognitive activities associated with revision behaviours. Table 3.5 provides a summary of the codes of writing behaviours and how they were operationalised in this study.

Data from six selected participants, three selected randomly from each group (10% of the total data), were coded by a second coder. The second coder was an L1 user of Mandarin Chinese, holding an MA degree in Chinese Linguistics and Applied Linguistics. The agreement between coders reached an excellent level, with Cohen's kappa being .91 for pausing behaviours and .93 for revision behaviours.

Table 3.5 *Coding scheme for writing behaviours*

Writing behaviours	Operationalisation
Speed fluency	
Production rate	Number of characters produced (including revised characters) divided by the total writing time excluding pauses (in minute)
P-burst length	Number of characters between two pauses on average
Pausing behaviours	
Pause frequency	Number of pauses by location
Pause duration	Median length of pause by location
• Pause location	• Pause within a word (e.g. [W][E][I]*唯* **[•02.171]** [Y][I]一 (only))

	• Pause between words (e.g. [N][A]*那* (that) **[•03.968]**[Z][H][A][N][L][A][N]*湛蓝* (blue)) • Pause between clauses (e.g. [M][A][M][A] *妈妈* (mom) [J][I][U] *就* (will) [J][I][N][G][H][C][A][N][G] *经常* (often) [D][A][I] *带* (take) [W][O]*我* (me) [Q][U]*去* (go) [Y][O][U][W][N][A]*游玩* (visit) **[•06.282]**, **[•02.625]**[W][O]*我* (I)⋯) • Pause between sentences (e.g. [X][I][N][G][X][I][A][N][G] *形象* (vividly) [L][E] *了* (a particle in Chinese used at the sentence end)。**[•34.891]**[Z][A][I]*在* (in)⋯) **Additional categories for pause location:** • Pause between Pinyin and character(s) (e.g. [L][A][N]**[•02.922]***懒* (lazy)) • Pause between contextual revisions (e.g. (Contextual revision 1) [J][I][G][E] *几个* (several)→→→**[•08.922]** (Contextual revision 2) [T][E][N][G][T][O][N][G][Z][H][O][N][G] *疼痛中* (the pain) [J][I][E][J][I][U][C][H][U][L][A][I]*解救出来* (rescue from))
External revision behaviours	
Linguistic domain	
Below word	Changes made to Pinyin before converting into a character (e.g. [K][A][I][S]◀[X][I][N]*开心* (happy)) Changes made to character(s) within a word (e.g. [Y][I][N][W][E][I]*因为* (because)◀[C][I]*此* (therefore)) Substitution of a character with another with same spelling in Pinyin (e.g. [Q][I][E]*切* (cut)◀[Q][I][E]*且* (in addition))
Word	Changes made to a word in Pinyin or character(s)
Below clause	Changes made to part of a clause but more than a word in Pinyin or characters
Clause and above	Changes made to more than clauses in Pinyin or characters
Context	
Pre-contextual	Changes made at the point of inscription in Pinyin or character(s)
Context	Changes made to already-written text in character(s)
Level of transcription	
Pinyin	Addition, deletion or substitution of Pinyin before selecting character(s)
Characters	Addition, deletion or substitution of character(s)

Note: The number in square brackets refers to pause duration in seconds. The symbol ◀ stands for the action of pressing the backspace/delete key during writing.

3.6.3 Stimulated recall comments

The stimulated recall data comprised 1,545 minutes (M = 48.28 minutes, range 29 to 66 minutes per participant) for L2 writers of Chinese and 1,413 minutes (M = 44.16 minutes, range 20 to 62 minutes per participant) for L1 writers. Four steps were involved in the analysis of stimulated recall comments. First, stimulated recall protocols were transcribed in full. Then, the protocol transcripts were segmented based on the content of the mental actions reported by the participants.[4] In other words, a new segment reflected a participant's switch to a different mental action (e.g. Gánem-Gutiérrez & Gilmore, 2018; Van Den Bergh & Rijlaarsdam, 2001). Example 1 in Table 3.6 shows an episode where an L2 writer of Chinese recalled his thoughts during a pause between clauses. This episode contains two segments, as it involves two mental actions, one related to content and the other concerning language.

Table 3.6 *Segmenting and coding stimulated recall data*

Example 1: (Segment 1) So you know, I was thinking, well, what other thoughts that I had at that time, so firstly, I was excited because I could make new friends and what other things I was excited about. I decided what I was excited about. (Segment 2) I was also looking for the correct word as well because I was fairly confident about using *接触* (have access to), but I wasn't sure whether I should use *知识* (knowledge) to describe the content of the class or not.
Example 2:

[4] Comments on revisions to typos, which accounted for 34% of the total comments on revisions for both groups, were excluded from further analysis.

写这段的时候我其实挺纠结的，反复思考整个段落的结构和语言上面的东西，还有内容上的问题。(I was struggling when writing this paragraph, including its structure, the language use as well as what to say.)

Next, the segments were grouped into categories of planning, translation and monitoring based on Kellogg's (1996) cognitive model of writing. Comments on planning were additionally distinguished as to whether they were idea- or organisation-related, whereas those on translation were further classified as to whether the participants referred to translation in general, lexical items, syntactic structures or cohesive devices. See Table 3.7 for an overview of the coding scheme and examples of each category. Notably, some segments were assigned to two or more subcategories (Barkaoui, 2015), as some mental actions could be broken up by, or occurred together with, other actions (Polio & Friedman, 2017), making it impossible to further divide the protocols into smaller segments. Example 2 in Table 3.6 contains the thoughts an L1 writer of Chinese reported when making a revision below the clause level. In this segment, the participant mentioned that she was struggling with several aspects of the paragraph (i.e. ideas, organisation and language) at the same time. The comment was made as one sentence, with three aspects intertwined with one another. Thus, this comment was treated as one segment but assigned to three subcategories, planning ideas, planning organisation and translation in general. Lastly, comments that fell into a specific group were added up, and a percentage for each group was calculated for the whole writing session and each stage. Data from six participants (randomly selected) from each group, i.e. 20% of the stimulated recall data, was double-coded by the same coder who coded the pausing and revision behaviours. Inter-coder reliability was found to be excellent (Cohen's kappa = .97). Disagreements between the coders were resolved through discussion.

Table 3.7 *Coding categories for stimulated recall comments*

Cognitive activity/ Subcategory	Example
Planning	
Idea	I think I just had a general thought, but I wanted to make an argument about how easy it is to change plans with mobile phones. (pausing)
	After putting down *同意* (agree), I was still thinking whether this was how it happened. Then I realised it was not, so I deleted it. (revision)
Organisation	I was thinking which one I should say first. (pausing)
	The paragraph in the middle was too long. I wanted to restructure the text to make the first paragraph as the opening, followed by another two paragraphs on the beginning and development of the story, and this paragraph as the ending of the story. (revision)
Translation	
Translation in general	I was thinking how to phrase the next sentence. I knew what to write but I didn't know how to phrase it. (pausing)
	I deleted that because I wanted to rephrase it. I remembered that the first paragraph did not flow well, but once I came to the story, I became much more fluent. (revision)
Lexical retrieval	Because I forgot the word for cartoon, so I wanted to change it to *电动* (electric), but I couldn't remember the noun after *电动* (electric), so I had to think of *电脑游戏* (computer games). (pausing)
	I thought this word was the wrong one, but I didn’t know where to find the right one. So basically trial and error, because I usually use a dictionary when writing. (revision)
Syntactic encoding	This sentence was long, and I didn't know how to write it. So I was thinking about the word order and where to put *的* (a particle in Chinese, usually used between an adjective and a noun). (pausing)
	I thought this part was redundant in terms of the sentence structure. (revision)
Cohesion	I was wondering whether to use *但是* (but) or *可是* (but). (pausing)
	I thought *每当* (whenever) was not appropriate, so I had to replace it with another cohesive device. (revision)

Monitoring: Reading for monitoring	I was reading back to see whether this sentence made any sense. (pausing)
	I was still reading that sentence and thinking how to make it better. (pausing)
Others	I felt like I was not writing an essay but whinging to others about a bad experience of my childhood. I was surprised that I would have become so emotional when writing this! (pausing)
	I was like should I write it in Chinese numbers instead. (revision)
No recall	I was just, just, yes… I didn't know what I was thinking originally. (pausing)
	I don't remember why I deleted that. (revision)

3.6.4 Text quality

Text quality was assessed using holistic rating scales. The rating categories included task response, cohesion and coherence, lexical resources, and grammatical range and accuracy. The rating scale for argumentative essays was adapted from the IELTS Writing Task Two assessment criteria, as both argumentative writing prompts used in the study were taken and translated from IELTS Writing Task Two. One adjustment was made to the criterion for lexical resources, given the difference between the English and Chinese languages; that is, the requirements for spelling, such as occasional errors in spelling and having some control over spelling, were replaced with those for Chinese characters. The rating scale for narrative essays was also adapted from the IELTS Writing Task Two assessment criteria. As for argumentative writing, the same adaptation was made to lexical resources. In addition, some adjustments were made to the criterion for task response, with reference to the rating scales for story writing in standardised Chinese tests (e.g. Advanced Placement exam on Chinese language and culture). Specifically, the requirements regarding ideas presentation and development were changed into those concerning how complete and elaborate the story is. The assessment criteria for English essays were used to evaluate Chinese essays, since good contemporary Chinese writing and good English writing share many common traits (X. Liu & Furneaux, 2014), and the existing assessment criteria for writing in Chinese did not balance between language and content (S. Liu & Cao, 2015). See Appendix F for the rating scales for argumentative and narrative essays.

Two independent raters were recruited to rate the essays. Both raters were L1 users of Chinese and had experience of teaching writing courses to L2 Chinese users. Before scoring the essays, each rater met the researcher individually for a face-to-face training session via Skype. The training session involved three steps adapted from Knoch, Fairbairn, Myford and Huisman's (2018) procedures for training new writing raters. First, the raters were given a brief introduction to the current study, including the aim, participants and data collection procedures. In particular, information about the participants' proficiency in L2 Chinese was given by providing the raters with the proficiency test used in the study and the descriptions of Bands A, B and C in TOCFL from the Taiwan Ministry of Education. Next, the raters familiarised themselves with the writing tasks and rating scales for argumentative and narrative essays. They were allowed sufficient time to read the writing prompts and rating scales and ask questions about any items that were not clear to them. Then, the raters participated in a rating practice session, in which they scored five argumentative and five narrative essays with different text qualities produced by L1 and L2 writers of Chinese. Their ratings of the essays were then compared with the scores given by the researcher. Absolute agreement on scores was reached for the majority of the essays (N = 8 for Rater 1 and N = 9 for Rater 2). After rating, the raters were also asked to explain their reasons for their scores to ensure that they understood and were interpreting the rating scales correctly. The training session took around three hours. The raters were given as much time as they wanted to rate all the essays.

Interater reliability was good, with interclass correlation being .81 for the essays by L2 writers of Chinese and .80 for the essays by L1 writers. One text, with a score discrepancy between two raters, was discussed and rescored. Text quality was approximated by averaging the scores given by the raters.

3.7 Statistical analyses

Statistical analyses were carried out in the *R* environment (*RStudio* version 1.1.463) using mixed-effects modelling. A mixed-effects model contains both fixed effects (independent variables) and random effects (participants and items). It allows for simultaneously generalising of the results for new participants and new items and modelling the relationship between fixed effects and the dependent variable in addition to participant- and item-related variables (Sonbul & Schmitt, 2013). Mixed-effects modelling is found to be more powerful than means-based parametric statistical techniques, such as *t* tests and *ANOVA*, in handling longitudinal data and data in repeated-measure design studies (Cunnings, 2012; Linck & Cunnings, 2015). As the current study employed a repeated-measure design and aimed to examine participants' writing behaviours from a temporal perspective, it was appropriate to use mixed-effects modelling to address the research questions.

Before I constructed the mixed-effects models, the data for all indices of writing behaviours were first inspected for outliers. The outliers were trimmed to values of three standard deviations from the mean for each index. Next, a series of linear mixed-effects regression models were constructed using the *lmer* function in the *lme4* package. In most models, the random effects were participant and writing prompt. Random slopes were retained only when the models succeeded in converging. Bonferroni post hoc tests were performed with some models when pairwise comparisons were needed to confirm where differences occurred. The linearity, homoscedasticity and normality assumptions for the models were checked using residual plots. If the residual plots suggested the data did not meet the assumptions, a log or square root transformation, depending on which one resulted in a better-fit model, was performed on the numeric variables in the model to improve the goodness of fit. The threshold value for alpha was set at .05 for all mixed-effects models, but lowered to .01 for Bonferroni post hoc tests due to the multiple comparisons of the data that could increase Type I errors. Both marginal and conditional R^2 values were calculated as measures of variance explained by fixed and random effects using the *r.squared GLMM* function in the *MuMIn* package. The *cohensD* function in the *lsr* package was used to obtain effect sizes (Cohen's *d*) for Bonferroni post hoc tests. *D* values of .60, 1.00 and 1.40 were considered to be small, medium and large effect sizes (Plonsky & Oswald, 2014).

Data for stimulated recall comments were analysed qualitatively.

Chapter 4 Results

In Chapter 4, the results of the statistical analyses performed to address the research questions are presented. The chapter consists of four sections. In each section, an explanation of the statistical analysis procedure for the research question is given first, followed by a detailed overview of the results. For the reader's convenience, the results for Research Question 5, that examined L1 Chinese writing, are presented after the corresponding research question that investigated L2 Chinese writing. Each section ends with a brief summary of the findings.

4.1 Stages of writing and cognitive processes in Chinese writing

First, the extent to which stages of writing affected behaviours in Chinese writing was investigated. Two sets of linear mixed-effects regression analyses with stage as the fixed effect and a writing behaviour index as the dependent variable were conducted with data collected from L1 and L2 Chinese writers separately. The random effects were initially participant and prompt for all models. However, some models failed to converge with two random effects, so prompt, the random effect that contributed less variance, was removed. No random slope was added to ensure model convergence. A log transformation of the dependent variable was done when necessary.

Descriptive data for general information about task performance, i.e. time on task and text length, are displayed in Tables 4.1 and 4.2. Tables 4.3–4.5 provide descriptive data for speed fluency, pausing and revision measures. Complete tables are available in Tables 1–3 in Appendix G. As shown in Table 4.6, stage was found to have an influence on both fluency measures in L2 Chinese writing, explaining 3% and 2% of the variance in production rate and P-burst length, respectively. In L1 Chinese writing, stage only had an impact on the production rate measure. The variance in L1 production rate explained by stage was 5%, slightly larger than that for L2 writing.

Table 4.1 *Total time on task (in minutes)*

	L2 Chinese (N = 32)				L1 Chinese (N = 32)			
	M	*SD*	95% *IC* lower	95% *IC* upper	*M*	*SD*	95% *IC* lower	95% *IC* upper
Total	28.78	3.67	28.14	29.42	22.66	5.36	21.73	23.58
A1	28.85	3.48	27.64	30.06	22.81	5.48	20.91	23.71
A2	28.20	3.82	26.86	29.52	22.49	4.84	20.81	24.17
N1	28.89	3.44	27.70	30.08	23.15	4.98	20.75	24.19
N2	29.18	4.01	27.79	30.57	22.18	6.24	20.02	24.34

Note: A1 = Argumentative writing task 1, A2 = Argumentative writing task 2, N1 = Narrative writing task 1, N2 = Narrative writing task 2

Table 4.2 *Text length (number of characters in final texts)*

	L2 Chinese (*N* = 32)				L1 Chinese (*N* = 32)			
	M	*SD*	95% *IC* lower	95% *IC* upper	*M*	*SD*	95% *IC* lower	95% *IC* upper
Total	541.28	199.86	506.66	575.90	741.30	235.24	700.54	782.05
A1	547.97	201.16	478.27	617.67	746.09	198.60	677.28	814.90
A2	517.97	189.72	452.24	583.70	733.34	217.69	657.92	808.77
N1	555.38	218.37	479.71	631.04	734.88	263.20	643.68	826.07
N2	543.81	196.67	475.67	611.95	750.88	265.28	658.96	842.79

Note: A1 = Argumentative writing task 1, A2 = Argumentative writing task 2, N1 = Narrative writing task 1, N2 = Narrative writing task 2

Table 4.3 *Speed fluency in Chinese writing*

	L2 Chinese (*N* = 32)				L1 Chinese (*N* = 32)			
Measure	*M*	*SD*	95% *IC* lower	95% *IC* upper	*M*	*SD*	95% *IC* lower	95% *IC* upper
Speed fluency								
Production rate	43.17	8.74	41.70	44.70	63.28	11.12	61.35	65.21
P-burst length	4.36	1.39	4.12	4.60	5.71	1.64	5.42	6.00

Note: Data by stages are available in Table 1 in Appendix G, Production rate = Number of characters produced divided by active writing time (in minutes), P-burst length = Number of characters between two pauses

Table 4.4 *Pause frequency and duration by location in Chinese writing*

	L2 Chinese (*N* = 32)				L1 Chinese (*N* = 32)			
Measure	*M*	*SD*	95% *IC* lower	95% *IC* upper	*M*	*SD*	95% *IC* lower	95% *IC* upper
Frequency (number of pauses per minute)								
Total	4.17	.85	4.02	4.32	3.50	.70	3.38	3.62
Within a word	.17	.10	.15	.19	.06	.07	.05	.07
Between Pinyin and character	.49	.49	.41	.58	.15	.17	.13	.19
Between words	1.97	.55	1.87	2.07	1.57	.47	1.49	1.65
Between clauses	.47	.18	.44	.50	.71	.30	.66	.76
Between sentences	.73	.21	.69	.77	.74	.23	.70	.78
Between contextual revisions	.34	.31	.29	.39	.25	.21	.21	.29
Duration (median length of pause)								
Total	4.19	.89	4.04	4.34	3.81	.63	3.70	3.92
Within a word	3.20	1.31	2.97	3.43	3.19	1.64	2.83	3.55
Between Pinyin and character	2.96	.59	2.86	3.06	2.94	1.45	2.67	3.21
Between words	3.75	.75	3.62	3.88	3.36	.51	3.27	3.45
Between clauses	4.64	1.78	4.32	4.94	4.18	1.41	3.93	4.42
Between sentences	8.97	5.23	8.06	9.88	6.74	2.76	6.26	7.22
Between contextual revisions	9.17	9.08	7.56	10.80	6.07	5.53	5.05	7.09

Note: Data by stages are available in Table 2 in Appendix G

Table 4.5 *Revision frequency (number of revisions per minute) by dimension in Chinese writing*

	L2 Chinese (*N* = 32)				L1 Chinese (*N* = 32)			
Measure	*M*	*SD*	95% *IC* lower	95% *IC* upper	*M*	*SD*	95% *IC* lower	95% *IC* upper
Total revision	3.86	2.55	3.24	4.12	6.18	2.00	5.83	6.53
Linguistic domain								
Below word	1.85	1.51	1.59	2.11	2.51	.95	2.34	2.67
Word	1.06	.65	.95	1.17	1.62	.63	1.51	1.73
Below clause	.65	.40	.58	.72	1.56	.66	1.45	1.67
Clause and above	.15	.13	.13	.17	.27	.21	.23	.31
Context								
Pre-contextual	3.30	2.51	2.86	3.73	5.55	1.93	5.22	5.88
Contextual	.57	.42	.50	.64	.63	.48	.55	.71
Level of transcription								
Pinyin	2.28	1.62	2.00	2.56	2.70	1.08	2.51	2.89
Character	1.42	1.04	1.24	1.60	3.25	1.36	3.01	3.49

Note: Data by stages are available in Table 3 in Appendix G

Table 4.6 *Results from linear mixed-effects regressions examining the effects of stage on speed fluency measures in Chinese writing*

Dependent variable	Fixed effect: stage	*Est*	*SE*	*t*	*p*
L2 Chinese (N = 32)					
Production rate	2	-.07	.03	-2.41	.02*
	3	-.11	.03	-3.61	< .01*
	4	-.05	.03	-1.56	.12
	5	-.15	.03	-5.05	< .01*
Model = lmer(log(production_rate) ~ stage + (1|participant) + (1|prompt), data, REML = F)					
$R^2{}_m = .03$, $R^2{}_c = .37$					
P-burst length	2	-.05	.04	-1.23	.22
	3	-.01	.04	-.32	.75
	4	< .01	.04	.10	.92
	5	-.14	.04	-3.43	< .01*
Model = lmer(log(pburst_length) ~ stage + (1|participant) + (1|prompt), data, REML = F)					
$R^2{}_m = .02$, $R^2{}_c = .37$					
L1 Chinese (N = 32)					
Production rate	2	-.09	.02	-3.71	< .01*
	3	-.09	.02	-3.64	< .01*
	4	-.05	.02	-2.18	.03*
	5	-.19	.02	-8.04	< .01*
Model = lmer(log(production_rate) ~ stage + (1|participant) + (1|prompt), data, REML = F)					
$R^2{}_m = .05$, $R^2{}_c = .50$					
P-burst length	2	.03	.06	.58	.57
	3	.06	.06	.99	.33
	4	.03	.06	.55	.58
	5	-.06	.06	-1.08	.28
Model = lmer(log(pburst_length) ~ stage + (1|participant) + (1|prompt), data, REML = F)					

Note: Production rate = Number of characters produced divided by active writing time (in minutes), P-burst length = Number of characters between two pauses, $R^2{}_m$ = marginal *R*-squared, $R^2{}_c$ = conditional *R*-squared, * = statistically significant result

Post hoc Bonferroni tests (see Table 4.7 for significant results, and Tables 1 and 2 in Appendix H for non-significant results) revealed that, in L2 writing, Stages 1 and 4 featured greater production rates and longer P-bursts than Stage 5. In addition, a higher production rate was found in Stage 1 than Stage 3. The results indicated that L2 writers tended be more fluent at the beginning but less fluent at the end of writing. In terms of L1 writing, the writers produced more characters per minute in Stage 1 but fewer in Stage 5 than Stages 2, 3 and 4, suggesting a decreasing trend of speed fluency across the stages. The effect sizes of the differences between stages were in the small range.

Table 4.7 *Significant results from post hoc Bonferroni tests examining the effects of stage on speed fluency in Chinese writing (p < .01)*

	Stage	*SE*	*t*	*d*
L2 Chinese (N = 32)				
P-burst length	S3 < S1	.03	-3.60	.34
	S5 < S1	.03	-5.03	.35
	S5 < S4	.03	-3.49	.23
P-burst length	S5 < S1	.04	-3.42	.24
	S5 < S4	.04	-3.55	.23
L1 Chinese (N = 32)				
P-burst length	S2 < S1	.02	-3.69	.22
	S3 < S1	.02	-3.62	.34
	S5 < S1	.02	-8.01	.35
	S5 < S2	.02	-4.34	.19
	S5 < S3	.02	-4.40	.08
	S5 < S4	.02	-5.86	.23

Note: Production rate = Number of characters produced divided by active writing time (in minutes), P-burst length = Number of characters between two pauses

Turning to pause frequency (Table 4.8), stage effects were identified on pause at all locations apart from within-word pause for both writer groups. Most of the variances explained by stage were less than 10%, except for pause frequency between contextual revisions in L2 writing (17%). The results of post hoc Bonferroni tests (see Table 4.9 for significant results, and Tables 1 and 2 in Appendix H for non-significant results) showed that L2 writers paused more frequently overall and between clauses in the middle stages than in Stages 1 and 5, whereas L1 writers only paused less frequently overall and at the clause boundary at the beginning. L2 writers paused less often between Pinyin and character(s) and at the word boundary in Stage 5 than in the earlier stages, while L1 writers made fewer pauses at these two locations in Stages 1 and 5. Both groups tended to pause more often at the sentence boundary in Stage 1 and between contextual revisions (See Table 3.4 in Chapter 3 for a definition and example of pauses between contextual revisions) in Stage 5. The effect sizes of the differences were small.

Table 4.8 *Results from linear mixed-effects regressions examining the effects of stage on pause frequency in Chinese writing*

	L2 Chinese (N = 32)					L1 Chinese (N = 32)				
Dependent variable	Fixed effect: stage	*Est*	*SE*	*t*	*p*	Fixed effect: stage	*Est*	*SE*	*t*	*p*
Total pause	2	.68	.11	6.01	< .01*	2	.53	.11	4.96	< .01*
	3	.59	.11	5.19	< .01*	3	.51	.11	4.67	< .01*
	4	.45	.11	3.97	< .01*	4	.38	.11	3.53	< .01*
	5	-.10	.11	-.91	.36	5	.30	.11	2.72	.01*
	Model = lmer(total_pause_frequency ~ stage + (1|participant) + (1|prompt), data, REML = F), R^2_m = .07, R^2_c = .43					Model = lmer(total_pause_frequency ~ stage + (1|participant) + (1|prompt), data, REML = F), R^2_m = .03, R^2_c = .32				
Ww pause	2	.03	.02	1.76	.08	2	< .01	.01	.10	.92
	3	.03	.02	1.41	.16	3	.01	.01	.53	.59
	4	-.02	.02	-1.04	.30	4	-.02	.01	-1.15	.25
	5	-.03	.02	-1.40	.16	5	-.01	.01	-.61	.54
	Model = lmer(log(ww_pause_frequency) ~ stage + (1|participant) + (1|prompt), data, REML = F)					Model = lmer(ww_pause_frequency ~ stage + (1|participant), data, REML = F)				
Bpc pause	2	.06	.03	2.44	.01*	2	.08	.02	3.66	< .01*
	3	.03	.03	1.23	.22	3	.03	.02	1.46	.14
	4	< .01	.03	.09	.93	4	.05	.02	2.39	.02*
	5	.08	.03	-3.33	< .01*	5	-.01	.02	-.41	.68
	Model = lmer(log(bpc_pause_frequency) ~ stage + (1|participant), data, REML = F), R^2_m = .03, R^2_c = .54					Model = lmer(bpc_pause_frequency ~ stage + (1|participant) + (1|prompt), data, REML = F), R^2_m = .02, R^2_c = .33				
Bw pause	2	.36	.09	3.97	< .01*	2	.38	.08	4.59	< .01*
	3	.19	.09	2.09	.04*	3	.38	.08	4.58	< .01*
	4	.15	.09	1.67	.10	4	.26	.08	3.19	< .01*

	5	-.36	.09	-3.96	< .01*	5	-.01	.08	-.07	.94
	Model = lmer(bw_pause_frequency ~ stage + (1|participant) + (1|prompt), data, REML = F), R^2_m = .08, R^2_c = .31					Model = lmer(bw_pause_frequency ~ stage + (1|participant) + (1|prompt), data, REML = F), R^2_m = .05, R^2_c = .25				
Bc pause	2	.13	.04	3.55	< .01*	2	.16	.05	3.07	< .01*
	3	.17	.04	4.65	< .01*	3	.21	.05	4.12	< .01*
	4	.13	.04	3.51	< .01*	4	.22	.05	4.34	< .01*
	5	-.02	.04	-.59	.55	5	.09	.05	1.72	.09
	Model = lmer(bc_pause_frequency ~ stage + (1|participant) + (1|prompt), data, REML = F), R^2_m = .05, R^2_c = .21					Model = lmer(bc_pause_frequency ~ stage + (1|participant) + (1|prompt), data, REML = F), R^2_m = .03, R^2_c = .25				
Bs pause	2	-.04	.04	-.97	.33	2	-.10	.05	-2.05	.04*
	3	-.04	.04	-.99	.32	3	-.16	.05	-3.39	< .01*
	4	-.12	.04	-3.22	< .01*	4	-.11	.05	-2.42	.02*
	5	-.15	.04	-3.90	< .01*	5	-.07	.05	-1.52	.13
	Model = lmer(bs_pause_frequency ~ stage + (1|participant) + (1|prompt), data, REML = F), R^2_m = .03, R^2_c = .25					Model = lmer(bs_pause_frequency ~ stage + (1|participant) + (1|prompt), data, REML = F), R^2_m = .02, R^2_c = .16				
Bcr pause	2	.06	.03	1.98	.05*	2	.02	.02	.91	.36
	3	.11	.03	3.88	< .01*	3	.03	.02	1.10	.27
	4	.18	.03	6.14	< .01*	4	.01	.02	.20	.84
	5	.36	.03	12.45	< .01*	5	.19	.02	6.94	< .01*
	Model = lmer(log(bcr_pause_frequency) ~ stage + (1|participant) + (1|prompt), data, REML = F), R^2_m = .17, R^2_c = .41					Model = lmer(log(bcr_pause_frequency) ~ stage + (1|participant) + (1|prompt), data, REML = F), R^2_m = .08, R^2_c = .23				

Note: Ww = Within a word, Bpc = Between Pinyin and character(s), Bw = Between words, Bc = Between clauses, Bs = Between sentences, Bcr = Between contextual revisions, R^2_m = marginal *R*-squared, R^2_c = conditional *R*-squared, * = statistically significant result

Table 4.9 *Significant results from post hoc Bonferroni tests examining the effects of stage on pause frequency in Chinese writing (p < .01)*

	L2 Chinese (N = 32)				L1 Chinese (N = 32)			
	Stage	*SE*	*t*	*d*	Stage	*SE*	*t*	*d*
Total pause	S1 < S2	.11	-5.99	.52	S1 < S2	.11	-4.95	.43
	S1 < S3	.11	-5.18	.43	S1 < S3	.11	-4.65	.37
	S1 < S4	.11	-3.96	.32	S1 < S4	.11	-3.52	.27
	S5 < S2	.11	-6.90	.64				
	S5 < S3	.11	-6.09	.56				
	S5 < S4	.11	-4.87	.45				
Bpc pause	S5 < S1	.03	-3.32	.27	S1 < S2	.02	-3.65	.34
	S5 < S2	.03	-5.75	.44	S5 < S2	.02	-4.06	.35
	S5 < S3	.03	-4.54	.41				
	S5 < S4	.03	-3.41	.31				
Bw pause	S1 < S2	.09	-3.96	.38	S1 < S2	.08	-4.58	.45
	S5 < S1	.09	-3.95	.31	S1 < S3	.08	-4.56	.43
	S5 < S2	.09	-7.90	.77	S5 < S2	.08	-4.65	.41
	S5 < S3	.09	-6.03	.60	S5 < S3	.08	-4.64	.43
	S5 < S4	.09	-5.61	.45				
Bc pause	S1 < S2	.04	-3.53	.32	S1 < S3	.05	-4.11	.38
	S1 < S3	.04	-4.63	.42	S1 < S4	.05	-4.32	.37
	S1 < S4	.04	-3.50	.30				
	S5 < S2	.04	-4.12	.38				
	S5 < S3	.04	-5.22	.45				
	S5 < S4	.04	-4.09	.38				
Bs pause	S5 < S1	.04	-3.89	.29	S3 < S1	.05	-3.38	.27

Bcr pause	S1 < S3	.03	-3.86	.39	S1 < S5	.03	-6.92	.47
	S1 < S4	.03	-6.12	.54	S2 < S5	.03	-6.01	.40
	S1 < S5	.03	-12.41	.87	S3 < S5	.03	-5.82	.41
	S2 < S4	.03	-4.14	.35	S4 < S5	.03	-6.72	.51
	S2 < S5	.03	-1.43	.75				
	S3 < S5	.03	-8.54	.61				
	S4 < S5	.03	-6.28	.43				

Note: Bpc = Between Pinyin and character(s), Bw = Between words, Bc = Between clauses, Bs = Between sentences, Bcr = Between contextual revisions

In terms of pause duration (Table 4.10), stage effects were found for pauses in total, between clauses, between sentences and between contextual revisions in L2 writing, while for pause in total, between sentences and between contextual revisions in L1 writing. Stage explained 18% of the variance in pause duration between contextual revisions in L2 writing, but less than 10% in the rest of the measures. In L1 writing, stage accounted for over 10% of the variance in pause duration between sentences and between contextual revisions, while only 3% of total pause duration.

Post hoc Bonferroni tests (see Table 4.11 for significant results, and Tables 1 and 2 in Appendix H for non-significant results) revealed that both groups of writers paused longer overall in Stage 5. When pauses were analysed in terms of location, I found longer between-sentence pauses in Stage 1 than in Stages 2, 3 and 4 in L2 writing, whereas L1 writers paused longer between sentences in both Stages 1 and 5 than in the middle stages. Although the effect of stage was significant for between-clause pause duration in L2 writing, no significant difference was found between any two stages in a post hoc test of this measure. There was a tendency for Stage 5 to exhibit longer pauses between contextual revisions than other stages. However, this finding should be interpreted with caution, given that 43% of the L2 data (278 out of 640) and 51% of the L1 data (327 out of 640) contained missing values for between-contextual-revision pause duration. The effect sizes of the differences were found to be small.

Table 4.10 *Results from linear mixed-effects regressions examining the effects of stage on pause duration in Chinese writing*

	L2 Chinese (N = 32)					L1 Chinese (N = 32)				
Dependent variable	Fixed effect: stage	*Est*	*SE*	*t*	*p*	Fixed effect: stage	*Est*	*SE*	*t*	*p*
Total pause	2	.04	.01	2.51	.01*	2	.03	.02	2.02	.04*
	3	.03	.01	1.91	.06	3	.05	.02	2.90	< .01*
	4	.03	.01	1.72	.09	4	.04	.02	2.50	.01*
	5	-.04	.01	-2.68	.01*	5	-.01	.02	-.72	.47
	Model = lmer(1/log(total_pause_duration) ~ stage + (1|participant) + (1|prompt), data, REML = F), R^2_m = .04, R^2_c = .36					Model = lmer(1/log(total_pause_duration) ~ stage + (1|participant), data, REML=F), R^2_m = .03, R^2_c = .20				
Ww pause	2	< .01	.04	.11	.91	2	.02	.07	.34	.74
	3	.05	.04	1.16	.25	3	-.09	.07	-1.32	.19
	4	.07	.05	1.51	.13	4	.01	.07	.11	.91
	5	.02	.05	.51	.61	5	.01	.07	.19	.85
	Model = lmer(1/log(ww_pause_duration) ~ stage + (1|participant), data, REML = F)					Model = lmer(1/log(ww_pause_duration) ~ stage + (1|participant), data, REML = F)				
Bpc pause	2	-.01	.03	-.34	.74	2	-.07	.05	-1.41	.15
	3	-.01	.03	-.19	.85	3	-.08	.05	-1.66	.10
	4	< -.01	.03	-.07	.95	4	-.05	.05	-1.03	.31
	5	.04	.03	1.27	.21	5	-.01	.05	-.20	.85
	Model = lmer(1/log(bpc_pause_duration) ~ stage + (1|participant), data, REML = F)					Model = lmer(1/log(bpc_pause_duration) ~ stage + (1|participant), data, REML = F)				
Bw pause	2	< .01	.02	.03	.98	2	-.01	.03	-.39	.70
	3	.01	.02	-.37	.71	3	-.03	.03	-.90	.37
	4	.01	.02	-.35	.73	4	-.02	.03	-.54	.59

	5	.01	.02	.69	.49	5	-.06	.03	-1.86	.06
	Model = lmer(1/log(bw_pause_duration) ~ stage + (1|participant) + (1|prompt), data, REML = F)					Model = lmer(1/log(bw_pause_duration) ~ stage + (1|participant), data, REML = F)				
Bc pause	2	-.06	.07	-.88	.38	2	-.01	.06	-.18	.85
	3	-.13	.07	-1.88	.06	3	-.04	.06	-.79	.43
	4	-.19	.07	-2.63	.01*	4	-.07	.06	-1.16	.25
	5	-.15	.07	-1.99	.05*	5	-.02	.06	-.33	.74
	Model = lmer(log(bc_pause_duration) ~ stage + (1|participant) + (1|prompt), data, REML = F), $R^2_m = .01$, $R^2_c = .10$					Model = lmer(log(bc_pause_duration) ~ stage + (1|participant) + (1|prompt), data, REML = F)				
Bs pause	2	-.47	.08	-5.65	< .01*	2	-.77	.08	-9.36	< .01*
	3	-.49	.08	-5.84	< .01*	3	-.93	.08	-11.30	< .01*
	4	-.40	.08	-4.79	< .01*	4	-.95	.08	-11.50	< .01*
	5	-.23	.08	-2.76	.01*	5	-.55	.08	-6.77	< .01*
	Model = lmer(log(bs_pause_duration) ~ stage + (1|participant), data, REML = F), $R^2_m = .06$, $R^2_c = .23$					Model = lmer(log(bs_pause_duration) ~ stage + (1|participant), data, REML=F), $R^2_m = .20$, $R^2_c = .32$				
Bcr pause	2	.21	.13	1.58	.11	2	-.06	.11	-.54	.59
	3	.11	.13	.83	.41	3	.11	.11	.93	.35
	4	.37	.13	2.91	< .01*	4	.16	.11	1.34	.18
	5	.88	.12	7.28	< .01*	5	.59	.11	5.55	< .01*
	Model = lmer(log(bcr_pause_duration) ~ stage + (1|participant) +(1|prompt), data, REML = F), $R^2_m = .18$, $R^2_c = .24$					Model = lmer(log(bcr_pause_duration) ~ stage + (1|participant) +(1|prompt), data, REML=F), $R^2_m = .14$, $R^2_c = .23$				

Note: Ww = Within a word, Bpc = Between Pinyin and character(s), Bw = Between words, Bc = Between clauses, Bs = Between sentences, Bcr = Between contextual revisions, R^2_m = marginal *R*-squared, R^2_c = conditional *R*-squared, * = statistically significant result

Table 4.11 *Significant results from post hoc Bonferroni tests examining the effects of stage on pause duration in Chinese writing (p < .01)*

	L2 Chinese (N = 32)				L1 Chinese (N = 32)			
	Stage	*SE*	*t*	*d*	Stage	*SE*	*t*	*d*
Total pause duration	S2 < S5	.01	5.17	.39	S3 < S5	.02	3.62	.35
	S3 < S5	.01	4.57	.39				
	S4 < S5	.01	4.38	.38				
Bs pause duration	S2 < S1	.08	-5.63	.37	S2 < S1	.08	-9.33	.49
	S3 < S1	.08	-5.82	.37	S3 < S1	.08	-11.26	.54
	S4 < S1	.08	-4.77	.31	S4 < S1	.08	-11.46	.54
					S5 < S1	.08	-6.75	.45
					S3 < S5	.08	-4.61	.42
					S4 < S5	.08	-4.80	.45
Bcr pause duration	S1 < S5	.12	-7.21	.70	S1 < S5	.11	-5.46	.48
	S2 < S5	.10	-6.12	.62	S2 < S5	.10	-6.46	.62
	S3 < S5	.10	-7.48	.77	S3 < S5	.11	-4.73	.46
	S4 < S5	.10	-5.08	.40	S4 < S5	.10	-4.17	.33

Note: Bs = Between sentences, Bcr = Between contextual revisions

An impact of stage was found on all revision measures, with stage contributing to less than 10% of the variance in all measures (Table 4.12). Table 4.13 shows the significant results obtained from post hoc Bonferroni tests (refer to Tables 1 and 2 in Appendix H for non-significant results). In general, L2 writers revised more often in Stages 2, 3 and 4, and less frequently in the initial and final stages. In terms of the linguistic domain, more below-word, word-level and below-clause revisions occurred in the middle stages. No significant difference was identified in the number of revisions at the clause level and above between stages in the post hoc tests. In terms of context, more pre-contextual revisions were made in Stages 2, 3 and 4, while contextual revisions increased from the beginning to the end. In terms of the level of transcription to which revisions were made, L2 writers revised Pinyin less often in the beginning and final stages, while they revised characters less frequently at the beginning only. Similar distribution patterns of revision behaviours were found across stages in L1 writing. Apart from contextual revision, other types of revision tended to increase after the initial stage, stay constant during the middle stages and decline when writing approached the end. Contextual revision also seemed to occur more frequently in Stage 5, but a difference was only observed between Stage 1 and Stage 5. The effect sizes of the differences were all small.

Table 4.12 *Results from linear mixed-effects regressions examining the effects of stage on revision in Chinese writing*

	L2 Chinese (*N* = 32)					L1 Chinese (*N* = 32)				
Dependent variable	Fixed effect: stage	*Est*	*SE*	*t*	*p*	Fixed effect: stage	*Est*	*SE*	*t*	*p*
Total	2	.20	.03	6.54	< .01*	2	1.08	.02	5.23	< .01*
	3	.22	.03	7.15	< .01*	3	1.25	.02	6.04	< .01*
	4	.22	.03	7.05	< .01*	4	1.26	.02	6.12	< .01*
	5	.11	.03	3.51	< .01*	5	.55	.02	2.66	.01*
	Model = lmer(log(total_revision) ~ stage + (1|participant), data, REML=F), $R^2_m = .04$, $R^2_c = .68$					Model = lmer(total_revision ~ stage + (1|participant), data, REML=F), $R^2_m = .04$, $R^2_c = .54$				
Linguistic domain										
Below word	2	.12	.03	3.49	< .01*	2	.14	.04	3.77	< .01*
	3	.14	.03	4.01	< .01*	3	.18	.04	4.80	< .01*
	4	.15	.03	4.47	< .01*	4	.18	.04	4.96	< .01*
	5	-.01	.03	-.19	.85	5	.07	.04	1.80	.07
	Model = lmer(log(below_word_revision) ~ stage + (1|participant), data, REML=F), $R^2_m = .02$, $R^2_c = .62$					Model = lmer(log(below_word_revision) ~ stage + (1|participant) + (1|prompt), data, REML=F), $R^2_m = .03$, $R^2_c = .42$				
Word	2	.10	.03	3.38	< .01*	2	.11	.03	3.30	< .01*
	3	.13	.03	4.22	< .01*	3	.14	.03	4.24	< .01*
	4	.11	.03	3.58	< .01*	4	.14	.03	4.15	< .01*
	5	.07	.03	2.50	.01*	5	.08	.03	2.50	.01*
	Model = lmer(log(word_revision) ~ stage + (1|participant), data, REML=F), $R^2_m = .02$, $R^2_c = .47$					Model = lmer(log(word_revision) ~ stage + (1|participant) + (1|prompt), data, REML=F), $R^2_m = .03$, $R^2_c = .33$				
Below clause	2	.11	.03	4.12	< .01*	2	.14	.03	4.28	< .01*

	3	.11	.03	4.37	< .01*	3	.13	.03	3.96	< .01*
	4	.10	.03	3.83	< .01*	4	.13	.03	3.94	< .01*
	5	.08	.03	3.12	< .01*	5	.05	.03	1.59	.11
	Model = lmer(log(below_clause_revision) ~ stage + (1|participant) + (1|prompt), data, REML=F), $R^2_m = .02$, $R^2_c = .43$					Model = lmer(log(below_clause_revision) ~ stage + (1|participant) + (1|prompt), data, REML=F), $R^2_m = .03$, $R^2_c = .41$				
Clause and above	2	.04	.02	2.26	.02*	2	.03	.02	1.08	.28
	3	.05	.02	2.69	.01*	3	.04	.02	1.75	.08
	4	.05	.02	2.87	< .01*	4	.03	.02	1.23	.22
	5	.05	.02	3.21	< .01*	5	.05	.02	2.18	.03*
	Model = lmer(log(clause_and_above_revision) ~ stage + (1|participant) + (1|prompt), data, REML=F), $R^2_m = .02$, $R^2_c = .22$					Model = lmer(log(clause_and_above_revision) ~ stage + (1|participant) + (1|prompt), data, REML=F), $R^2_m = .01$, $R^2_c = .21$				
Context										
Pre-contextual	2	.18	.04	4.77	< .01*	2	.09	.20	4.34	< .01*
	3	.19	.04	5.15	< .01*	3	1.08	.20	5.17	< .01*
	4	.14	.04	3.97	< .01*	4	1.17	.20	5.62	< .01*
	5	.05	.04	-1.33	.18	5	.24	.20	1.17	.24
	Model = lmer(log(precontextual_revision) ~ stage + (1|participant), data, REML=F), $R^2_m = .04$, $R^2_c = .62$					Model = lmer(precontextual_revision ~ stage + (1|participant) + (1|prompt), data, REML=F), $R^2_m = .04$, $R^2_c = .52$				
Contextual	2	.11	.03	3.37	< .01*	2	.10	.03	2.78	.01*
	3	.11	.03	3.56	< .01*	3	.10	.03	2.78	.01*
	4	.19	.03	5.84	< .01*	4	.07	.03	2.03	.04*
	5	.30	.03	9.33	< .01*	5	.17	.03	5.03	< .01*
	Model = lmer(log(contextual_revision) ~ stage + (1|participant), data, REML=F), $R^2_m = .09$, $R^2_c = .42$					Model = lmer(log(contextual_revision) ~ stage + (1|participant), data, REML=F), $R^2_m = .02$, $R^2_c = .41$				
Level of transcription										
Pinyin	2	.14	.04	4.01	< .01*	2	.14	.04	3.93	< .01*

	3	.17	.04	4.70	< .01*	3	.15	.04	4.26	< .01*
	4	.13	.04	3.61	< .01*	4	.17	.04	4.72	< .01*
	5	-.06	.04	-1.61	.11	5	.04	.04	1.09	.28
	Model = lmer(log(pinyin_revision) ~ stage + (1|participant), data, REML=F), R^2_m = .04, R^2_c = .60					Model = lmer(log(pinyin_revision) ~ stage + (1|participant) + (1|prompt), data, REML=F), R^2_m = .03, R^2_c = .48				
Character	2	.15	.03	5.07	< .01*	2	.58	.14	4.27	< .01*
	3	.15	.03	5.22	< .01*	3	.70	.14	5.20	< .01*
	4	.21	.03	7.19	< .01*	4	.66	.14	4.87	< .01*
	5	.21	.03	7.37	< .01*	5	.37	.14	2.73	.01*
	Model = lmer(log(character_revision) ~ stage + (1|participant), data, REML=F), R^2_m = .04, R^2_c = .66					Model = lmer(character_revision ~ stage + (1|participant) + (1|prompt), data, REML=F), R^2_m = .03, R^2_c = .55				

Note: R^2_m = marginal *R*-squared, R^2_c = conditional *R*-squared, * = statistically significant result

Table 4.13 *Significant results from post hoc Bonferroni tests examining the effects of stage on revision in Chinese writing (p < .01)*

	L2 Chinese (N = 32)				L1 Chinese (N = 32)			
	Stage	*SE*	*t*	*d*	Stage	*SE*	*t*	*d*
Total	S1 < S2	.03	-6.52	.56	S1 < S2	.21	-5.22	.51
	S1 < S3	.03	-7.13	.68	S1 < S3	.21	-6.02	.54
	S1 < S4	.03	-7.02	.66	S1 < S4	.21	-6.10	.54
	S1 < S5	.03	-3.50	.26	S5 < S3	.21	-3.37	.37
	S5 < S3	.03	-3.63	.32	S5 < S4	.21	-3.45	.38
	S5 < S4	.03	-3.52	.35				
Below word	S1 < S2	.03	-3.48	.32	S1 < S2	.04	-3.76	.35
	S1 < S3	.03	-3.99	.39	S1 < S3	.04	-4.78	.42
	S1 < S4	.03	-4.50	.43	S1 < S4	.04	-4.94	.41
	S5 < S2	.03	-3.67	.30				
	S5 < S3	.03	-4.18	.34				
	S5 < S4	.03	-4.65	.41				
Word	S1 < S2	.03	-3.37	.33	S1 < S3	.03	-4.23	.29
	S1 < S3	.03	-4.21	.43	S1 < S4	.03	-4.13	.36
	S1 < S4	.03	-3.57	.32				
Below clause	S1 < S2	.03	-4.11	.40	S1 < S2	.03	-4.26	.39
	S1 < S3	.03	-4.36	.39	S1 < S3	.03	-3.95	.33
	S1 < S4	.03	-3.82	.35	S1 < S4	.03	-3.93	.36
Pre-contextual	S1 < S2	.04	-4.76	.44	S1 < S2	.21	-4.32	.43
	S1 < S3	.04	-5.14	.55	S1 < S3	.21	-5.15	.48
	S1 < S4	.04	-3.95	.37	S1 < S4	.21	-5.60	.50
	S5 < S2	.04	-6.08	.47	S5 < S3	.21	-3.99	.41

	S5 < S3	.04	-6.46	.49	S5 < S4	.21	-4.44	.46
	S5 < S4	.04	-5.28	.43				
Contextual	S1 < S2	.03	-3.36	.37	S1 < S5	.03	-5.01	.39
	S1 < S3	.03	-3.55	.34				
	S1 < S4	.03	-5.82	.59				
	S1 < S5	.03	-9.30	.79				
	S2 < S5	.03	-5.94	.52				
	S3 < S5	.03	-5.75	.47				
	S4 < S5	.03	-3.48	.29				
Pinyin	S1 < S2	.04	-4.00	.37	S1 < S2	.04	-3.91	.37
	S1 < S3	.04	-4.69	.45	S1 < S3	.04	-4.25	.37
	S1 < S4	.04	-3.60	.34	S1 < S4	.04	-4.70	.40
	S5 < S2	.04	-5.60	.45	S5 < S4	.04	-3.61	.36
	S5 < S3	.04	-6.29	.49				
	S5 < S4	.04	-5.21	.46				
Character	S1 < S2	.03	-5.06	.48	S1 < S2	.14	-4.25	.42
	S1 < S3	.03	-5.21	.49	S1 < S3	.14	-5.18	.47
	S1 < S4	.03	-7.17	.67	S1 < S4	.14	-4.85	.45
	S1 < S5	.03	-7.34	.67				

Next, I explored the extent to which stages of writing affected cognitive activities associated with writing behaviours elicited via the stimulated recall interviews. Table 4.14 summarises the cognitive activities associated with pausing behaviours in L2 Chinese writing. For the whole writing session, the largest percentage of stimulated recall comments referred to translation processes (52%), followed by those for planning (27%) and monitoring (12%). Most planning comments were content-related (93%), while the majority of translation comments focused on lexical retrieval mechanisms (57%). Most pauses at smaller textual unit boundaries (i.e. within a word, between Pinyin and character(s) and between words) reflected translation processes (95%, 97% and 68%, respectively). Between-clause pauses were linked to either planning (36%) or translation (44%), with a slightly higher percentage for translation processes. Around half (49%) of between-sentence pauses were associated with planning, while the majority of pauses between contextual revisions (54%) were related to monitoring.

Similar to the findings for the whole writing session, L2 writers devoted most of the time (around half) in pausing to translation processes, followed by planning and reviewing, in each writing stage. The percentage of translation-related comments was slightly higher in Stages 2, 3 and 4 than in Stages 1 and 5. There was a sharp drop in the percentage of planning-related comments after Stage 1, but an increase in the percentage of comments referring to monitoring activities from the beginning to the end, with noticeable growth between Stage 4 and Stage 5. Taking pause location into account, I found pauses at the clause boundary and below were more likely to be linked to translation processes (lexical retrieval in particular), while pauses at the sentence boundary tended to be associated more with planning activities (primarily with content planning). Pauses between contextual revisions were mainly related to content planning in the initial stage, but to monitoring towards the latter stages of writing.

Table 4.15 provides information about pause-associated cognitive activities in L1 Chinese writing. Unlike L2 writers, L1 writers referred to planning processes most often (45%), followed by translation (39%) and monitoring processes (9%) for the whole writing session. The comments on planning were mainly content-related (95%), while the largest percentage of translation comments was associated with lexical retrieval (48%). When considering pause location, in parallel with L2 writers, L1 writers mentioned translation processes for most pauses at smaller textual unit boundaries (67% for pause within a word, 82% for pause between Pinyin and character(s), and 56% for pause between words). Between-sentence pauses were primarily linked to planning activities (62%). However, different from L2 writing, the majority of pauses between clauses and between contextual revisions in L1 writing (56% and 49% respectively) referred to planning processes rather than to translation or monitoring processes.

Different from L2 writers, L1 writers spent most time on planning in each stage. They were constantly engaged in planning for around half of the time from Stage 1 to Stage 3, while the percentage dropped below 45% in the last two stages. The percentage of comments on translation remained stable (around 40%) across stages.

L1 writers made most references to translation for pauses at the word boundary or below, but on planning for pauses at the clause boundary and above. Similar to L2 writing, monitoring activities were few in the first four writing stages, but they increased considerably in Stage 5. Pauses between contextual revisions were more often linked to monitoring in Stage 5 than in earlier stages.

Table 4.14 *Pause-associated cognitive activities by location and stage in L2 Chinese writing (N = 32)*

		Planning			Translation					Monitoring	Others	No recall	Total
Stage	Location	Content	Organisation	Total (%)	Translation in general	Lexical retrieval	Syntactic encoding	Cohesion	Total (%)	Total (%)	Total (%)	Total (%)	Total (%)
1	Ww	1	0	1 (0%)	0	2	0	0	2 (0%)	0 (0%)	0 (0%)	0 (0%)	3 (0%)
1	Bpc	0	0	0 (0%)	0	7	0	0	7 (1%)	0 (0%)	0 (0%)	1 (0%)	8 (1%)
1	Bw	66	5	71 (11%)	24	132	62	12	230 (37%)	17 (3%)	2 (0%)	37 (6%)	357 (57%)
1	Bc	24	2	26 (4%)	4	8	16	5	33 (5%)	3 (0%)	0 (0%)	12 (2%)	74 (12%)
1	Bs	63	5	68 (11%)	3	8	13	4	28 (4%)	9 (1%)	1 (0%)	7 (1%)	113 (18%)
1	Bcr	34	12	46 (7%)	5	1	1	2	9 (1%)	2 (0%)	17 (3%)	0 (0%)	74 (12%)
1	Total	188	24	212 (34%)	36	158	92	23	309 (49%)	31 (5%)	20 (3%)	57 (9%)	629 (100%)
2	Ww	0	0	0 (0%)	0	7	0	0	7 (1%)	0 (0%)	0 (0%)	0 (0%)	7 (1%)
2	Bpc	0	0	0 (0%)	0	9	0	0	9 (2%)	0 (0%)	0 (0%)	0 (0%)	9 (2%)
2	Bw	54	0	54 (9%)	34	158	57	9	258 (44%)	19 (3%)	6 (1%)	18 (3%)	355 (61%)
2	Bc	24	6	30 (5%)	5	17	6	7	35 (6%)	7 (1%)	1 (0%)	6 (1%)	79 (14%)
2	Bs	56	5	61 (11%)	12	8	7	2	29 (5%)	24 (4%)	1 (0%)	8 (1%)	123 (21%)
2	Bcr	2	0	2 (0%)	0	1	0	0	1 (0%)	4 (1%)	0 (0%)	0 (0%)	7 (1%)
2	Total	136	11	147 (25%)	51	200	70	18	339 (58%)	54 (9%)	8 (1%)	32 (6%)	580 (100%)
3	Ww	0	0	0 (0%)	0	5	0	1	6 (1%)	0 (0%)	0 (0%)	0 (0%)	6 (1%)
3	Bpc	0	0	0 (0%)	0	5	0	1	6 (1%)	0 (0%)	0 (0%)	0 (0%)	6 (1%)

3	Bw	56	4	60 (11%)	40	128	33	8	209 (39%)	19 (4%)	4 (1%)	30 (6%)	322 (60%)
3	Bc	22	0	22 (4%)	12	9	5	5	31 (6%)	9 (2%)	0 (0%)	7 (1%)	69 (13%)
3	Bs	49	4	53 (10%)	11	9	6	2	28 (5%)	25 (5%)	1 (0%)	16 (3%)	123 (23%)
3	Bcr	1	1	2 (0%)	1	0	0	0	1 (0%)	3 (1%)	0 (0%)	1 (0%)	7 (1%)
3	Total	128	9	137 (26%)	65	157	49	15	275 (52%)	62 (12%)	5 (1%)	54 (10%)	533 (100%)
4	Ww	0	0	0 (0%)	0	1	0	0	1 (0%)	0 (0%)	0 (0%)	0 (0%)	1 (0%)
4	Bpc	0	0	0 (0%)	0	4	0	0	4 (1%)	0 (0%)	0 (0%)	0 (0%)	4 (1%)
4	Bw	51	0	51 (10%)	34	135	35	6	210 (43%)	14 (3%)	2 (0%)	19 (4%)	296 (61%)
4	Bc	23	0	23 (5%)	8	14	8	5	35 (7%)	6 (1%)	0 (0%)	9 (2%)	73 (15%)
4	Bs	43	4	47 (10%)	4	4	2	0	10 (2%)	20 (4%)	0 (0%)	11 (2%)	88 (18%)
4	Bcr	1	0	1 (0%)	1	1	1	0	3 (1%)	21 (4%)	0 (0%)	0 (0%)	25 (5%)
4	Total	118	4	122 (25%)	47	159	46	11	263 (54%)	61 (13%)	2 (0%)	39 (8%)	487 (100%)
5	Ww	0	0	0 (0%)	0	3	0	0	3 (1%)	0 (0%)	0 (0%)	0 (0%)	3 (1%)
5	Bpc	0	0	0 (0%)	0	3	0	0	3 (1%)	0 (0%)	0 (0%)	0 (0%)	3 (1%)
5	Bw	48	0	48 (11%)	24	93	27	7	151 (35%)	11 (3%)	1 (0%)	11 (3%)	222 (51%)
5	Bc	19	1	20 (5%)	6	9	2	0	17 (4%)	3 (1%)	0 (0%)	5 (1%)	45 (10%)
5	Bs	30	4	34 (8%)	6	2	1	4	13 (3%)	36 (8%)	2 (0%)	2 (0%)	87 (20%)
5	Bcr	3	1	4 (1%)	0	0	0	0	0 (0%)	73 (17%)	0 (0%)	0 (0%)	77 (18%)
5	Total	100	6	106 (24%)	36	110	30	11	187 (43%)	123 (28%)	3 (1%)	18 (4%)	437 (100%)
Total	Ww	1	0	1 (0%)	0	18	0	1	19 (1%)	0 (0%)	0 (0%)	0 (0%)	20 (1%)
Total	Bpc	0	0	0 (0%)	0	29	0	0	29 (1%)	0 (0%)	0 (0%)	1 (0%)	30 (1%)
Total	Bw	275	9	284 (11%)	156	646	214	42	1058 (40%)	80 (3%)	15 (1%)	115 (4%)	1552 (58%)
Total	Bc	112	9	121 (5%)	35	57	37	22	151 (6%)	28 (1%)	1 (0%)	39 (1%)	340 (13%)
Total	Bs	241	22	263 (10%)	36	31	29	12	108 (4%)	114 (4%)	5 (0%)	44 (2%)	534 (20%)
Total	Bcr	41	14	55 (2%)	7	3	2	2	14 (1%)	103 (4%)	17 (0%)	1 (0%)	190 (7%)

Total	Total	670	54	724 (27%)	234	784	282	79	1379 (52%)	325 (12%)	38 (1%)	200 (8%)	2666 (100%)

Note: Ww = Within a word, Bpc = Between Pinyin and character(s), Bw = Between words, Bc = Between clauses, Bs = Between sentences, Bcr = Between contextual revisions

Table 4.15 *Pause-associated cognitive activities by location and stage in L1 Chinese writing (N = 32)*

		Planning			Translation					Monitoring	Others	No recall	Total
Stage	Location	Content	Organisation	Total (%)	Translation in general	Lexical retrieval	Syntactic encoding	Cohesion	Total (%)	Total (%)	Total (%)	Total (%)	Total (%)
1	Ww	0	0	0 (0%)	0	0	0	0	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
1	Bpc	0	0	0 (0%)	0	2	0	0	2 (1%)	0 (0%)	0 (0%)	1 (0%)	3 (1%)
1	Bw	60	2	62 (16%)	36	47	11	6	100 (26%)	11 (3%)	1 (0%)	16 (4%)	190 (50%)
1	Bc	27	4	31 (8%)	13	1	0	4	18 (5%)	4 (1%)	0 (0%)	4 (1%)	57 (15%)
1	Bs	36	6	42 (11%)	7	1	2	5	15 (4%)	8 (2%)	0 (0%)	5 (1%)	70 (18%)
1	Bcr	31	7	38 (10%)	2	0	0	0	2 (1%)	2 (1%)	18 (5%)	0 (0%)	60 (16%)
1	Total	154	19	173 (46%)	58	51	13	15	137 (36%)	25 (7%)	19 (5%)	26 (7%)	380 (100%)
2	Ww	0	0	0 (0%)	0	0	0	0	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
2	Bpc	0	0	0 (0%)	0	4	0	0	4 (1%)	0 (0%)	0 (0%)	0 (0%)	4 (1%)
2	Bw	96	3	99 (23%)	33	86	17	3	139 (32%)	5 (1%)	1 (0%)	15 (3%)	259 (59%)
2	Bc	46	0	46 (10%)	11	3	5	1	20 (5%)	7 (2%)	1 (0%)	7 (2%)	81 (18%)
2	Bs	54	9	63 (14%)	5	4	0	1	10 (2%)	16 (4%)	0 (0%)	3 (1%)	92 (21%)
2	Bcr	1	0	1 (0%)	0	1	0	1	2 (0%)	0 (0%)	0 (0%)	0 (0%)	3 (1%)
2	Total	197	12	209 (48%)	49	98	22	6	175 (40%)	28 (6%)	2 (0%)	25 (6%)	439 (100%)
3	Ww	0	0	0 (0%)	0	0	0	0	0 (0%)	1 (0%)	0 (0%)	0 (0%)	1 (0%)
3	Bpc	0	0	0 (0%)	0	5	0	0	5 (1%)	0 (0%)	0 (0%)	0 (0%)	5 (1%)

3	Bw	89	1	90 (21%)	39	69	10	1	119 (27%)	7 (2%)	1 (0%)	12 (3%)	229 (53%)
3	Bc	54	2	56 (13%)	11	6	0	5	22 (5%)	7 (2%)	0 (0%)	10 (2%)	95 (22%)
3	Bs	58	5	63 (14%)	7	1	2	4	14 (3%)	22 (5%)	1 (0%)	1 (0%)	101 (23%)
3	Bcr	1	0	1 (0%)	0	0	0	0	0 (0%)	3 (1%)	0 (0%)	0 (0%)	4 (1%)
3	Total	202	8	210 (48%)	57	81	12	10	160 (37%)	40 (9%)	2 (0%)	23 (5%)	435 (100%)
4	Ww	0	0	0 (0%)	1	0	0	0	1 (0%)	0 (0%)	0 (0%)	0 (0%)	1 (0%)
4	Bpc	2	0	2 (0%)	0	2	0	0	2 (0%)	0 (0%)	0 (0%)	0 (0%)	4 (1%)
4	Bw	64	1	65 (15%)	52	74	15	6	147 (34%)	8 (2%)	1 (0%)	14 (3%)	235 (55%)
4	Bc	51	1	52 (12%)	21	6	1	2	30 (7%)	6 (1%)	1 (0%)	4 (1%)	93 (22%)
4	Bs	60	2	62 (15%)	5	0	0	3	8 (2%)	13 (3%)	1 (0%)	5 (1%)	89 (21%)
4	Bcr	3	0	3 (1%)	0	1	0	0	1 (0%)	1 (0%)	0 (0%)	0 (0%)	5 (1%)
4	Total	180	4	184 (43%)	79	83	16	11	189 (44%)	28 (7%)	3 (1%)	23 (5%)	427 (100%)
5	Ww	0	0	0 (0%)	0	1	0	0	1 (0%)	0 (0%)	0 (0%)	0 (0%)	1 (0%)
5	Bpc	0	0	0 (0%)	0	1	0	0	1 (0%)	0 (0%)	0 (0%)	0 (0%)	1 (0%)
5	Bw	61	1	62 (16%)	41	68	12	1	122 (31%)	11 (3%)	3 (1%)	10 (3%)	208 (52%)
5	Bc	39	0	39 (10%)	10	3	3	0	16 (4%)	12 (3%)	0 (0%)	4 (1%)	71 (18%)
5	Bs	40	6	46 (12%)	2	3	2	2	9 (2%)	36 (9%)	1 (0%)	4 (1%)	96 (24%)
5	Bcr	2	0	2 (1%)	0	1	0	0	1 (0%)	17 (4%)	0 (0%)	0 (0%)	20 (5%)
5	Total	142	7	149 (38%)	53	77	17	3	150 (38%)	76 (19%)	4 (1%)	18 (5%)	397 (100%)
Total	Ww	0	0	0 (0%)	1	1	0	0	2 (0%)	1 (0%)	0 (0%)	0 (0%)	3 (0%)
Total	Bpc	2	0	2 (0%)	0	14	0	0	14 (1%)	0 (0%)	0 (0%)	1 (0%)	17 (1%)
Total	Bw	370	8	378 (18%)	201	344	65	17	627 (30%)	42 (2%)	7 (0%)	67 (3%)	1121 (54%)
Total	Bc	217	7	224 (11%)	66	19	9	12	106 (5%)	36 (2%)	2 (0%)	29 (1%)	397 (19%)
Total	Bs	248	28	276 (13%)	26	9	6	15	56 (3%)	95 (5%)	3 (0%)	18 (1%)	448 (22%)
Total	Bcr	38	7	45 (2%)	2	3	0	1	6 (0%)	23 (1%)	18 (1%)	0 (0%)	92 (4%)

Total	Total	875	50	925 (45%)	296	390	80	45	811 (39%)	197 (9%)	30 (1%)	115 (6%)	2078 (100%)

Note: Ww = Within a word, Bpc = Between Pinyin and character(s), Bw = Between words, Bc = Between clauses, Bs = Between sentences, Bcr = Between contextual revisions

Stimulated recall comments that shed light on cognitive activities associated with revision behaviours by linguistic domain and context are presented in Tables 4.16 and 4.17. L2 writers referred to translation processes (67%) more often than planning processes (27%) for the whole writing session. Among comments on planning, 93% were associated with content, whereas among comments on translation, lexical retrieval (40%) and syntactic encoding (36%) were the two major reasons for revision. In terms of the linguistic domain, most revisions below the word level (92%), at the word level (66%) and below the clause level (67%) were translation-related. Equally frequent references were made to planning and translation processes (46% and 45%, respectively) for revisions at the clause level and above. In terms of context, the majority of pre-contextual (68%) and contextual (66%) revisions were translation-related.

In line with the trends observed for the whole writing session, L2 writers referred to translation about 70% of the time in each writing stage, with the majority of comments mentioning lexical retrieval and syntactic encoding. A gradual decrease was found in the percentage of comments on planning from the beginning to the end. When considering the linguistic domain, below-word, word-level and below-clause revisions were mostly translation-related, whereas revisions at the clause level and above were primarily linked to translation in Stages 1 and 2, but associated more with planning after Stage 3. In terms of context, the majority of pre-contextual and contextual revisions described translation processes regardless of the stage. The percentage of contextual revision mentioning translation processes increased considerably in Stage 5, indicating a dominant focus on language when L2 writers were systematically reviewing and editing their texts.

In terms of revision-associated cognitive activities in L1 writing (Tables 18 and 19), the main reason for revisions was, again, related to translation processes (56%) for the whole session. However, L1 writers seemed to have a more balanced allocation of focus on language, with lexical retrieval accounting for the largest proportion (37%), followed by translation in general (27%) and syntactic encoding (25%). For 37% of the time, L1 writers mentioned planning activities when revising, for which 89% of the comments were associated with making changes to the content of writing. In line with the trend reported by L2 writers, L1 writers made most references to translation when revising smaller textual units (65% for below-word revision, 63% for word-level revision and 58% for below-clause revision). However, different from L2 writers, L1 writers attended to content and organisation (58%) more often than language (35%) when making revisions at the clause level and above. In terms of context, the majority of pre-contextual (56%) and contextual (57%) revisions were linked to translation, which is similar to L2 writing.

Like L2 writers, L1 writers referred to translation processes more often than planning processes in each stage when making revisions. However, countering the trends observed in L2 writing, Stage 1 in L1 writing featured a larger percentage of comments on translation than in the latter stages; proportionally fewer planning-related comments were made in the initial stage, though the percentage increased to around 37% from Stage 2 onwards. When considering the linguistic domain, most revisions to small textual units in L1 writing were also found to describe translation, regardless of the stage. Nevertheless, L1 writers revised larger textual units mainly for planning in every stage, different from L2 writers who started doing that after Stage 2. Similar to L2 writing, both pre-contextual and contextual revisions in L1 writing were primarily translation-related across stages. However in L1 writing, the percentage of comments on translation processes for contextual revisions stayed relatively stable from Stage 1 to Stage 5. This suggests that L1 writers tended to constantly go back and make language changes to already-written text rather than postponing them to the end of writing.

Table 4.16 *Revision-associated cognitive activities by linguistic domain and stage in L2 Chinese writing (N = 32)*

		Planning			Translation					Others	No recall	Total
Stage	Linguistic domain	Content	Organisation	Total (%)	Translation in general	Lexical retrieval	Syntactic encoding	Cohesion	Total (%)	Total (%)	Total (%)	Total (%)
1	Below word	2	0	2 (1%)	0	8	0	0	8 (5%)	0 (0%)	0 (0%)	10 (6%)
1	Word	8	0	8 (5%)	6	19	15	4	44 (26%)	0 (0%)	5 (3%)	57 (34%)
1	Below clause	23	2	25 (15%)	7	13	17	3	40 (24%)	0 (0%)	5 (3%)	70 (41%)
1	Clause and above	5	7	12 (7%)	0	3	12	4	19 (11%)	1 (1%)	1 (1%)	33 (19%)
1	Total	38	9	47 (28%)	13	43	44	11	111 (65%)	1 (1%)	11 (6%)	170 (100%)
2	Below word	0	0	0 (0%)	0	14	1	0	15 (7%)	0 (0%)	0 (0%)	15 (7%)
2	Word	16	1	17 (8%)	7	20	17	3	47 (23%)	0 (0%)	9 (4%)	73 (35%)
2	Below clause	15	2	17 (8%)	14	20	19	3	56 (27%)	0 (0%)	5 (2%)	78 (38%)
2	Clause and above	6	6	12 (6%)	3	2	13	5	23 (11%)	1 (0%)	4 (2%)	40 (19%)
2	Total	37	9	46 (22%)	24	56	50	11	141 (68%)	1 (0%)	18 (9%)	206 (100%)
3	Below word	0	0	0 (0%)	0	14	0	1	15 (8%)	0 (0%)	3 (2%)	18 (10%)
3	Word	7	3	10 (5%)	8	27	20	2	57 (31%)	0 (0%)	6 (3%)	73 (40%)
3	Below clause	10	2	12 (7%)	8	8	13	3	32 (18%)	2 (1%)	7 (4%)	53 (29%)
3	Clause and above	14	4	18 (10%)	5	3	10	1	19 (10%)	0 (0%)	1 (1%)	38 (21%)
3	Total	31	9	40 (22%)	21	52	43	7	123 (68%)	2 (1%)	17 (9%)	182 (100%)
4	Below word	0	0	0 (0%)	0	6	2	1	9 (5%)	0 (0%)	0 (0%)	9 (5%)
4	Word	13	0	13 (7%)	10	19	18	5	52 (29%)	0 (0%)	6 (3%)	71 (39%)

4	Below clause	13	0	13 (7%)	8	16	19	2	45 (25%)	0 (0%)	6 (3%)	64 (36%)
4	Clause and above	20	4	24 (13%)	0	2	4	1	7 (4%)	0 (0%)	5 (3%)	36 (20%)
4	Total	46	4	50 (28%)	18	43	43	9	113 (63%)	0 (0%)	17 (9%)	180 (100%)
5	Below word	0	0	0 (0%)	0	7	2	1	10 (5%)	0 (0%)	0 (0%)	10 (5%)
5	Word	6	0	6 (3%)	3	16	31	10	60 (29%)	0 (0%)	9 (4%)	75 (36%)
5	Below clause	17	2	19 (9%)	9	12	29	2	52 (25%)	0 (0%)	1 (0%)	72 (35%)
5	Clause and above	21	4	25 (12%)	6	4	10	3	23 (11%)	0 (0%)	1 (0%)	50 (24%)
5	Total	44	6	50 (24%)	18	39	72	16	145 (70%)	0 (0%)	11 (5%)	207 (100%)
Total	Below word	2	0	2 (0%)	0	49	5	3	57 (6%)	0 (0%)	3 (0%)	62 (7%)
Total	Word	50	4	54 (6%)	34	101	101	24	260 (28%)	0 (0%)	35 (4%)	349 (37%)
Total	Below clause	78	8	86 (9%)	46	69	97	13	225 (24%)	2 (0%)	24 (3%)	337 (36%)
Total	Clause and above	66	25	91 (10%)	14	11	50	14	89 (9%)	5 (1%)	12 (1%)	197 (21%)
Total	Total	196	37	233 (25%)	94	230	253	54	631 (67%)	7 (1%)	74 (8%)	945 (100%)

Table 4.17 *Revision-associated cognitive activities by context and stage in L2 Chinese writing (N = 32)*

		Planning			Translation					Others	No recall	Total
Stage	Context	Content	Organisation	Total (%)	Translation in general	Lexical retrieval	Syntactic encoding	Cohesion	Total (%)	Total (%)	Total (%)	Total (%)
1	Pre-contextual	28	3	31 (18%)	9	38	34	8	89 (52%)	1 (1%)	8 (5%)	129 (76%)
1	Contextual	10	6	16 (9%)	4	5	10	3	22 (13%)	0 (0%)	3 (2%)	41 (24%)
1	Total	38	9	47 (28%)	13	43	44	11	111 (65%)	1 (1%)	11 (6%)	170 (100%)
2	Pre-contextual	19	4	23 (11%)	15	47	22	6	90 (44%)	0 (0%)	15 (7%)	128 (62%)
2	Contextual	18	5	23 (11%)	9	9	28	5	51 (25%)	1 (0%)	3 (1%)	78 (38%)
2	Total	37	9	46 (22%)	24	56	50	11	141 (68%)	1 (0%)	18 (9%)	206 (100%)
3	Pre-contextual	17	4	21 (12%)	13	46	26	4	89 (49%)	2 (1%)	13 (7%)	125 (69%)
3	Contextual	14	5	19 (10%)	8	6	17	3	34 (19%)	0 (0%)	4 (2%)	57 (31%)
3	Total	31	9	40 (22%)	21	52	43	7	123 (68%)	2 (1%)	17 (9%)	182 (100%)
4	Pre-contextual	26	0	26 (14%)	12	35	15	3	65 (36%)	0 (0%)	11 (6%)	102 (57%)
4	Contextual	20	4	24 (13%)	6	8	28	6	48 (27%)	0 (0%)	6 (3%)	78 (43%)
4	Total	46	4	50 (28%)	18	43	43	9	113 (63%)	0 (0%)	17 (9%)	180 (100%)
5	Pre-contextual	19	1	20 (10%)	6	20	17	5	48 (23%)	1 (0%)	8 (4%)	77 (37%)
5	Contextual	25	5	30 (14%)	12	19	55	11	97 (47%)	0 (0%)	3 (1%)	130 (63%)
5	Total	44	6	50 (24%)	18	39	72	16	145 (70%)	1 (0%)	11 (5%)	207 (100%)
Total	Pre-contextual	109	12	121 (13%)	55	183	115	26	379 (40%)	6 (1%)	55 (6%)	561 (59%)
Total	Contextual	87	25	112 (12%)	39	47	138	28	252 (27%)	1 (0%)	19 (2%)	384 (41%)

Total	Total	196	37	233 (25%)	94	230	253	54	631 (67%)	7 (1%)	74 (8%)	945 (100%)

Table 4.18 *Revision-associated cognitive activities by linguistic domain and stage in L1 Chinese writing (N = 32)*

		Planning			Translation					Others	No recall	Total
Stage	Linguistic domain	Content	Organisation	Total (%)	Translation in general	Lexical retrieval	Syntactic encoding	Cohesion	Total (%)	Total (%)	Total (%)	Total (%)
1	Below word	3	0	3 (1%)	0	3	0	0	3 (1%)	0 (0%)	0 (0%)	6 (3%)
1	Word	26	3	29 (12%)	10	33	18	11	72 (30%)	0 (0%)	7 (3%)	108 (45%)
1	Below clause	20	1	21 (9%)	25	18	10	4	57 (24%)	0 (0%)	9 (4%)	87 (36%)
1	Clause and above	14	8	22 (9%)	5	1	9	2	17 (7%)	0 (0%)	0 (0%)	39 (16%)
1	Total	63	12	75 (31%)	40	55	37	17	149 (62%)	0 (0%)	16 (7%)	240 (100%)
2	Below word	0	0	0 (0%)	0	9	2	0	11 (3%)	0 (0%)	0 (0%)	11 (3%)
2	Word	41	5	46 (14%)	11	35	24	8	78 (24%)	2 (1%)	13 (4%)	139 (42%)
2	Below clause	44	5	49 (15%)	21	21	25	5	72 (22%)	0 (0%)	12 (4%)	133 (40%)
2	Clause and above	25	2	27 (8%)	4	2	6	3	15 (5%)	0 (0%)	4 (1%)	46 (14%)
2	Total	110	12	122 (37%)	36	67	57	16	176 (53%)	2 (1%)	29 (9%)	329 (100%)
3	Below word	2	0	2 (1%)	0	6	0	0	6 (2%)	0 (0%)	0 (0%)	8 (3%)
3	Word	38	1	39 (12%)	15	46	19	11	91 (28%)	1 (0%)	6 (2%)	137 (43%)
3	Below clause	40	4	44 (14%)	17	21	16	3	57 (18%)	0 (0%)	8 (3%)	109 (34%)
3	Clause and above	34	7	41 (13%)	8	4	7	3	22 (7%)	0 (0%)	3 (1%)	66 (21%)
3	Total	114	12	126 (39%)	40	77	42	17	176 (55%)	1 (0%)	17 (5%)	320 (100%)
4	Below word	3	0	3 (1%)	2	0	0	0	2 (1%)	0 (0%)	0 (0%)	5 (2%)

4	Word	36	1	37 (13%)	16	28	17	7	68 (24%)	2 (1%)	7 (2%)	114 (40%)
4	Below clause	33	0	33 (12%)	36	14	15	4	69 (24%)	0 (0%)	8 (3%)	110 (39%)
4	Clause and above	31	7	38 (13%)	4	0	6	5	15 (5%)	0 (0%)	3 (1%)	56 (20%)
4	Total	103	8	111 (39%)	58	42	38	16	154 (54%)	2 (1%)	18 (6%)	285 (100%)
5	Below word	3	0	3 (1%)	0	1	1	0	2 (1%)	0 (0%)	2 (1%)	7 (2%)
5	Word	25	3	28 (9%)	10	39	14	13	76 (25%)	0 (0%)	8 (3%)	112 (36%)
5	Below clause	39	1	40 (13%)	29	20	5	11	65 (21%)	0 (0%)	8 (3%)	113 (37%)
5	Clause and above	27	9	36 (12%)	7	4	15	4	30 (10%)	2 (1%)	7 (2%)	75 (24%)
5	Total	94	13	107 (35%)	46	64	35	28	173 (56%)	2 (1%)	25 (8%)	307 (100%)
Total	Below word	11	0	11 (1%)	2	19	3	0	24 (2%)	0 (0%)	2 (0%)	37 (2%)
Total	Word	166	13	179 (12%)	62	181	92	50	385 (26%)	5 (0%)	41 (3%)	610 (41%)
Total	Below clause	176	11	187 (13%)	128	94	71	27	320 (22%)	0 (0%)	45 (3%)	552 (37%)
Total	Clause and above	131	33	164 (11%)	28	11	43	17	99 (7%)	2 (0%)	17 (1%)	282 (19%)
Total	Total	484	57	541 (37%)	220	305	209	94	828 (56%)	7 (0%)	105 (7%)	1481 (100%)

Table 4.19 *Revision-associated cognitive activities by context and stage in L1 Chinese writing (N = 32)*

		Planning			Translation					Others	No recall	Total
Stage	Context	Content	Organisation	Total (%)	Translation in general	Lexical retrieval	Syntactic encoding	Cohesion	Total (%)	Total (%)	Total (%)	Total (%)
1	Pre-contextual	50	7	57 (24%)	35	47	31	9	122 (51%)	0 (0%)	14 (6%)	193 (80%)
1	Contextual	13	5	18 (8%)	5	8	6	8	27 (11%)	0 (0%)	2 (1%)	47 (20%)
1	Total	63	12	75 (31%)	40	55	37	17	149 (62%)	0 (0%)	16 (7%)	240 (100%)
2	Pre-contextual	82	6	88 (27%)	28	56	38	10	132 (40%)	2 (1%)	27 (8%)	249 (76%)
2	Contextual	28	6	34 (10%)	8	11	19	6	44 (13%)	0 (0%)	2 (1%)	80 (24%)
2	Total	110	12	122 (37%)	36	67	57	16	176 (53%)	2 (1%)	29 (9%)	329 (100%)
3	Pre-contextual	92	8	100 (31%)	35	61	35	10	141 (44%)	1 (0%)	16 (5%)	258 (81%)
3	Contextual	22	4	26 (8%)	5	16	7	7	35 (11%)	0 (0%)	1 (0%)	62 (19%)
3	Total	114	12	126 (39%)	40	77	42	17	176 (55%)	1 (0%)	17 (5%)	320 (100%)
4	Pre-contextual	83	5	88 (31%)	49	37	30	10	126 (44%)	2 (1%)	17 (6%)	233 (82%)
4	Contextual	20	3	23 (8%)	9	5	8	6	28 (10%)	0 (0%)	1 (0%)	52 (18%)
4	Total	103	8	111 (39%)	58	42	38	16	154 (54%)	2 (1%)	18 (6%)	285 (100%)
5	Pre-contextual	75	5	80 (26%)	36	50	22	14	122 (40%)	0 (0%)	21 (7%)	223 (73%)
5	Contextual	19	8	27 (9%)	10	14	13	14	51 (17%)	2 (1%)	4 (1%)	84 (27%)
5	Total	94	13	107 (35%)	46	64	35	28	173 (56%)	2 (1%)	25 (8%)	307 (100%)
Total	Pre-contextual	382	31	413 (28%)	183	251	156	53	643 (43%)	5 (0%)	95 (6%)	1156 (78%)
Total	Contextual	102	26	128 (9%)	37	54	53	41	185 (12%)	2 (0%)	10 (1%)	325 (22%)

Total	Total	484	57	541 (37%)	220	305	209	94	828 (56%)	7 (0%)	105 (7%)	1481 (100%)

Summary: the role of stage in Chinese writing Stages were found to have an influence on L2 Chinese writing behaviours, and the differences in writing behaviours mainly set apart Stage 1 and/or Stage 5 from the middle stages. Speed fluency tended to be highest at the beginning and lowest at the end. More pauses occurred in the middle than in the initial and final stages of writing, apart from between-contextual-revision pauses, which occurred more often at the end, whereas longer pauses were found at the beginning and end than in the middle stages. L2 writers revised more frequently in the middle stages than in the first and last stages of writing, except that most contextual revisions were made at the end of writing. The distribution of stimulated recall comments that provided information on cognitive activities associated with writing behaviours also showed some changes across writing stages. L2 writers reported proportionally more planning processes during pausing in Stage 1 and proportionally more monitoring processes in Stage 5 than in other stages. The percentage of comments on translation was higher in the middle than at the beginning or the end. In terms of the reasons for revision, references made to translation processes remained relatively stable across stages, while those made to planning processes dropped from the beginning to the end.

The effects of stage also emerged for L1 Chinese writing behaviours. Similar to L2 writers, L1 writers tended to be more fluent at the beginning and less fluent at the end. The distribution of pauses at different locations across stages was largely similar for L1 and L2 writing; however, some slight differences were found in terms of the frequencies of pauses between Pinyin and character(s), between words, and the duration of pauses between sentences in the initial stages, as well as the frequencies of pauses overall and between clauses in the final stage. L1 writers were also similar to L2 writers in their temporal distribution of revisions in terms of linguistic domain, context and level of transcription. As for cognitive activities associated with pausing behaviours, the temporal distribution of stimulated recall comments revealed differences between L1 and L2 writers in terms of planning and translation processes. L1 writers reported a main focus on planning before Stage 4, whereas the dominant focus for L2 writers was on translation throughout the five stages. With regard to cognitive activities associated with revision behaviours, the distribution patterns were also different, although revisions were primarily related to translation processes in both L1 and L2 writing. Unlike L2 writing, L1 writing featured a lower percentage of planning but a higher percentage of translation in Stage 1 than in subsequent stages.

4.2 Genre, writing behaviours and cognitive activities in Chinese writing

This section deals with the impact of genre on the cognitive processes involved in Chinese writing. To begin with, I examined the extent to which genre (i.e. argumentative versus narrative) affected writing behaviours. Tables 4.20–4.22 summarise the descriptive statistics for speed fluency, pausing and revision measures by genre (for complete tables, see Tables 4–6 in Appendix G).

First, the effect of genre on writing behaviours was explored for the whole session. Two sets of linear mixed-effects regressions were performed, respectively, on the data obtained from L1 and L2 writers, with writing behaviour indices as the dependent variables and genre as the fixed effect. Random effects were the participants (and prompt) depending on whether or not the model was able to converge with prompt as a second random effect. A by-participant random slope for genre was added to the models to take into account participant-by-genre variation. However, it was only retained in two analyses, as the rest of the models failed to converge with a by-participant random slope. A log transformation was made to a dependent variable if residual plots suggested poorly fitting models. For some models, a square-root transformation was performed instead when the log transformation failed to improve the goodness of fit.

Table 4.20 *Speed fluency by genre in Chinese writing*

	Argumentative				Narrative			
Measure	*M*	*SD*	95% *IC* lower	95% *IC* upper	*M*	*SD*	95% *IC* lower	95% *IC* upper
L2 Chinese (N =32)								
Production rate	43.95	8.57	41.85	46.05	42.38	8.73	40.24	44.52
P-burst length	4.47	1.37	4.13	4.81	4.26	1.42	3.91	4.61
L1 Chinese (N =32)								
Production rate	64.63	11.28	61.87	67.40	61.92	10.87	59.26	64.58
P-burst length	5.77	1.59	5.38	6.16	5.65	1.69	5.23	6.07

Note: Data by stages are available in Table 4 in Appendix G, Production rate = Number of characters produced divided by active writing time (in minutes), P-burst length = Number of characters between two pauses

Table 4.21 *Pause frequency and duration by genre and location in Chinese writing*

	Argumentative								Narrative							
	Frequency (Number of pauses per minute)				Duration (median length of pause)				Frequency (Number of pauses per minute)				Duration (median length of pause)			
Location	*M*	*SD*	95% *CI* low	95% *CI* up	*M*	*SD*	95% *CI* low	95% *CI* up	*M*	*SD*	95% *CI* low	95% *CI* up	*M*	*SD*	95% *CI* low	95% *CI* up
L2 Chinese (N =32)																
Total	4.21	.84	4.00	4.42	4.27	.85	4.06	4.48	4.14	.86	3.93	4.35	4.11	.93	3.88	4.34
Ww	.18	.11	.15	.21	3.22	1.31	2.89	3.55	.15	.09	.13	.17	3.17	1.33	2.84	3.50
Bpc	.48	.53	.35	.61	3.04	.63	2.88	3.20	.51	.46	.40	.62	2.88	.54	2.75	3.01
Bw	2.02	.58	1.88	2.16	3.84	.81	3.64	4.04	1.92	.53	1.79	2.05	3.67	.69	3.50	3.84
Bc	.46	.16	.42	.50	4.78	1.88	4.32	5.24	.49	.20	.44	.54	4.50	1.67	4.09	4.91
Bs	.73	.21	.68	.78	9.39	5.46	8.05	1.73	.73	.22	.68	.78	8.56	5.01	7.33	9.79
Bcr	.34	.30	.27	.41	9.70	11.06	6.90	12.50	.34	.32	.26	.42	8.66	6.69	6.99	1.33
L1 Chinese (N =32)																
Total	3.61	.72	3.44	3.79	3.83	.62	3.68	3.99	3.39	.67	3.22	3.55	3.79	.64	3.63	3.94
Ww	.07	.07	.05	.08	3.17	1.19	2.82	3.53	.06	.08	.04	.08	3.20	2.06	2.55	3.86
Bpc	.13	.17	.09	.17	2.85	.94	2.59	3.11	.17	.16	.13	.21	3.03	1.79	2.57	3.50
Bw	1.68	.53	1.55	1.81	3.39	.55	3.25	3.52	1.47	.38	1.37	1.56	3.32	.47	3.21	3.44
Bc	.73	.30	.65	.80	4.32	1.61	3.92	4.71	.70	.30	.63	.77	4.04	1.18	3.75	4.33
Bs	.76	.21	.71	.81	7.18	3.06	6.43	7.93	.73	.26	.67	.79	6.31	2.38	5.73	6.90
Bcr	.25	.20	.20	.30	5.21	2.55	4.55	5.88	.26	.22	.20	.31	6.93	7.33	5.01	8.85

Note: Data by stages are available in Table 5 in Appendix G, Ww = Within a word, Bpc = Between Pinyin and character(s), Bw = Between words, Bc = Between clauses, Bs = Between sentences, Bcr = Between contextual revisions

Table 4.22 *Revision measures by genre and dimension in Chinese writing*

	Argumentative				Narrative			
	M	*SD*	95% *CI* low	95% *CI* up	*M*	*SD*	95% *CI* low	95% *CI* up
L2 Chinese (N =32)								
Total revision	3.85	2.63	3.21	4.49	3.86	2.47	3.25	4.46
Linguistic domain								
Below word	1.84	1.54	1.46	2.22	1.86	1.48	1.50	2.22
Word	1.06	.68	.89	1.23	1.04	.62	.89	1.19
Below clause	.64	.43	.54	.75	.65	.38	.56	.74
Clause and above	.16	.14	.13	.19	.14	.11	.11	.17
Context								
Pre-contextual	3.29	2.60	2.65	3.93	3.29	2.41	2.70	3.88
Contextual	.56	.40	.46	.66	.57	.44	.46	.68
Level of transcription								
Pinyin	2.27	1.63	1.87	2.67	2.28	1.61	1.89	2.67
Character	1.43	1.12	1.16	1.70	1.41	.95	1.18	1.64
L1 Chinese (N =32)								
Total revision	6.30	2.06	5.80	6.81	6.05	1.96	5.57	6.53
Linguistic domain								
Below word	2.57	1.02	2.32	2.82	2.44	.89	2.23	2.66
Word	1.70	.68	1.53	1.87	1.53	.56	1.39	1.67
Below clause	1.53	.59	1.38	1.67	1.59	.72	1.41	1.77
Clause and above	.28	.22	.22	.33	.26	.20	.21	.31
Context								
Pre-contextual	5.66	2.00	5.17	6.14	5.44	1.88	4.98	5.90

Contextual	.65	.50	.53	.77	.61	.47	.50	.73
Level of transcription								
Pinyin	2.73	1.10	2.46	3.00	2.67	1.06	2.41	2.93
Character	3.34	1.41	3.00	3.69	3.15	1.30	2.83	3.47

Note: Data by stages are available in Table 6 in Appendix G

As shown in Table 4.23, the effects of genre were found in three writing behaviour indices in L2 writing, i.e. production rate, P-burst length and within-word pause frequency, and two indices in L1 writing, i.e. production rate and between-word pause frequency. The results indicate that L2 writers tended to show greater speed fluency when performing argumentative writing tasks as compared to narrative writing tasks. However, they paused more often within a word when writing argumentative essays than narrative essays. Similarly, L1 writers showed greater speed fluency, as measured by production rate, in argumentative than narrative writing. Slightly different from L2 writers, L1 writers paused more frequently between words rather than within a word when producing argumentative than narrative essays. The amount of variance contributed by genre was less than 5% for all five writing behaviour indices. Non-significant results for the effect of genre on writing behaviour indices are reported in Tables 3 and 4 in Appendix H.

Table 4.23 *Significant results from linear mixed-effects regressions examining the effects of genre on writing behaviours in Chinese*

Dependent variable	Fixed effect: genre	*Est*	*SE*	*t*	*p*
L2 Chinese (N = 32)					
Production rate	Narrative	-1.57	.69	-2.30	.02
Model = lmer(production_rate ~ genre + (1|participant), data, REML=F)					
$R^2_m = .01$, $R^2_c = .80$					
P-burst length	Narrative	-.22	.10	-2.12	.04
Model = lmer(pburst_length ~ genre + (1|participant), data, REML=F)					
$R^2_m = .01$, $R^2_c = .80$					
Ww pause frequency	Narrative	-.03	.01	-2.34	.02
Model = lmer(log(ww_pause_frequency) ~ genre + (1|participant), data, REML=F)					
$R^2_m = .02$, $R^2_c = .48$					
L1 Chinese (N = 32)					
Production rate	Narrative	-2.71	.79	-3.44	.04
Model = lmer(production_rate ~ genre + (1|participant) + (1|prompt), data, REML=F)					

R^2_m = .02, R^2_c = .90					
Bw pause frequency	Narrative	-.21	.06	-3.47	.01
Model = lmer(bw_pause_frequency ~ genre + (1|participant), data, REML=F)					
R^2_m = .05, R^2_c = .45					

Note: Production rate = Number of characters produced divided by active writing time (in minutes), P-burst length = Number of characters between two pauses, Ww = Within a word, Bw = Between words, R^2_m = marginal *R*-squared, R^2_c = conditional *R*-squared

Second, I looked at the extent to which genre affected writing behaviours at different stages of writing. In the series of linear mixed-effects regression models that I constructed, the dependent variable was, again, an index of writing behaviour, and the fixed effects were genre, stages of writing and their interaction. Participant was the random effect of all models. Prompt was included as the second random effect in some models when convergence was achieved. Participant-by-genre and participant-by-stage random slopes were added to account for potentially differential genre and stage effects on the participants. However, only a by-participant random slope was retained in some models to ensure convergence. Dependent variables were either log- or square-root-transformed when needed.

The effects of interaction between genre and stage were identified in nine writing behaviour indices in L2 Chinese writing (Table 4.24): total pause frequency, between-Pinyin-and-character(s) pause frequency, between-clause pause frequency, between-Pinyin-and-character(s) pause duration, and the amount of revision in total, below the word level, below the clause level and in Pinyin. In L1 writing, interactions were found for 12 behavioural measures (Table 4.25), including the frequency and duration of pauses overall, between-Pinyin-and-character(s) and between clauses, between-sentence pause duration, total revision, below-word revision, pre-contextual revision, Pinyin revision, and character revision. Refer to Tables 5 and 6 in Appendix

H for non-significant effects of interaction between genre and stage in L2 and L1 writing, respectively. The amount of variance explained by the interaction ranged between 1% and 4%. This suggests that L1 and L2 writers of Chinese did, to some extent, behave differently when performing argumentative and narrative writing tasks at different stages.

Table 4.24 *Significant results from linear mixed-effects regressions examining the effects of interaction between genre and stage on writing behaviours in L2 Chinese (N = 32)*

Dependent variable	Predictor: genre * stage	*Est*	*SE*	*t*	*p*
Total pause frequency	Narrative * 2	.51	.22	2.25	.02*
	Narrative * 3	.70	.22	3.12	< .01*
	Narrative * 4	.24	.22	1.06	.29
	Narrative * 5	.36	.22	1.59	.11
Model = lmer(total_pause_frequency ~ genre*stage + (1|participant) + (1|prompt), data, REML = F)					
Amount of variance explained by the interaction: $R^2_m = .01$, $R^2_c = .01$					
Bpc pause frequency	Narrative * 2	.07	.05	1.41	.16
	Narrative * 3	.12	.05	2.31	.02*
	Narrative * 4	.04	.05	.89	.37
	Narrative * 5	.04	.05	.74	.46
Model = lmer(log(bpc_pause_frequency) ~ genre*stage + (1|participant) + (1|prompt), data, REML = F)					
Amount of variance explained by the interaction: $R^2_m < .01$, $R^2_c = .01$					
Bc pause frequency	Narrative * 2	.08	.05	1.68	.09
	Narrative * 3	.09	.05	1.73	.08
	Narrative * 4	.12	.05	2.46	.01*
	Narrative * 5	.10	.05	1.93	.05
Model = lmer(log(bc_pause_frequency) ~ genre*stage + (1|participant), data, REML = F)					
Amount of variance explained by the interaction: $R^2_m = .01$, $R^2_c = .01$					
Bpc pause duration	Narrative * 2	.19	.06	3.04	< .01*
	Narrative * 3	.19	.06	3.14	< .01*
	Narrative * 4	.06	.06	.98	.33
	Narrative * 5	.09	.06	1.37	.17
Model = lmer(1/log(bpc_pause_duration) ~ genre*stage + (1|participant), data, REML = F)					
Amount of variance explained by the interaction: $R^2_m = .03$, $R^2_c = .03$					
Total revision	Narrative * 2	.20	.06	3.23	< .01*

	Narrative * 3	.21	.06	3.40	< .01*
	Narrative * 4	.18	.06	2.90	< .01*
	Narrative * 5	.11	.06	1.81	.07
Model = lmer(log(total_revision) ~ genre*stage + (1|participant), data, REML=F)					
Amount of variance explained by the interaction: R^2_m = .01, R^2_c = .01					
Below word revision	Narrative * 2	.16	.07	2.43	.02*
	Narrative * 3	.20	.07	3.01	< .01*
	Narrative * 4	.18	.07	2.72	< .01*
	Narrative * 5	.09	.07	1.40	.16
Model = lmer(log(below_word_revision) ~ genre*stage + (1+genre|participant), data, REML=F)					
Amount of variance explained by the interaction: R^2_m = .01, R^2_c = .01					
Below clause revision	Narrative * 2	.09	.05	1.67	.09
	Narrative * 3	.11	.05	2.19	.03*
	Narrative * 4	.04	.05	.75	.45
	Narrative * 5	.04	.05	.84	.40
Model = lmer(log(below_clause_revision) ~ genre*stage + (1+genre|participant) + (1|prompt), data, REML=F)					
Amount of variance explained by the interaction: R^2_m = .01, R^2_c = .01					
Pre-contextual revision	Narrative * 2	.21	.07	2.90	< .01*
	Narrative * 3	.19	.07	2.69	< .01*
	Narrative * 4	.17	.07	2.32	.02*
	Narrative * 5	.04	.07	.58	.56
Model = lmer(log(precontextual_revision) ~ genre*stage + (1+genre|participant), data, REML=F)					
Amount of variance explained by the interaction: R^2_m = .01, R^2_c = .01					
Pinyin revision	Narrative * 2	.21	.07	2.96	< .01*
	Narrative * 3	.22	.07	3.18	< .01*
	Narrative * 4	.18	.07	2.58	.01*
	Narrative * 5	.10	.07	1.38	.17
Model = lmer(log(pinyin_revision) ~ genre*stage + (1+genre|participant), data, REML=F)					
Amount of variance explained by the interaction: R^2_m = .01, R^2_c = .01					

Note: Bpc = Between Pinyin and character(s), Bc = Between clauses, R^2_m = marginal *R*-squared, R^2_c = conditional *R*-squared, * = statistically significant results

Table 4.25 *Significant results from linear mixed-effects regressions examining the effects of interaction between genre and stage on writing behaviours in L1 Chinese (N = 32)*

	Predictor: genre * stage	*E*	*SE*	*t*	*p*
Total pause frequency	Narrative * 2	.46	.21	2.17	.03*
	Narrative * 3	.47	.21	2.21	.02*
	Narrative * 4	.31	.21	1.47	.14
	Narrative * 5	.45	.21	2.11	.04*
Model = lmer(total_pause_frequency ~ genre*stage + (1+genre|participant) + (1|prompt), data,					

REML = F)

Amount of variance explained by the interaction: $R^2_m = .01$, $R^2_c = .01$

Bpc pause frequency	Narrative * 2	.06	.05	1.33	.18
	Narrative * 3	.03	.05	.70	.48
	Narrative * 4	.10	.05	2.25	.03*
	Narrative * 5	.02	.05	.53	.60

Model = lmer(bpc_pause_frequency ~ genre*stage + (1|participant) + (1|prompt), data, REML = F)

Amount of variance explained by the interaction: $R^2_m = .01$, $R^2_c = .01$

Bc pause frequency	Narrative * 2	.80	.10	.78	.43
	Narrative * 3	.29	.10	2.87	< .01*
	Narrative * 4	.25	.10	2.43	.02*
	Narrative * 5	.07	.10	.73	.47

Model = lmer(bc_pause_frequency ~ genre*stage + (1|participant) + (1|prompt), data, REML = F)

Amount of variance explained by the interaction: $R^2_m = .01$, $R^2_c = .01$

Total pause duration	Narrative * 2	.03	.03	1.06	.29
	Narrative * 3	.07	.03	2.14	.03*
	Narrative * 4	.04	.03	1.35	.18
	Narrative * 5	.01	.03	.22	.83

Model = lmer(1/log(total_pause_duration) ~ genre*stage + (1+genre|participant), data, REML = F)

Amount of variance explained by the interaction: $R^2_m = .01$, $R^2_c = .01$

Bpc pause duration	Narrative * 2	< .01	.09	.04	.97
	Narrative * 3	.02	.09	2.26	.02*
	Narrative * 4	-.07	.09	.72	.47
	Narrative * 5	.13	.10	1.30	.20

Model = lmer(1/log(bpc_pause_duration) ~ genre*stage + (1|participant), data, REML = F)

Amount of variance explained by the interaction: $R^2_m = .04$, $R^2_c = .06$

Bc pause duration	Narrative * 2	-.30	.11	-2.68	.01*
	Narrative * 3	-.10	.11	-.94	.35
	Narrative * 4	-.09	.11	-.83	.41
	Narrative * 5	-.17	.11	-1.52	.13

Model = lmer(log(bc_pause_duration) ~ genre*stage + (1+genre|participant) + (1|prompt), data, REML = F)

Amount of variance explained by the interaction: $R^2_m = .01$, $R^2_c = .01$

Bs pause duration	Narrative * 2	-.41	.16	-2.53	.01*
	Narrative * 3	-.65	.16	-4.01	< .01*
	Narrative * 4	-.42	.16	-2.57	.01*
	Narrative * 5	-.32	.16	-1.99	.05*

Model = lmer(log(bs_pause_duration) ~ genre*stage + (1|participant) + (1|prompt), data, REML = F)

Amount of variance explained by the interaction: $R^2_m = .02$, $R^2_c = .02$

Total revision	Narrative * 2	.06	.40	1.50	.13
	Narrative * 3	1.30	.40	3.25	< .01*

	Narrative * 4	.75	.40	1.88	.06
	Narrative * 5	.47	.40	1.17	.4
Model = lmer(total_revision ~ genre*stage + (1|participant) + (1|prompt), data, REML = F)					
Amount of variance explained by the interaction: R^2_m = .01, R^2_c = .01					
Below word revision	Narrative * 2	.17	.07	2.39	.02*
	Narrative * 3	.27	.07	3.89	< .01*
	Narrative * 4	.20	.07	2.78	.01*
	Narrative * 5	.05	.07	.75	.45
Model = lmer(log(below_word_revision) ~ genre*stage + (1+genre|participant) + (1|prompt), data, REML = F)					
Amount of variance explained by the interaction: R^2_m = .02, R^2_c = .02					
Pre-contextual revision	Narrative * 2	.70	.40	1.72	.09
	Narrative * 3	1.37	.40	3.39	< .01*
	Narrative * 4	.87	.40	2.15	.03*
	Narrative * 5	.47	.40	1.16	.24
Model = lmer(precontextual_revision ~ genre*stage + (1+genre|participant) + (1|prompt), data, REML=F)					
Amount of variance explained by the interaction: R^2_m = .01, R^2_c = .01					
Pinyin revision	Narrative * 2	.16	.07	2.36	.02*
	Narrative * 3	.23	.07	3.35	< .01*
	Narrative * 4	.18	.07	2.58	.01*
	Narrative * 5	.08	.07	1.14	.26
Model = lmer(log(pinyin_revision) ~ genre*stage + (1+genre|participant) + (1|prompt), data, REML = F)					
Amount of variance explained by the interaction: R^2_m = .01, R^2_c = .01					
Character revision	Narrative * 2	.07	.07	1.07	.28
	Narrative * 3	.17	.07	2.60	.01*
	Narrative * 4	.11	.07	1.62	.11
	Narrative * 5	.10	.07	1.50	.14
Model = lmer(log(character_revision) ~ genre*stage + (1+genre|participant) + (1|prompt), data, REML = F)					
Amount of variance explained by the interaction: R^2_m = .01, R^2_c = .01					

Note: Bpc = Between Pinyin and character(s), Bc = Between clauses, Bs = Between sentences, R^2_m = marginal R-squared, R^2_c = conditional R-squared, * = statistically significant results

To further examine interactions, linear mixed-effects regression analyses were performed separately for argumentative and narrative writing. The fixed effect, this time, was the writing stage only. The random effects were participant (and prompt). No by-participant random slope for writing stage was included, as the models failed to converge with a random slope. Log transformations were carried out on the dependent variables when necessary. Post hoc Bonferroni tests were performed when significant effects of stage were found in behaviour indices.

The results for L2 Chinese writing are demonstrated in Tables 4.26 and 4.27. Stage had a greater influence on narrative writing, as the fixed effect, stage, explained a larger amount of variances in the behaviour indices with considerably more significant differences across stages revealed by post hoc Bonferroni tests for this genre. Although an interaction between genre and stage was identified for between-Pinyin-and-character(s) pause duration, post hoc analyses did not reveal an effect of stage on this index for either argumentative or narrative writing.

For argumentative writing, most differences were observed between the last and middle stage(s). L2 writers paused more often between Pinyin and character(s) in Stage 2 than in Stage 5. They also made more stops between clauses, and more changes at the point of inscription and to Pinyin in Stage 3 than in Stage 5. In addition, the pause frequency overall was higher in Stages 2, 3 and 4 than in Stage 5. For narrative writing, the majority of differences were found between the middle, and the initial and final stages. In terms of pausing measures, L2 writers paused more often overall and between clauses in Stages 2, 3 and 4 than in Stage 1. Stages 2 and 3 featured more pauses overall and between Pinyin and character(s) than in Stage 5, while Stages 3 and 4 exhibited more between-clause pauses than Stage 5. In terms of revision measures, L2 writers made fewer revisions overall in Stage 1 than in sequential stages and in Stage 5 than in Stage 3. In terms of the linguistic domain, fewer below-word revisions were found in Stages 1 than in Stages 2, 3 and 4, and in

Stage 5 than in Stages 3 and 4. L2 writers also made fewer below-clause revisions in Stage 1 than in Stages 2 and 3. In addition, they made fewer changes at the point of inscription and to Pinyin in Stages 1 and 5 than in the middle writing stages. The effect sizes of the differences were all small.

Table 4.26 *Results from linear mixed-effects regressions examining the effects of stage on writing behaviours where interactions were significant by genre in L2 Chinese (N = 32)*

	Argumentative					Narrative				
Dependent variable	Fixed effect: stage	*Est*	*SE*	*t*	*p*	Fixed effect: stage	*Est*	*SE*	*t*	*p*
Total pause frequency	2	.43	.15	2.85	< .01*	2	.93	.17	5.59	< .01*
	3	.24	.15	1.59	.11	3	.94	.17	5.62	< .01*
	4	.33	.15	2.20	.03*	4	.57	.17	3.41	< .01*
	5	-.28	.15	-1.88	.06	5	.08	.17	.45	.65
	Model = lmer(total_pause_frequency_arg ~ stage + (1|participant), data, REML = F), R^2_m = .05, R^2_c = .43					Model = lmer(total_pause_frequency_nar ~ stage + (1|participant) + (1|prompt), data, REML = F), R^2_m = .10, R^2_c = .45				
Bpc pause frequency	2	.03	.03	.78	.43	2	.10	.04	2.66	< .01*
	3	-.03	.03	-.78	.44	3	.09	.04	2.44	.02*
	4	-.02	.03	-.58	.56	4	.02	.04	.67	.50
	5	-.10	.03	-3.02	< .01*	5	-.07	.04	-1.81	.07
	Model = lmer(log(bpc_pause_frequency_arg) ~ stage + (1|participant), data, REML = F), R^2_m = .02, R^2_c = .59					Model = lmer(log(bpc_pause_frequency_nar) ~ stage + (1|participant) + (1|prompt), data, REML = F), R^2_m = .04, R^2_c = .51				
Bc pause frequency	2	.08	.05	1.52	.13	2	.14	.04	3.85	< .01*
	3	.12	.05	2.18	.03*	3	.16	.04	4.50	< .01*
	4	.04	.05	.81	.41	4	.15	.04	4.24	< .01*
	5	-.09	.05	-1.71	.09	5	.03	.04	.87	.39
	Model = lmer(bc_pause_frequency_arg ~ stage + (1|participant), data, REML = F), R^2_m = .05, R^2_c = .14					Model = lmer(bc_pause_frequency_nar ~ stage + (1|participant) + (1|prompt), data, REML = F), R^2_m = .08, R^2_c = .28				
Bpc pause duration	2	-.11	.04	-2.38	.02*	2	.08	.04	1.90	.06

	3	-.10	.04	-2.37	.02*	3	.09	.04	2.05	.04*
	4	-.03	.04	-.67	.51	4	.03	.04	.71	.48
	5	< -.01	.04	-.03	.98	5	.09	.04	1.89	.06
	Model = lmer(1/log(bpc_pause_duration_arg) ~ stage + (1|participant), data, REML = F), R^2_m = .04, R^2_c = .06					Model = lmer(1/log(bpc_pause_duration_nar) ~ stage + (1|participant), data, REML = F), R^2_m = .02, R^2_c = .04				
Total revision	2	.11	.04	2.49	.01*	2	.30	.04	7.03	< .01*
	3	.12	.04	2.81	< .01*	3	.33	.04	7.60	< .01*
	4	.13	.04	3.10	< .01*	4	.31	.04	7.16	< .01*
	5	.05	.04	1.28	.20	5	.17	.04	3.83	< .01*
	Model = lmer(log(total_revision_arg) ~ stage + (1|participant), data, REML = F), R^2_m = .01, R^2_c = .69					Model = lmer(log(total_revision_nar) ~ stage + (1|participant), data, REML = F), R^2_m = .08, R^2_c = .71				
Below-word revision	2	.04	.05	.82	.41	2	.20	.05	4.10	< .01*
	3	.04	.05	.78	.44	3	.24	.05	4.86	< .01*
	4	.06	.05	1.34	.18	4	.25	.05	4.99	< .01*
	5	-.05	.05	-1.16	.25	5	.04	.05	.82	.41
	Model = lmer(log(below_word_revision_arg) ~ stage + (1|participant), data, REML = F)					Model = lmer(log(below_word_revision_nar) ~ stage + (1|participant), data, REML = F), R^2_m = .05, R^2_c = .63				
Below-clause revision	2	.06	.04	1.74	.08	2	.15	.04	4.18	< .01*
	3	.06	.04	1.55	.12	3	.17	.04	4.73	< .01*
	4	.08	.04	2.18	.03*	4	.12	.04	3.31	< .01*
	5	.06	.04	1.62	.11	5	.10	.04	2.86	< .01*
	Model = lmer(log(below_clause_revision_arg) ~ stage + (1|participant), data, REML = F), R^2_m = .01, R^2_c = .44					Model = lmer(log(below_clause_revision_nar) ~ stage + (1|participant) + (1|prompt), data, REML = F), R^2_m = .05, R^2_c = .44				
Pre-contextual revision	2	.07	.05	1.48	.14	2	.28	.05	5.17	< .01*
	3	.09	.05	1.93	.05*	3	.29	.05	5.29	< .01*
	4	.06	.05	1.29	.20	4	.22	.05	4.24	< .01*

	5	-.07	.05	-1.46	.14	5	-.03	.05	-.52	.61
	Model = lmer(log(precontextual_revision_arg) ~ stage + (1|participant), data, REML = F), $R^2_m = .02$, $R^2_c = .66$					Model = lmer(log(precontextual_revision_nar) ~ stage + (1|participant) + (1|prompt), data, REML = F), $R^2_m = .08$, $R^2_c = .61$				
Pinyin revision	2	.04	.05	.86	.39	2	.25	.05	4.66	< .01*
	3	.06	.05	1.23	.22	3	.28	.05	5.28	< .01*
	4	.04	.05	.84	.40	4	.22	.05	4.14	< .01*
	5	-.11	.05	-2.30	.02*	5	-.01	.05	-.17	.86
	Model = lmer(log(pinyin_revision_arg) ~ stage + (1|participant), data, REML = F), $R^2_m = .02$, $R^2_c = .64$					Model = lmer(log(pinyin_revision_nar) ~ stage + (1|participant), data, REML = F), $R^2_m = .07$, $R^2_c = .60$				

Note: Bpc = Between Pinyin and character(s), Bc = Between clauses, R^2_m = marginal *R*-squared, R^2_c = conditional *R*-squared, * = statistically significant result

Table 4.27 *Significant results*[5] *from post hoc Bonferroni tests examining the effects of stage on writing behaviours by genre in L2 Chinese (N = 32, p < .01)*

	Argumentative				Narrative			
	Stage	*SE*	*t*	*d*	Stage	*SE*	*t*	*d*
Pause frequency								
Total pause	S5 < S2	.15	-4.69	.56	S1 < S2	.17	-5.55	.60
	S5 < S3	.15	-3.44	.45	S1 < S3	.17	-5.59	.60
	S5 < S4	.15	-4.05	.46	S1 < S4	.17	-3.39	.37
					S5 < S2	.17	-5.11	.73
					S5 < S3	.17	-5.14	.66
Bpc pause	S5 < S2	.04	-3.77	.28	S5 < S2	.04	-4.44	.51
					S5 < S3	.04	-4.21	.52
Bc pause	S5 < S3	.05	-3.87	.49	S1 < S2	.04	-3.83	.47
					S1 < S3	.04	-4.47	.58
					S1 < S4	.04	-4.21	.47
					S5 < S3	.04	-3.61	.40
					S5 < S4	.04	-3.35	.39
Revision								
Total					S1 < S2	.04	-6.98	.75
					S1 < S3	.04	-7.54	.99
					S1 < S4	.04	-7.11	.83
					S1 < S5	.04	-3.80	.38

[5] Refer to Table 7 in Appendix H for non-significant results from post hoc Bonferroni tests of stage effects on writing behaviours by genre in L2 Chinese.

					S5 < S3	.04	-3.74	.49
Below word					S1 < S2	.05	-4.07	.49
					S1 < S3	.05	-4.83	.63
					S1 < S4	.05	-4.95	.63
					S5 < S3	.05	-4.01	.48
					S5 < S4	.05	-4.13	.48
Below clause					S1 < S2	.04	-4.15	.59
					S1 < S3	.04	-4.70	.60
Pre-contextual	S5 < S3	.05	-3.37	.35	S1 < S2	.05	-5.14	.59
					S1 < S3	.05	-5.26	.60
					S1 < S4	.05	-4.21	.59
					S5 < S2	.05	-5.65	.83
					S5 < S3	.05	-5.77	.85
					S5 < S4	.05	-4.72	.48
Pinyin	S5 < S3	.05	-3.51	.37	S1 < S2	.05	-4.63	.60
					S1 < S3	.05	-5.24	.70
					S1 < S4	.05	-4.12	.53
					S5 < S2	.05	-4.80	.58
					S5 < S3	.05	-5.41	.60
					S5 < S4	.05	-4.29	.49

Note: Bpc = Between Pinyin and character(s), Bc = Between clauses

In parallel with the findings for L2 writing, a greater effect of stage was observed on narrative than argumentative writing in L1 Chinese, as illustrated in Tables 4.28 and 4.29. However, the way in which stage affected the behaviours in L1 argumentative and narrative writing seems to be different from that in L2 writing. First, stage had no effect on pause frequency in L1 argumentative writing. On the other hand, the effect of stage was found on one pause duration measure, i.e. L1 writers paused longer between sentences in Stage 1 than in the latter stages and in Stage 5 than in Stage 4 in argumentative writing. In narrative writing, L1 writers only paused (overall and between-clauses) less often in Stage 1 than in Stages 2, 3 and 4, while L2 writers paused less frequently in both the initial and final stages. In addition, L1 narrative writing featured longer pauses (overall and between-sentence) in Stages 1 and 5 than in the middle stages. In terms of revision behaviours, unlike in L2 argumentative writing, stage only influenced the total amount of revision in L1 argumentative writing, i.e. there were fewer revisions overall in Stage 1 than in Stages 2 and 4. Similar to L2 narrative writing, L1 narrative writing featured fewer revisions (overall, below-word, pre-contextual and Pinyin) at the beginning and the end than in the middle of writing. L1 writers additionally revised characters more often after the initial stage when producing narrative essays. Again, the effect sizes of the differences were in the range of small

Table 4.28 *Results from linear mixed-effects regressions examining the effects of stage on writing behaviours where interactions were significant by genre in L1 Chinese (N = 32)*

	Argumentative					Narrative				
Dependent variable	Fixed effect: stage	*Est*	*SE*	*t*	*p*	Fixed effect: stage	*Est*	*SE*	*t*	*p*
Total pause frequency	2	.31	.14	2.17	.03*	2	.77	.16	4.82	< .01*
	3	.27	.14	1.91	.06	3	.74	.16	4.64	< .01*
	4	.23	.14	1.60	.11	4	.54	.16	3.37	< .01*
	5	.07	.14	.50	.62	5	.52	.16	3.25	< .01*
	Model = lmer(total_pause_frequency_arg ~ stage + (1|participant) + (1|prompt), data, REML = F), R^2_m = .01, R^2_c = .38					Model = lmer(total_pause_frequency_nar ~ stage + (1|participant) + (1|prompt), data, REML = F), R^2_m = .07, R^2_c = .30				
Bpc pause frequency	2	.05	.03	1.81	.07	2	.12	.03	3.32	< .01*
	3	.02	.03	.59	.55	3	.05	.03	1.44	.15
	4	< .01	.03	.12	.91	4	.11	.03	3.07	< .01*
	5	-.02	.03	-.73	.47	5	< .01	.03	.08	.94
	Model = lmer(bpc_pause_frequency_arg ~ stage + (1|participant) + (1|prompt), data, REML = F)					Model = lmer(bpc_pause_frequency_nar ~ stage + (1|participant), data, REML = F), R^2_m = .04, R^2_c = .34				
Bc pause frequency	2	.12	.07	1.62	.11	2	.20	.07	2.78	.01*
	3	.07	.07	.90	.37	3	.36	.07	5.03	< .01*
	4	.10	.07	1.36	.17	4	.35	.07	4.87	< .01*
	5	.05	.07	.70	.48	5	.13	.07	1.76	.08
	Model = lmer(bc_pause_frequency_arg ~ stage + (1|participant) + (1|prompt), data, REML = F)					Model = lmer(bc_pause_frequency_nar ~ stage + (1|participant), data, REML = F), R^2_m = .08, R^2_c = .30				
Total pause duration	2	.02	.02	.75	.46	2	.05	.02	2.17	.03*

	3	.01	.02	.62	.54	3	.08	.02	3.53	< .01*
	4	.02	.02	.90	.37	4	.06	.02	2.70	.01*
	5	-.02	.02	-.69	.49	5	-.01	.02	-.35	.73
	Model = lmer(1/log(total_pause_duration_arg) ~ stage + (1|participant), data, REML = F)					Model = lmer(1/log(total_pause_duration_nar) ~ stage + (1|participant), data, REML = F), R^2_m = .05, R^2_c = .25				
Bpc pause duration	2	-.07	.06	-1.02	.30	2	-.08	.06	-1.27	.21
	3	-.19	.06	-2.85	.01*	3	.02	.07	.33	.74
	4	< .01	.06	< .01	1.00	4	-.08	.06	-1.21	.23
	5	-.07	.06	-1.07	.29	5	.05	.07	.71	.48
	Model = lmer(1/log(bpc_pause_duration_arg) ~ stage + (1|participant), data, REML = F), R^2_m = .08, R^2_c = .19					Model = lmer(1/log(bpc_pause_duration_nar) ~ stage + (1|participant), data, REML = F)				
Bc pause duration	2	-.07	.04	-1.72	.09	2	-.16	.08	-2.10	.04*
	3	-.01	.04	-.22	.83	3	-.10	.08	-1.28	.20
	4	< .01	.04	.12	.90	4	-.12	.08	-1.52	.13
	5	-.05	.04	-1.23	.22	5	-.10	.08	-1.35	.18
	Model = lmer(1/log(bc_pause_duration_arg) ~ stage + (1|participant) + (1|prompt), data, REML = F)					Model = lmer(log(bc_pause_duration_nar) ~ stage + (1|participant) + (1|prompt), data, REML = F), R^2_m = .01, R^2_c = .20				
Bs pause duration	2	-.57	.10	-5.51	< .01*	2	-.98	.13	-7.76	< .01*
	3	-.61	.10	-5.97	< .01*	3	-1.26	.13	-9.98	< .01*
	4	-.74	.10	-7.21	< .01*	4	-1.16	.13	-9.22	< .01*
	5	-.39	.10	-3.87	< .01*	5	-.71	.13	-5.73	< .01*
	Model = lmer(log(bs_pause_duration_arg) ~ stage + (1|participant) + (1|prompt),data, REML = F), R^2_m = .15, R^2_c = .28					Model = lmer(log(bs_pause_duration_nar) ~ stage + (1|participant) + (1|prompt), data, REML = F), R^2_m = .26, R^2_c = .38				
Total revision	2	.13	.04	3.55	< .01*	2	1.39	.29	4.80	< .01*
	3	.10	.04	2.72	.01*	3	1.88	.29	6.53	< .01*
	4	.14	.04	3.62	< .01*	4	1.64	.29	5.67	< .01*

	5	.06	.04	1.71	.09	5	.79	.29	2.73	.01*
	Model = lmer(log(total_revision_arg) ~ stage + (1|participant), data, REML = F), R^2_m = .02, R^2_c = .60					Model = lmer(total_revision_nar) ~ stage + (1|participant) + (1|prompt), data, REML = F), R^2_m = .08, R^2_c = .56				
Below-word revision	2	.05	.05	1.14	.25	2	.22	.05	4.25	< .01*
	3	.04	.05	.82	.41	3	.31	.05	5.98	< .01*
	4	.08	.05	1.78	.08	4	.28	.05	5.35	< .01*
	5	.04	.05	.83	.41	5	.09	.05	1.77	.08
	Model = lmer(log(below_word_revision_arg) ~ stage + (1|participant), data, REML = F)					Model = lmer(log(below_word_revision_nar) ~ stage + (1|participant) + (1|prompt), data, REML = F), R^2_m = .09, R^2_c = .44				
Pre-contextual revision	2	.11	.04	2.69	.01*	2	.28	.06	5.02	< .01*
	3	.08	.04	1.78	.08	3	.36	.06	6.50	< .01*
	4	.12	.04	2.83	< .01*	4	.34	.06	6.07	< .01*
	5	< .01	.04	.09	.93	5	.13	.06	2.37	.02*
	Model = lmer(log(precontextual_revision_arg) ~ stage + (1|participant) + (1|prompt), data, REML = F), R^2_m = .02, R^2_c = .58					Model = lmer(log(precontextual_revision_nar) ~ stage + (1|participant) + (1|prompt), data, REML = F), R^2_m = .10, R^2_c = .47				
Pinyin revision	2	.06	.05	1.32	.19	2	.22	.05	4.31	< .01*
	3	.04	.05	.83	.41	3	.27	.05	5.21	< .01*
	4	.08	.05	.77	.08	4	.26	.05	5.01	< .01*
	5	< .01	.05	< .01	1.00	5	.08	.05	1.52	.13
	Model = lmer(log(pinyin_revision_arg) ~ stage + (1|participant), data, REML = F)					Model = lmer(log(pinyin_revision_nar) ~ stage + (1|participant) + (1|prompt), data, REML = F), R^2_m = .07, R^2_c = .51				
Character revision	2	.13	.04	2.96	< .01*	2	.20	.05	3.98	< .01*
	3	.12	.04	2.73	.01*	3	.29	.05	5.81	< .01*
	4	.14	.04	3.09	< .01*	4	.24	.05	4.82	< .01*
	5	.08	.04	1.94	.05	5	.19	.05	3.66	< .01*
	Model = lmer(log(character_revision_arg) ~ stage + (1|participant)					Model = lmer(log(character_revision_nar) ~ stage + (1|participant)				

+ (1|prompt), data, REML = F), R^2_m = .02, R^2_c = .59	+ (1|prompt), data, REML = F), R^2_m = .06, R^2_c = .51

Note: Bpc = Between Pinyin and character(s), Bc = Between clauses, Bs = Between sentences, R^2_m = marginal *R*-squared, R^2_c = conditional *R*-squared, * = statistically significant result

Table 4.29 *Significant results*[6] *from post hoc Bonferroni tests examining the effects of stage on writing behaviours by genre in L1 Chinese (N = 32, p < .01)*

	Argumentative				Narrative			
	Stage	*SE*	*t*	*d*	Stage	*SE*	*t*	*d*
Pause frequency								
Total pause					S1 < S2	.16	-4.78	.54
					S1 < S3	.16	-4.61	.51
					S1 < S4	.16	-3.35	.35
Bc pause					S1 < S3	.01	-5.00	.66
					S1 < S4	.01	-4.84	.57
Pause duration								
Total pause					S3 < S1	.02	-3.51	.24
					S3 < S5	.02	-3.87	.46
Bs pause	S2 < S1	.10	-5.47	.56	S2 < S1	.13	-7.71	.59
	S3 < S1	.10	-5.93	.66	S3 < S1	.13	-9.91	.65
	S4 < S1	.10	-7.16	.69	S4 < S1	.13	-9.16	.64
	S5 < S1	.10	-3.84	.44	S5 < S1	.13	-5.69	.55
	S4 < S5	.10	-3.36	.49	S3 < S5	.13	-4.32	.67
Revision								
Total	S1 < S2	.04	-3.52	.44	S1 < S2	.29	-4.77	.63
	S1 < S4	.04	-3.59	.48	S1 < S3	.29	-6.48	.75
					S1 < S4	.29	-5.64	.60

[6] Refer to Table 8 in Appendix H for non-significant results from post hoc Bonferroni tests of stage effects on writing behaviours by genre in L1 Chinese.

	S5 < S3	.29	-3.77	.80
Below word	S1 < S2	.05	-4.22	.55
	S1 < S3	.05	-5.94	.75
	S1 < S4	.05	-5.31	.56
	S5 < S4	.05	-4.18	.61
	S5 < S3	.05	-3.56	.53
Pre-contextual	S1 < S2	.06	-4.99	.62
	S1 < S3	.06	-6.45	.70
	S1 < S4	.06	-6.03	.62
	S5 < S4	.06	-4.10	.70
	S5 < S3	.06	-3.68	.61
Pinyin	S1 < S2	.05	-4.28	.55
	S1 < S3	.05	-5.18	.64
	S1 < S4	.05	-4.98	.53
	S5 < S4	.05	-3.66	.51
	S5 < S3	.05	-3.46	.52
Character	S1 < S2	.05	-4.28	.50
	S1 < S3	.05	-5.18	.61
	S1 < S4	.05	-4.98	.50
	S1 < S5	.05	-3.66	.37

Note: Bc = Between clauses, Bs = Between sentences

Next, I looked at the extent to which genre affected cognitive activities that are associated with writing behaviours in L1 and L2 Chinese. The stimulated recall comments on pausing behaviours in argumentative and narrative writing are displayed in Tables 4.30–4.33. As shown in Tables 4.30 and 4.31, L2 writers reported proportionally fewer planning and monitoring activities but more translation activities in argumentative than narrative writing for the whole session. In L1 writing, however, the proportions of comments on planning, translation and monitoring were similar between argumentative and narrative writing (Tables 4.32 and 4.33).

Considering the five stages, similar distributions of cognitive activities were observed between genres in L2 writing. Specifically, translation-related comments accounted for the greatest share in each stage. A considerably larger percentage of planning-related comments was observed in Stage 1 than in subsequent stages, while Stage 5 had proportionally more monitoring-related comments than previous stages. Regardless of genre, pauses at the clause boundary and below mainly reflected translation processes, whereas those at the sentence boundary were more likely to be associated with planning processes. Most pauses between contextual revisions were related to planning in Stage 1 for both genres. However, the majority of pauses between contextual revisions were linked to monitoring in Stage 5 for argumentative writing, but in Stages 4 and 5 for narrative writing.

With regard to L1 writing, the largest percentage of comments was on planning processes in almost every stage for both genres. In argumentative writing, planning-related comments accounted for around 45% from Stage 1 to Stage 4, but dropped to 39% in the last stage. In narrative writing, L1 writers mentioned planning processes with a slightly higher percentage (around 50%) than that in argumentative writing from Stage 1 to Stage 3, but the percentage decreased considerably in the last two stages. Translation processes took up about 40% in each stage in argumentative writing, whereas in narrative writing, the percentage fluctuated across stages. Most monitoring activities were found in Stage 5, regardless of genre. For argumentative and narrative writing, pauses at the word boundary and below tended to be associated with translation, while pauses at the clause boundary and above with planning. Most pauses between contextual revisions were linked to planning in Stage 1 and to monitoring in Stage 5.

Table 4.30 *Pause-associated cognitive activities by stage and location in L2 Chinese argumentative writing (N = 16)*

		Planning				Translation						Monitoring		Others		Not recall		Total	
Stage	Location	Content	Organisation	Total (%)		Translation in general	Lexical retrieval	Syntactic encoding	Cohesion	Total (%)		Total (%)		Total (%)		Total (%)		Total (%)	
1	Ww	1	0	1	(0%)	0	1	0	0	1	(0%)	0	(0%)	0	(0%)	0	(0%)	2	(1%)
1	Bpc	0	0	0	(0%)	0	4	0	0	4	(1%)	0	(0%)	0	(0%)	1	(0%)	5	(1%)
1	Bw	23	4	27	(8%)	13	73	35	8	129	(38%)	9	(3%)	2	(1%)	21	(6%)	188	(56%)
1	Bc	13	1	14	(4%)	2	3	9	5	19	(6%)	2	(1%)	0	(0%)	10	(3%)	45	(13%)
1	Bs	35	3	38	(11%)	2	4	5	2	13	(4%)	5	(1%)	0	(0%)	7	(2%)	63	(19%)
1	Bcr	17	5	22	(7%)	2	0	0	0	2	(1%)	2	(1%)	9	(3%)	0	(0%)	35	(10%)
1	Total	89	13	102	(30%)	19	85	49	15	168	(50%)	18	(5%)	11	(3%)	39	(12%)	338	(100%)
2	Ww	0	0	0	(0%)	0	0	0	0	0	(0%)	0	(0%)	0	(0%)	0	(0%)	0	(0%)
2	Bpc	0	0	0	(0%)	0	6	0	0	6	(2%)	0	(0%)	0	(0%)	0	(0%)	6	(2%)
2	Bw	23	0	23	(8%)	16	74	26	5	121	(44%)	7	(3%)	2	(1%)	12	(4%)	165	(60%)
2	Bc	11	4	15	(5%)	1	3	3	4	11	(4%)	2	(1%)	0	(0%)	3	(1%)	31	(11%)
2	Bs	28	2	30	(11%)	6	6	3	1	16	(6%)	14	(5%)	1	(0%)	6	(2%)	67	(25%)
2	Bcr	0	0	0	(0%)	0	1	0	0	1	(0%)	3	(1%)	0	(0%)	0	(0%)	4	(1%)
2	Total	62	6	68	(25%)	23	90	32	10	155	(57%)	26	(10%)	3	(1%)	21	(8%)	273	(100%)
3	Ww	0	0	0	(0%)	0	3	0	1	4	(2%)	0	(0%)	0	(0%)	0	(0%)	4	(2%)
3	Bpc	0	0	0	(0%)	0	4	0	0	4	(2%)	0	(0%)	0	(0%)	0	(0%)	4	(2%)

3	Bw	26	1	27	(10%)	21	65	12	5	103	(40%)	10	(4%)	1	(0%)	18	(7%)	159	(61%)
3	Bc	7	0	7	(3%)	7	2	1	5	15	(6%)	2	(1%)	0	(0%)	4	(2%)	28	(11%)
3	Bs	23	4	27	(10%)	5	4	5	0	14	(5%)	10	(4%)	0	(0%)	10	(4%)	61	(23%)
3	Bcr	0	1	1	(0%)	1	0	0	0	1	(0%)	1	(0%)	0	(0%)	1	(0%)	4	(2%)
3	Total	56	6	62	(24%)	34	78	18	11	141	(54%)	23	(9%)	1	(0%)	33	(13%)	260	(100%)
4	Ww	0	0	0	(0%)	0	1	0	0	1	(0%)	0	(0%)	0	(0%)	0	(0%)	1	(0%)
4	Bpc	0	0	0	(0%)	0	4	0	0	4	(2%)	0	(0%)	0	(0%)	0	(0%)	4	(2%)
4	Bw	21	0	21	(9%)	24	77	14	3	118	(48%)	5	(2%)	1	(0%)	15	(6%)	160	(65%)
4	Bc	11	0	11	(4%)	4	9	4	5	22	(9%)	1	(0%)	0	(0%)	5	(2%)	39	(16%)
4	Bs	15	3	18	(7%)	3	2	1	0	6	(2%)	9	(4%)	0	(0%)	7	(3%)	40	(16%)
4	Bcr	1	0	1	(0%)	0	0	0	0	0	(0%)	0	(0%)	0	(0%)	0	(0%)	1	(0%)
4	Total	48	3	51	(21%)	31	93	19	8	151	(62%)	15	(6%)	1	(0%)	27	(11%)	245	(100%)
5	Ww	0	0	0	(0%)	0	2	0	0	2	(1%)	0	(0%)	0	(0%)	0	(0%)	2	(1%)
5	Bpc	0	0	0	(0%)	0	2	0	0	2	(1%)	0	(0%)	0	(0%)	0	(0%)	2	(1%)
5	Bw	20	0	20	(10%)	15	45	11	2	73	(35%)	5	(2%)	0	(0%)	6	(3%)	104	(50%)
5	Bc	11	1	12	(6%)	4	3	1	0	8	(4%)	1	(0%)	0	(0%)	3	(1%)	24	(12%)
5	Bs	15	3	18	(9%)	5	0	1	4	10	(5%)	16	(8%)	1	(0%)	2	(1%)	47	(23%)
5	Bcr	0	1	1	(0%)	0	0	0	0	0	(0%)	26	(13%)	0	(0%)	0	(0%)	27	(13%)
5	Total	46	5	51	(25%)	24	52	13	6	95	(46%)	48	(23%)	1	(0%)	11	(5%)	206	(100%)
Total	Ww	1	0	1	(0%)	0	7	0	1	8	(1%)	0	(0%)	0	(0%)	0	(0%)	9	(1%)
Total	Bpc	0	0	0	(0%)	0	20	0	0	20	(2%)	0	(0%)	0	(0%)	1	(0%)	21	(2%)
Total	Bw	113	5	118	(9%)	89	334	98	23	544	(41%)	36	(3%)	6	(0%)	72	(5%)	776	(59%)
Total	Bc	53	6	59	(4%)	18	20	18	19	75	(6%)	8	(1%)	0	(0%)	25	(2%)	167	(13%)
Total	Bs	116	15	131	(10%)	21	16	15	7	59	(4%)	54	(4%)	2	(0%)	32	(2%)	278	(21%)
Total	Bcr	18	7	25	(2%)	3	1	0	0	4	(0%)	32	(2%)	9	(1%)	1	(0%)	71	(5%)

Total	Total	301	33	334	(25%)	131	398	131	50	710	(54%)	130	(10%)	17	(1%)	131	(10%)	1322	(100%)

Note: Ww = Within a word, Bpc = Between Pinyin and character(s), Bw = Between words, Bc = Between clauses, Bs = Between sentences, Bcr = Between contextual revisions

Table 4.31 *Pause-associated cognitive activities by stage and location in L2 Chinese narrative writing (N = 16)*

		Planning				Translation						Monitoring		Others		Not recall		Total	
Stage	Location	Content	Organisation	Total (%)		Translation in general	Lexical retrieval	Syntactic encoding	Cohesion	Total (%)		Total (%)		Total (%)		Total (%)		Total (%)	
1	Ww	0	0	0	(0%)	0	1	0	0	1	(0%)	0	(0%)	0	(0%)	0	(0%)	1	(0%)
1	Bpc	0	0	0	(0%)	0	3	0	0	3	(1%)	0	(0%)	0	(0%)	0	(0%)	3	(1%)
1	Bw	43	1	44	(15%)	11	59	27	4	101	(35%)	8	(3%)	0	(0%)	16	(5%)	169	(58%)
1	Bc	11	1	12	(4%)	2	5	7	0	14	(5%)	1	(0%)	0	(0%)	2	(1%)	29	(10%)
1	Bs	28	2	30	(10%)	1	4	8	2	15	(5%)	4	(1%)	1	(0%)	0	(0%)	50	(17%)
1	Bcr	17	7	24	(8%)	3	1	1	2	7	(2%)	0	(0%)	8	(3%)	0	(0%)	39	(13%)
1	Total	99	11	110	(38%)	17	73	43	8	141	(48%)	13	(4%)	9	(3%)	18	(6%)	291	(100%)
2	Ww	0	0	0	(0%)	0	7	0	0	7	(2%)	0	(0%)	0	(0%)	0	(0%)	7	(2%)
2	Bpc	0	0	0	(0%)	0	3	0	0	3	(1%)	0	(0%)	0	(0%)	0	(0%)	3	(1%)
2	Bw	31	0	31	(10%)	18	84	31	4	137	(45%)	12	(4%)	4	(1%)	6	(2%)	190	(62%)
2	Bc	13	2	15	(5%)	4	14	3	3	24	(8%)	5	(2%)	1	(0%)	3	(1%)	48	(16%)
2	Bs	28	3	31	(10%)	6	2	4	1	13	(4%)	10	(3%)	0	(0%)	2	(1%)	56	(18%)
2	Bcr	2	0	2	(1%)	0	0	0	0	0	(0%)	1	(0%)	0	(0%)	0	(0%)	3	(1%)
2	Total	74	5	79	(26%)	28	110	38	8	184	(60%)	28	(9%)	5	(2%)	11	(4%)	307	(100%)
3	Ww	0	0	0	(0%)	0	2	0	0	2	(1%)	0	(0%)	0	(0%)	0	(0%)	2	(1%)
3	Bpc	0	0	0	(0%)	0	2	0	0	2	(1%)	0	(0%)	0	(0%)	0	(0%)	2	(1%)

3	Bw	30	3	33	(12%)	19	63	21	3	106	(39%)	9	(3%)	3	(1%)	12	(4%)	163	(60%)
3	Bc	15	0	15	(5%)	5	7	4	0	16	(6%)	7	(3%)	0	(0%)	3	(1%)	41	(15%)
3	Bs	26	0	26	(10%)	6	5	1	2	14	(5%)	15	(5%)	1	(0%)	6	(2%)	62	(23%)
3	Bcr	1	0	1	(0%)	0	0	0	0	0	(0%)	2	(1%)	0	(0%)	0	(0%)	3	(1%)
3	Total	72	3	75	(27%)	30	79	26	5	140	(51%)	33	(12%)	4	(1%)	21	(8%)	273	(100%)
4	Ww	0	0	0	(0%)	0	0	0	0	0	(0%)	0	(0%)	0	(0%)	0	(0%)	0	(0%)
4	Bpc	0	0	0	(0%)	0	0	0	0	0	(0%)	0	(0%)	0	(0%)	0	(0%)	0	(0%)
4	Bw	30	0	30	(12%)	10	58	21	3	92	(38%)	9	(4%)	1	(0%)	4	(2%)	136	(56%)
4	Bc	12	0	12	(5%)	4	5	4	0	13	(5%)	5	(2%)	0	(0%)	4	(2%)	34	(14%)
4	Bs	28	1	29	(12%)	1	2	1	0	4	(2%)	11	(5%)	0	(0%)	4	(2%)	48	(20%)
4	Bcr	0	0	0	(0%)	1	1	1	0	3	(1%)	21	(9%)	0	(0%)	0	(0%)	24	(10%)
4	Total	70	1	71	(29%)	16	66	27	3	112	(46%)	46	(19%)	1	(0%)	12	(5%)	242	(100%)
5	Ww	0	0	0	(0%)	0	1	0	0	1	(0%)	0	(0%)	0	(0%)	0	(0%)	1	(0%)
5	Bpc	0	0	0	(0%)	0	1	0	0	1	(0%)	0	(0%)	0	(0%)	0	(0%)	1	(0%)
5	Bw	28	0	28	(12%)	9	48	16	5	78	(34%)	6	(3%)	1	(0%)	5	(2%)	118	(51%)
5	Bc	8	0	8	(3%)	2	6	1	0	9	(4%)	2	(1%)	0	(0%)	2	(1%)	21	(9%)
5	Bs	15	1	16	(7%)	1	2	0	0	3	(1%)	20	(9%)	1	(0%)	0	(0%)	40	(17%)
5	Bcr	3	0	3	(1%)	0	0	0	0	0	(0%)	47	(20%)	0	(0%)	0	(0%)	50	(22%)
5	Total	54	1	55	(24%)	12	58	17	5	92	(40%)	75	(32%)	2	(1%)	7	(3%)	231	(100%)
Total	Ww	0	0	0	(0%)	0	11	0	0	11	(1%)	0	(0%)	0	(0%)	0	(0%)	11	(1%)
Total	Bpc	0	0	0	(0%)	0	9	0	0	9	(1%)	0	(0%)	0	(0%)	0	(0%)	9	(1%)
Total	Bw	162	4	166	(12%)	67	312	116	19	514	(38%)	44	(3%)	9	(1%)	43	(3%)	776	(58%)
Total	Bc	59	3	62	(5%)	17	37	19	3	76	(6%)	20	(1%)	1	(0%)	14	(1%)	173	(13%)
Total	Bs	125	7	132	(10%)	15	15	14	5	49	(4%)	60	(4%)	3	(0%)	12	(1%)	256	(19%)
Total	Bcr	23	7	30	(2%)	4	2	2	2	10	(1%)	71	(5%)	8	(1%)	0	(0%)	119	(9%)

Total	Total	369	21	390	(29%)	103	386	151	29	669	(50%)	195	(15%)	21	(2%)	69	(5%)	1344	(100%)

Note: Ww = Within a word, Bpc = Between Pinyin and character(s), Bw = Between words, Bc = Between clauses, Bs = Between sentences, Bcr = Between contextual revisions

Table 4.32 *Pause-associated cognitive activities by stage and location in L1 Chinese argumentative writing (N = 16)*

		Planning			Translation					Monitoring	Others	No recall	Total
Stage	Location	Content	Organisation	Total (%)	Translation in general	Lexical retrieval	Syntactic encoding	Cohesion	Total (%)	Total (%)	Total (%)	Total (%)	Total (%)
1	Ww	0	0	0 (0%)	0	0	0	0	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
1	Bpc	0	0	0 (0%)	0	2	0	0	2 (1%)	0 (0%)	0 (0%)	1 (0%)	3 (1%)
1	Bw	26	2	28 (13%)	24	31	3	1	59 (27%)	5 (2%)	1 (0%)	14 (6%)	107 (50%)
1	Bc	16	2	18 (8%)	11	0	0	1	12 (6%)	2 (1%)	0 (0%)	1 (0%)	33 (15%)
1	Bs	24	6	30 (14%)	3	0	2	3	8 (4%)	5 (2%)	0 (0%)	0 (0%)	43 (20%)
1	Bcr	13	4	17 (8%)	2	0	0	0	2 (1%)	1 (0%)	10 (5%)	0 (0%)	30 (14%)
1	Total	79	14	93 (43%)	40	33	5	5	83 (38%)	13 (6%)	11 (5%)	16 (7%)	216 (100%)
2	Ww	0	0	0 (0%)	0	0	0	0	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
2	Bpc	0	0	0 (0%)	0	0	0	0	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
2	Bw	42	1	43 (19%)	18	46	10	2	76 (34%)	4 (2%)	0 (0%)	8 (4%)	131 (58%)
2	Bc	20	0	20 (9%)	7	0	0	1	8 (4%)	4 (2%)	0 (0%)	2 (1%)	34 (15%)
2	Bs	31	8	39 (17%)	2	4	0	0	6 (3%)	11 (5%)	0 (0%)	3 (1%)	59 (26%)
2	Bcr	0	0	0 (0%)	0	1	0	0	1 (0%)	0 (0%)	0 (0%)	0 (0%)	1 (0%)
2	Total	93	9	102 (45%)	27	51	10	3	91 (40%)	19 (8%)	0 (0%)	13 (6%)	225 (100%)
3	Ww	0	0	0 (0%)	0	0	0	0	0 (0%)	1 (0%)	0 (0%)	0 (0%)	1 (0%)
3	Bpc	0	0	0 (0%)	0	2	0	0	2 (1%)	0 (0%)	0 (0%)	0 (0%)	2 (1%)

3	Bw	40	1	41 (18%)	24	44	4	0	72 (32%)	5 (2%)	1 (0%)	7 (3%)	126 (56%)
3	Bc	23	1	24 (11%)	4	1	0	3	8 (4%)	1 (0%)	0 (0%)	3 (1%)	36 (16%)
3	Bs	32	5	37 (17%)	4	0	1	3	8 (4%)	10 (4%)	1 (0%)	0 (0%)	56 (25%)
3	Bcr	1	0	1 (0%)	0	0	0	0	0 (0%)	2 (1%)	0 (0%)	0 (0%)	3 (1%)
3	Total	96	7	103 (46%)	32	47	5	6	90 (40%)	19 (8%)	2 (1%)	10 (4%)	224 (100%)
4	Ww	0	0	0 (0%)	1	0	0	0	1 (0%)	0 (0%)	0 (0%)	0 (0%)	1 (0%)
4	Bpc	0	0	0 (0%)	0	1	0	0	1 (0%)	0 (0%)	0 (0%)	0 (0%)	1 (0%)
4	Bw	40	1	41 (18%)	25	36	6	3	70 (31%)	4 (2%)	1 (0%)	9 (4%)	125 (56%)
4	Bc	17	0	17 (8%)	11	1	0	2	14 (6%)	5 (2%)	0 (0%)	2 (1%)	38 (17%)
4	Bs	37	2	39 (17%)	3	0	0	2	5 (2%)	6 (3%)	0 (0%)	1 (0%)	54 (24%)
4	Bcr	3	0	3 (1%)	0	0	0	0	0 (0%)	1 (0%)	0 (0%)	0 (0%)	4 (2%)
4	Total	97	3	100 (45%)	40	38	6	7	91 (41%)	16 (7%)	1 (0%)	15 (7%)	223 (100%)
5	Ww	0	0	0 (0%)	0	1	0	0	1 (0%)	0 (0%)	0 (0%)	0 (0%)	1 (0%)
5	Bpc	0	0	0 (0%)	0	1	0	0	1 (0%)	0 (0%)	0 (0%)	0 (0%)	1 (0%)
5	Bw	34	0	34 (17%)	22	35	1	1	59 (29%)	4 (2%)	1 (0%)	8 (4%)	106 (52%)
5	Bc	22	0	22 (11%)	8	2	0	0	10 (5%)	3 (1%)	0 (0%)	3 (1%)	38 (19%)
5	Bs	18	4	22 (11%)	1	2	0	2	5 (2%)	18 (9%)	0 (0%)	3 (1%)	48 (24%)
5	Bcr	1	0	1 (0%)	0	0	0	0	0 (0%)	9 (4%)	0 (0%)	0 (0%)	10 (5%)
5	Total	75	4	79 (39%)	31	41	1	3	76 (37%)	34 (17%)	1 (0%)	14 (7%)	204 (100%)
Total	Ww	0	0	0 (0%)	1	1	0	0	2 (0%)	1 (0%)	0 (0%)	0 (0%)	3 (0%)
Total	Bpc	0	0	0 (0%)	0	6	0	0	6 (1%)	0 (0%)	0 (0%)	1 (0%)	7 (1%)
Total	Bw	182	5	187 (17%)	113	192	24	7	336 (31%)	22 (2%)	4 (0%)	46 (4%)	595 (54%)
Total	Bc	98	3	101 (9%)	41	4	0	7	52 (5%)	15 (1%)	0 (0%)	11 (1%)	179 (16%)
Total	Bs	142	25	167 (15%)	13	6	3	10	32 (3%)	50 (5%)	1 (0%)	10 (1%)	260 (24%)
Total	Bcr	18	4	22 (2%)	2	1	0	0	3 (0%)	13 (1%)	10 (1%)	0 (0%)	48 (4%)

Total	Total	440	37	477 (44%)	170	210	27	24	431 (39%)	101 (9%)	15 (1%)	68 (6%)	1092 (100%)

Note: Ww = Within a word, Bpc = Between Pinyin and character(s), Bw = Between words, Bc = Between clauses, Bs = Between sentences, Bcr = Between contextual revisions

Table 4.33 *Pause-associated cognitive activities by stage and location in L1 Chinese narrative writing (N = 16)*

		Planning			Translation					Monitoring	Others	No recall	Total
Stage	Location	Content	Organisation	Total (%)	Translation in general	Lexical retrieval	Syntactic encoding	Cohesion	Total (%)	Total (%)	Total (%)	Total (%)	Total (%)
1	Ww	0	0	0 (0%)	0	0	0	0	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
1	Bpc	0	0	0 (0%)	0	0	0	0	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
1	Bw	34	0	34 (21%)	12	16	8	5	41 (25%)	6 (4%)	0 (0%)	2 (1%)	83 (51%)
1	Bc	11	2	13 (8%)	2	1	0	3	6 (4%)	2 (1%)	0 (0%)	3 (2%)	24 (15%)
1	Bs	12	0	12 (7%)	4	1	0	2	7 (4%)	3 (2%)	0 (0%)	5 (3%)	27 (16%)
1	Bcr	18	3	21 (13%)	0	0	0	0	0 (0%)	1 (1%)	8 (5%)	0 (0%)	30 (18%)
1	Total	75	5	80 (49%)	18	18	8	10	54 (33%)	12 (7%)	8 (5%)	10 (6%)	164 (100%)
2	Ww	0	0	0 (0%)	0	0	0	0	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
2	Bpc	0	0	0 (0%)	0	4	0	0	4 (2%)	0 (0%)	0 (0%)	0 (0%)	4 (2%)
2	Bw	54	2	56 (26%)	15	40	7	1	63 (29%)	1 (0%)	1 (0%)	7 (3%)	128 (60%)
2	Bc	26	0	26 (12%)	4	3	5	0	12 (6%)	3 (1%)	1 (0%)	5 (2%)	47 (22%)
2	Bs	23	1	24 (11%)	3	0	0	1	4 (2%)	5 (2%)	0 (0%)	0 (0%)	33 (15%)
2	Bcr	1	0	1 (0%)	0	0	0	1	1 (0%)	0 (0%)	0 (0%)	0 (0%)	2 (1%)
2	Total	104	3	107 (50%)	22	47	12	3	84 (39%)	9 (4%)	2 (1%)	12 (6%)	214 (100%)
3	Ww	0	0	0 (0%)	0	0	0	0	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
3	Bpc	0	0	0 (0%)	0	3	0	0	3 (1%)	0 (0%)	0 (0%)	0 (0%)	3 (1%)

3	Bw	49	0	49 (23%)	15	25	6	1	47 (22%)	2 (1%)	0 (0%)	5 (2%)	103 (49%)
3	Bc	31	1	32 (15%)	7	5	0	2	14 (7%)	6 (3%)	0 (0%)	7 (3%)	59 (28%)
3	Bs	26	0	26 (12%）	3	1	1	1	6 (3%)	12 (6%)	0 (0%)	1 (0%)	45 (21%)
3	Bcr	0	0	1 (0%)	0	0	0	0	0 (0%)	1 (0%)	0 (0%)	0 (0%)	1 (0%)
3	Total	106	1	107 (51%)	25	34	7	4	70 (33%)	21 (10%)	0 (0%)	13 (6%)	211 (100%)
4	Ww	0	0	0 (0%)	0	0	0	0	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
4	Bpc	2	0	2 (1%)	0	1	0	0	1 (0%)	0 (0%)	0 (0%)	0 (0%)	3 (1%)
4	Bw	24	0	24 (12%)	27	38	9	3	77 (38%)	4 (2%)	0 (0%)	5 (2%)	110 (54%)
4	Bc	34	1	35 (17%)	10	5	1	0	16 (8%)	1 (0%)	1 (0%)	2 (1%)	55 (27%)
4	Bs	23	0	23 (11%)	2	0	0	1	3 (1%)	7 (3%)	1 (0%)	1 (0%)	35 (17%)
4	Bcr	0	0	0 (0%)	0	1	0	0	1 (0%)	0 (0%)	0 (0%)	0 (0%)	1 (0%)
4	Total	83	1	84 (41%)	39	45	10	4	98 (48%)	12 (6%)	2 (1%)	8 (4%)	204 (100%)
5	Ww	0	0	0 (0%)	0	0	0	0	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
5	Bpc	0	0	0 (0%)	0	0	0	0	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
5	Bw	27	1	28 (15%)	19	33	11	0	63 (33%)	7 (4%)	2 (1%)	2 (1%)	102 (53%)
5	Bc	17	0	17 (9%)	2	1	3	0	6 (3%)	9 (5%)	0 (0%)	1 (1%)	33 (17%)
5	Bs	22	2	24 (12%)	1	1	2	0	4 (2%)	18 (9%)	1 (1%)	1 (1%)	48 (25%)
5	Bcr	1	0	1 (1%)	0	1	0	0	1 (1%)	8 (4%)	0 (0%)	0 (0%)	10 (5%)
5	Total	67	3	70 (36%)	22	36	16	0	74 (38%)	42 (22%)	3 (2%)	4 (2%)	193 (100%)
Total	Ww	0	0	0 (0%)	0	0	0	0	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Total	Bpc	2	0	2 (0%)	0	8	0	0	8 (1%)	0 (0%)	0 (0%)	0 (0%)	10 (1%)
Total	Bw	188	3	191 (19%)	88	152	41	10	291 (30%)	20 (2%)	3 (0%)	21 (2%)	526 (53%)
Total	Bc	119	4	123 (12%)	25	15	9	5	54 (5%)	21 (2%)	2 (0%)	18 (2%)	218 (22%)
Total	Bs	106	3	109 (11%)	13	3	3	5	24 (2%)	45 (5%)	2 (0%)	8 (1%)	188 (19%)
Total	Bcr	20	3	23 (2%)	0	2	0	1	3 (0%)	10 (1%)	8 (1%)	0 (0%)	44 (4%)

Total	Total	435	13	448 (45%)	126	180	53	21	380 (39%)	96 (10%)	15 (2%)	47 (5%)	986 (100%)

Note: Ww = Within a word, Bpc = Between Pinyin and character(s), Bw = Between words, Bc = Between clauses, Bs = Between sentences, Bcr = Between contextual revisions

Tables 4.34–4.41 describe the reasons for revisions in Chinese argumentative and narrative writing. For the whole session, L2 writers made more references to planning processes but fewer to translation processes in argumentative than narrative writing (Tables 4.34 and 4.37), while similar percentages of references were made to planning and translation processes in L1 argumentative and narrative writing (Tables 4.38 and 4.41).

In terms of the temporal distribution of cognitive activities associated with revisions, L2 writers referred more to planning in Stages 2, 3 and 4 when writing argumentative essays, whereas the middle stages in narrative writing featured proportionally fewer planning activities. Proportionately more comments on translation were found in Stages 1 and 5 in argumentative writing, but in Stages 2, 3 and 5 in narrative writing. Revisions in different linguistic domains were mainly linked to translation processes for both genres. The only exceptions are that L2 writers mentioned more planning processes when making revisions at the clause level and above in Stages 3 and 4 in argumentative writing, and in Stages 4 and 5 in narrative writing. Both pre-contextual and contextual revisions were primarily translation-related in the majority of stages, irrespective of genre.

Conversely, similar distribution patterns across stages were found in the cognitive activities in which L1 writers engaged when revising argumentative and narrative essays. In general, L1 writers referred more frequently to planning processes in Stages 2, 3 and 4 than in Stages 1 and 5, whereas they made more references to translation processes at the beginning and the end. Regardless of genre, revisions below the clause level were mainly linked to translation, while most revisions at the clause level and above were planning-related. The majority of pre-contextual and contextual revisions described translation processes.

Table 4.34 *Revision-associated cognitive activities by stage and linguistic domain in L2 Chinese argumentative writing (N = 16)*

		Planning				Translation						Others		Not recall		Total	
Stage	Linguistic domain	Content	Organisation	Total (%)		Translation in general	Lexical retrieval	Syntactic encoding	Cohesion	Total (%)		Total (%)		Total (%)		Total (%)	
1	Below word	0	0	0	(0%)	0	5	0	0	5	(5%)	0	(0%)	0	(0%)	5	(5%)
1	Word	4	0	4	(4%)	4	12	6	2	24	(26%)	0	(0%)	4	(4%)	32	(34%)
1	Below clause	9	0	9	(10%)	4	6	9	2	21	(22%)	0	(0%)	4	(4%)	34	(36%)
1	Clause and above	2	6	8	(9%)	0	3	8	3	14	(15%)	0	(0%)	1	(1%)	23	(24%)
1	Total	15	6	21	(22%)	8	26	23	7	64	(68%)	0	(0%)	9	(10%)	94	(100%)
2	Below word	0	0	0	(0%)	0	7	1	0	8	(7%)	0	(0%)	0	(0%)	8	(7%)
2	Word	11	0	11	(10%)	3	13	8	2	26	(23%)	0	(0%)	8	(7%)	45	(40%)
2	Below clause	9	2	11	(10%)	6	9	7	1	23	(20%)	0	(0%)	3	(3%)	37	(33%)
2	Clause and above	5	4	9	(8%)	2	1	6	2	11	(10%)	0	(0%)	3	(3%)	23	(20%)
2	Total	25	6	31	(27%)	11	30	22	5	68	(60%)	0	(0%)	14	(12%)	113	(100%)
3	Below word	0	0	0	(0%)	0	7	0	1	8	(8%)	0	(0%)	3	(3%)	11	(10%)
3	Word	4	3	7	(7%)	5	14	10	2	31	(29%)	0	(0%)	5	(5%)	43	(41%)
3	Below clause	8	1	9	(8%)	4	3	8	3	18	(17%)	1	(1%)	5	(5%)	33	(31%)
3	Clause and above	10	2	12	(11%)	3	0	4	0	7	(7%)	0	(0%)	0	(0%)	19	(18%)
3	Total	22	6	28	(26%)	12	24	22	6	64	(60%)	1	(1%)	13	(12%)	106	(100%)
4	Below word	0	0	0	(0%)	0	3	1	0	4	(5%)	0	(0%)	0	(0%)	4	(5%)

4	Word	7	0	7	(8%)	4	13	3	1	21	(24%)	0	(0%)	5	(6%)	33	(38%)
4	Below clause	7	0	7	(8%)	3	9	7	1	20	(23%)	0	(0%)	5	(6%)	32	(37%)
4	Clause and above	12	1	13	(15%)	0	1	1	1	3	(3%)	0	(0%)	1	(1%)	17	(20%)
4	Total	26	1	27	(31%)	7	26	12	3	48	(56%)	0	(0%)	11	(13%)	86	(100%)
5	Below word	0	0	0	(0%)	0	6	1	1	8	(8%)	0	(0%)	0	(0%)	8	(8%)
5	Word	2	0	2	(2%)	1	7	12	7	27	(25%)	0	(0%)	7	(7%)	36	(34%)
5	Below clause	7	2	9	(8%)	2	5	14	2	23	(22%)	0	(0%)	1	(1%)	33	(31%)
5	Clause and above	9	4	13	(12%)	4	1	8	2	15	(14%)	0	(0%)	1	(1%)	29	(27%)
5	Total	18	6	24	(23%)	7	19	35	12	73	(69%)	0	(0%)	9	(8%)	106	(100%)
Total	Below word	0	0	0	(0%)	0	28	3	2	33	(7%)	0	(0%)	3	(1%)	36	(7%)
Total	Word	28	3	31	(6%)	17	59	39	14	129	(26%)	0	(0%)	29	(6%)	189	(37%)
Total	Below clause	40	5	45	(9%)	19	32	45	9	105	(21%)	1	(0%)	18	(4%)	169	(33%)
Total	Clause and above	38	17	55	(11%)	9	6	27	8	50	(10%)	0	(0%)	6	(1%)	111	(22%)
Total	Total	106	25	131	(26%)	45	125	114	33	317	(63%)	1	(0%)	56	(11%)	505	(100%)

Table 4.35 *Revision-associated cognitive activities by stage and linguistic domain in L2 Chinese narrative writing (N = 16)*

		Planning			Translation					Others	Not recall	Total
Stage	Linguistic domain	Content	Organisation	Total (%)	Translation in general	Lexical retrieval	Syntactic encoding	Cohesion	Total (%)	Total (%)	Total (%)	Total (%)
1	Below word	2	0	2 (3%)	0	3	0	0	3 (4%)	0 (0%)	0 (0%)	5 (7%)
1	Word	4	0	4 (5%)	2	7	9	2	20 (26%)	0 (0%)	1 (1%)	25 (33%)
1	Below clause	14	2	16 (21%)	3	7	8	1	19 (25%)	0 (0%)	1 (1%)	36 (47%)
1	Clause and above	3	1	4 (5%)	0	0	4	1	5 (7%)	1 (1%)	0 (0%)	10 (13%)
1	Total	23	3	26 (34%)	5	17	21	4	47 (62%)	1 (1%)	2 (3%)	76 (100%)
2	Below word	0	0	0 (0%)	0	7	0	0	7 (8%)	0 (0%)	0 (0%)	7 (8%)
2	Word	5	1	6 (6%)	4	7	9	1	21 (23%)	0 (0%)	1 (1%)	28 (30%)
2	Below clause	6	0	6 (6%)	8	11	12	2	33 (35%)	0 (0%)	2 (2%)	41 (44%)
2	Clause and above	1	2	3 (3%)	1	1	7	3	12 (13%)	1 (1%)	1 (1%)	17 (18%)
2	Total	12	3	15 (16%)	13	26	28	6	73 (78%)	1 (1%)	4 (4%)	93 (100%)
3	Below word	0	0	0 (0%)	0	7	0	0	7 (9%)	0 (0%)	0 (0%)	7 (9%)
3	Word	3	0	3 (4%)	3	13	10	0	26 (34%)	0 (0%)	1 (1%)	30 (39%)
3	Below clause	2	1	3 (4%)	4	5	5	0	14 (18%)	1 (1%)	2 (3%)	20 (26%)
3	Clause and above	4	2	6 (8%)	2	3	6	1	12 (16%)	0 (0%)	1 (1%)	19 (25%)
3	Total	9	3	12 (16%)	9	28	21	1	59 (78%)	1 (1%)	4 (5%)	76 (100%)
4	Below word	0	0	0 (0%)	0	3	1	1	5 (5%)	0 (0%)	0 (0%)	5 (5%)

4	Word	6	0	6	(6%)	6	6	15	4	31	(33%)	0	(0%)	1	(1%)	38	(40%)
4	Below clause	6	0	6	(6%)	5	7	12	1	25	(27%)	0	(0%)	1	(1%)	32	(34%)
4	Clause and above	8	3	11	(12%)	0	1	3	0	4	(4%)	0	(0%)	4	(4%)	19	(20%)
4	Total	20	3	23	(24%)	11	17	31	6	65	(69%)	0	(0%)	6	(6%)	94	(100%)
5	Below word	0	0	0	(0%)	0	1	1	0	2	(2%)	0	(0%)	0	(0%)	2	(2%)
5	Word	4	0	4	(4%)	2	9	19	3	33	(33%)	0	(0%)	2	(2%)	39	(39%)
5	Below clause	10	0	10	(10%)	7	7	15	0	29	(29%)	0	(0%)	0	(0%)	39	(39%)
5	Clause and above	12	0	12	(12%)	2	3	2	1	8	(8%)	1	(1%)	0	(0%)	21	(21%)
5	Total	26	0	26	(26%)	11	20	37	4	72	(71%)	1	(1%)	2	(2%)	101	(100%)
Total	Below word	2	0	2	(0%)	0	21	2	1	24	(5%)	0	(0%)	0	(0%)	26	(6%)
Total	Word	22	1	23	(5%)	17	42	62	10	131	(30%)	0	(0%)	6	(1%)	160	(36%)
Total	Below clause	38	3	41	(9%)	27	37	52	4	120	(27%)	1	(0%)	6	(1%)	168	(38%)
Total	Clause and above	28	8	36	(8%)	5	8	22	6	41	(9%)	3	(1%)	6	(1%)	86	(20%)
Total	Total	90	12	102	(23%)	49	108	138	21	316	(72%)	4	(1%)	18	(4%)	440	(100%)

Table 4.36 *Revision-associated cognitive activities by stage and context in L2 Chinese argumentative writing (N = 16)*

		Planning				Translation						Others		Not recall		Total	
Stage	Context	Content	Organisation	Total (%)		Translation in general	Lexical retrieval	Syntactic encoding	Cohesion	Total (%)		Total (%)		Total (%)		Total (%)	
1	Pre-contextual	11	0	11	(12%)	5	24	19	7	55	(59%)	0	(0%)	7	(7%)	73	(78%)
1	Contextual	4	6	10	(11%)	3	2	4	0	9	(10%)	0	(0%)	2	(2%)	21	(22%)
1	Total	15	6	21	(22%)	8	26	23	7	64	(68%)	0	(0%)	9	(10%)	94	(100%)
2	Pre-contextual	12	3	15	(13%)	5	26	6	3	40	(35%)	0	(0%)	13	(12%)	68	(60%)
2	Contextual	13	3	16	(14%)	6	4	16	2	28	(25%)	0	(0%)	1	(1%)	45	(40%)
2	Total	25	6	31	(27%)	11	30	22	5	68	(60%)	0	(0%)	14	(12%)	113	(100%)
3	Pre-contextual	12	3	15	(14%)	8	21	14	3	46	(43%)	1	(1%)	9	(8%)	71	(67%)
3	Contextual	10	3	13	(12%)	4	3	8	3	18	(17%)	0	(0%)	4	(4%)	35	(33%)
3	Total	22	6	28	(26%)	12	24	22	6	64	(60%)	1	(1%)	13	(12%)	106	(100%)
4	Pre-contextual	15	0	15	(17%)	6	22	5	1	34	(40%)	0	(0%)	8	(9%)	57	(66%)
4	Contextual	11	1	12	(14%)	1	4	7	2	14	(16%)	0	(0%)	3	(3%)	29	(34%)
4	Total	26	1	27	(31%)	7	26	12	3	48	(56%)	0	(0%)	11	(13%)	86	(100%)
5	Pre-contextual	8	1	9	(8%)	0	14	10	5	29	(27%)	0	(0%)	7	(7%)	45	(42%)
5	Contextual	10	5	15	(14%)	7	5	25	7	44	(42%)	0	(0%)	2	(2%)	61	(58%)
5	Total	18	6	24	(23%)	7	19	35	12	73	(69%)	0	(0%)	9	(8%)	106	(100%)
Total	Pre-contextual	58	7	65	(13%)	24	107	54	19	204	(40%)	1	(0%)	44	(9%)	314	(62%)

Total	Contextual	48	18	66	(13%)	21	18	60	14	113	(22%)	0	(0%)	12	(2%)	191	(38%)
Total	Total	106	25	131	(26%)	45	125	114	33	317	(63%)	1	(0%)	56	(11%)	505	(100%)

Table 4.37 *Revision-associated cognitive activities by stage and context in L2 Chinese narrative writing (N = 16)*

		Planning				Translation						Others		Not recall		Total	
Stage	Context	Content	Organisation	Total (%)		Translation in general	Lexical retrieval	Syntactic encoding	Cohesion	Total (%)		Total (%)		Total (%)		Total (%)	
1	Pre-contextual	17	3	20	(26%)	4	14	15	1	34	(45%)	1	(1%)	1	(1%)	56	(74%)
1	Contextual	6	0	6	(8%)	1	3	6	3	13	(17%)	0	(0%)	1	(1%)	20	(26%)
1	Total	23	3	26	(34%)	5	17	21	4	47	(62%)	1	(1%)	2	(3%)	76	(100%)
2	Pre-contextual	7	1	8	(9%)	10	21	16	3	50	(54%)	0	(0%)	2	(2%)	60	(65%)
2	Contextual	5	2	7	(8%)	3	5	12	3	23	(25%)	1	(1%)	2	(2%)	33	(35%)
2	Total	12	3	15	(16%)	13	26	28	6	73	(78%)	1	(1%)	4	(4%)	93	(100%)
3	Pre-contextual	5	1	6	(8%)	5	25	12	1	42	(55%)	1	(1%)	4	(5%)	54	(71%)
3	Contextual	4	2	6	(8%)	4	3	9	0	16	(21%)	0	(0%)	0	(0%)	22	(29%)
3	Total	9	3	12	(16%)	9	28	21	1	59	(78%)	1	(1%)	4	(5%)	76	(100%)
4	Pre-contextual	11	0	11	(12%)	6	13	10	2	31	(33%)	0	(0%)	3	(3%)	45	(48%)
4	Contextual	9	3	12	(13%)	5	4	21	4	34	(36%)	0	(0%)	3	(3%)	49	(52%)
4	Total	20	3	23	(24%)	11	17	31	6	65	(69%)	0	(0%)	6	(6%)	94	(100%)
5	Pre-contextual	11	0	11	(11%)	6	6	7	0	19	(19%)	1	(1%)	1	(1%)	32	(32%)
5	Contextual	15	0	15	(15%)	5	14	30	4	53	(52%)	0	(0%)	1	(1%)	69	(68%)
5	Total	26	0	26	(26%)	11	20	37	4	72	(71%)	1	(1%)	2	(2%)	101	(100%)
Total	Pre-contextual	51	5	56	(13%)	31	79	60	7	176	(40%)	3	(1%)	11	(3%)	247	(56%)

Total	Contextual	39	7	46	(10%)	18	29	78	14	139	(32%)	1	(0%)	7	(2%)	193	(44%)
Total	Total	90	12	102	(23%)	49	108	138	21	316	(72%)	4	(1%)	18	(4%)	440	(100%)

Table 4.38 *Revision-associated cognitive activities by stage and linguistic domain in L1 Chinese argumentative writing (N = 16)*

		Planning			Translation					Others	No recall	Total
Stage	Linguistic domain	Content	Organisation	Total (%)	Translation in general	Lexical retrieval	Syntactic encoding	Cohesion	Total (%)	Total (%)	Total (%)	Total (%)
1	Below word	2	0	2 (1%)	0	1	0	0	1 (1%)	0 (0%)	0 (0%)	3 (2%)
1	Word	19	2	21 (15%)	7	24	12	5	48 (34%)	0 (0%)	2 (1%)	71 (50%)
1	Below clause	12	1	13 (9%)	8	16	5	2	31 (22%)	0 (0%)	4 (3%)	48 (34%)
1	Clause and above	7	4	11 (8%)	2	1	5	1	9 (6%)	0 (0%)	0 (0%)	20 (14%)
1	Total	40	7	47 (33%)	17	42	22	8	89 (63%)	0 (0%)	6 (4%)	142 (100%)
2	Below word	0	0	0 (0%)	0	3	1	0	4 (2%)	0 (0%)	0 (0%)	4 (2%)
2	Word	21	5	26 (13%)	9	26	12	7	54 (28%)	0 (0%)	6 (3%)	86 (45%)
2	Below clause	27	4	31 (16%)	10	10	11	3	34 (18%)	0 (0%)	8 (4%)	73 (38%)
2	Clause and above	12	2	14 (7%)	4	2	5	3	14 (7%)	0 (0%)	2 (1%)	30 (16%)
2	Total	60	11	71 (37%)	23	41	29	13	106 (55%)	0 (0%)	16 (8%)	193 (100%)
3	Below word	0	0	0 (0%)	0	2	0	0	2 (1%)	0 (0%)	0 (0%)	2 (1%)
3	Word	23	0	23 (14%)	8	27	12	6	53 (32%)	0 (0%)	3 (2%)	79 (47%)
3	Below clause	15	3	18 (11%)	7	12	9	2	30 (18%)	0 (0%)	5 (3%)	53 (32%)
3	Clause and above	17	6	23 (14%)	2	2	5	0	9 (5%)	0 (0%)	2 (1%)	34 (20%)
3	Total	55	9	64 (38%)	17	43	26	8	94 (56%)	0 (0%)	10 (6%)	168 (100%)
4	Below word	1	0	1 (1%)	2	0	0	0	2 (1%)	0 (0%)	0 (0%)	3 (2%)
4	Word	22	1	23 (14%)	7	17	10	4	38 (23%)	2 (1%)	4 (2%)	67 (41%)

4	Below clause	22	0	22 (13%)	18	9	4	1	32 (20%)	0 (0%)	6 (4%)	60 (37%)
4	Clause and above	19	6	25 (15%)	2	0	4	1	7 (4%)	0 (0%)	2 (1%)	34 (21%)
4	Total	64	7	71 (43%)	29	26	18	6	79 (48%)	2 (1%)	12 (7%)	164 (100%)
5	Below word	2	0	2 (1%)	0	1	1	0	2 (1%)	0 (0%)	0 (0%)	4 (2%)
5	Word	10	2	12 (7%)	6	28	5	10	49 (29%)	0 (0%)	6 (4%)	67 (40%)
5	Below clause	23	0	23 (14%)	15	11	2	5	33 (20%)	0 (0%)	4 (2%)	60 (36%)
5	Clause and above	13	3	16 (10%)	1	4	8	1	14 (8%)	2 (1%)	5 (3%)	37 (22%)
5	Total	48	5	53 (32%)	22	44	16	16	98 (58%)	2 (1%)	15 (9%)	168 (100%)
Total	Below word	5	0	5 (1%)	2	7	2	0	11 (1%)	0 (0%)	0 (0%)	16 (2%)
Total	Word	95	10	105 (13%)	37	122	51	32	242 (29%)	2 (0%)	21 (3%)	370 (44%)
Total	Below clause	99	8	107 (13%)	58	58	31	13	160 (19%)	0 (0%)	27 (3%)	294 (35%)
Total	Clause and above	68	21	89 (11%)	11	9	27	6	53 (6%)	2 (0%)	11 (1%)	155 (19%)
Total	Total	267	39	306 (37%)	108	196	111	51	466 (56%)	4 (0%)	59 (7%)	835 (100%)

Table 4.39 *Revision-associated cognitive activities by stage and linguistic domain in L1 Chinese narrative writing (N = 16)*

		Planning			Translation					Others	No recall	Total
Stage	Linguistic domain	Content	Organisation	Total (%)	Translation in general	Lexical retrieval	Syntactic encoding	Cohesion	Total (%)	Total (%)	Total (%)	Total (%)
1	Below word	1	0	1 (1%)	0	2	0	0	2 (2%)	0 (0%)	0 (0%)	3 (3%)
1	Word	7	1	8 (8%)	3	9	6	6	24 (24%)	0 (0%)	5 (5%)	37 (38%)
1	Below clause	8	0	8 (8%)	17	2	5	2	26 (27%)	0 (0%)	5 (5%)	39 (40%)
1	Clause and above	7	4	11 (11%)	3	0	4	1	8 (8%)	0 (0%)	0 (0%)	19 (19%)
1	Total	23	5	28 (29%)	23	13	15	9	60 (61%)	0 (0%)	10 (10%)	98 (100%)
2	Below word	0	0	0 (0%)	0	6	1	0	7 (5%)	0 (0%)	0 (0%)	7 (5%)
2	Word	20	0	20 (15%)	2	9	12	1	24 (18%)	2 (1%)	7 (5%)	53 (39%)
2	Below clause	17	1	18 (13%)	11	11	14	2	38 (28%)	0 (0%)	4 (3%)	60 (44%)
2	Clause and above	13	0	13 (10%)	0	0	1	0	1 (1%)	0 (0%)	2 (1%)	16 (12%)
2	Total	50	1	51 (38%)	13	26	28	3	70 (51%)	2 (1%)	13 (10%)	136 (100%)
3	Below word	2	0	2 (1%)	0	4	0	0	4 (3%)	0 (0%)	0 (0%)	6 (4%)
3	Word	15	1	16 (11%)	7	19	7	5	38 (25%)	1 (1%)	3 (2%)	58 (38%)
3	Below clause	25	1	26 (17%)	10	9	7	1	27 (18%)	0 (0%)	3 (2%)	56 (37%)
3	Clause and above	17	1	18 (12%)	6	2	2	3	13 (9%)	0 (0%)	1 (1%)	32 (21%)
3	Total	59	3	62 (41%)	23	34	16	9	82 (54%)	1 (1%)	7 (5%)	152 (100%)
4	Below word	2	0	2 (2%)	0	0	0	0	0 (0%)	0 (0%)	0 (0%)	2 (2%)
4	Word	14	0	14 (12%)	9	11	7	3	30 (25%)	0 (0%)	3 (2%)	47 (39%)

4	Below clause	11	0	11 (9%)	18	5	11	3	37 (31%)	0 (0%)	2 (2%)	50 (41%)
4	Clause and above	12	1	13 (11%)	2	0	2	4	8 (7%)	0 (0%)	1 (1%)	22 (18%)
4	Total	39	1	40 (33%)	29	16	20	10	75 (62%)	0 (0%)	6 (5%)	121 (100%)
5	Below word	1	0	1 (1%)	0	0	0	0	0 (0%)	0 (0%)	2 (1%)	3 (2%)
5	Word	15	1	16 (12%)	4	11	9	3	27 (19%)	0 (0%)	2 (1%)	45 (32%)
5	Below clause	16	1	17 (12%)	14	9	3	6	32 (23%)	0 (0%)	4 (3%)	53 (38%)
5	Clause and above	14	6	20 (14%)	6	0	7	3	16 (12%)	0 (0%)	2 (1%)	38 (27%)
5	Total	46	8	54 (39%)	24	20	19	12	75 (54%)	0 (0%)	10 (7%)	139 (100%)
Total	Below word	6	0	6 (1%)	0	12	1	0	13 (2%)	0 (0%)	2 (0%)	21 (3%)
Total	Word	71	3	74 (11%)	25	59	41	18	143 (22%)	3 (0%)	20 (3%)	240 (37%)
Total	Below clause	77	3	80 (12%)	70	36	40	14	160 (25%)	0 (0%)	18 (3%)	258 (40%)
Total	Clause and above	63	12	75 (12%)	17	2	16	11	46 (7%)	0 (0%)	6 (1%)	127 (20%)
Total	Total	217	18	235 (36%)	112	109	98	43	362 (56%)	3 (0%)	46 (7%)	646 (100%)

Table 4.40 *Revision-associated cognitive activities by stage and context in L1 Chinese argumentative writing (N = 16)*

		Planning			Translation					Others	No recall	Total
Stage	Context	Content	Organisation	Total (%)	Translation in general	Lexical retrieval	Syntactic encoding	Cohesion	Total (%)	Total (%)	Total (%)	Total (%)
1	Pre-contextual	30	3	33 (23%)	16	35	18	2	71 (50%)	0 (0%)	4 (3%)	108 (76%)
1	Contextual	10	4	14 (10%)	1	7	4	6	18 (13%)	0 (0%)	2 (1%)	34 (24%)
1	Total	40	7	47 (33%)	17	42	22	8	89 (63%)	0 (0%)	6 (4%)	142 (100%)
2	Pre-contextual	45	5	50 (26%)	18	31	21	7	77 (40%)	0 (0%)	14 (7%)	141 (73%)
2	Contextual	15	6	21 (11%)	5	10	8	6	29 (15%)	0 (0%)	2 (1%)	52 (27%)
2	Total	60	11	71 (37%)	23	41	29	13	106 (55%)	0 (0%)	16 (8%)	193 (100%)
3	Pre-contextual	41	5	46 (27%)	16	30	20	4	70 (42%)	0 (0%)	10 (6%)	126 (75%)
3	Contextual	14	4	18 (11%)	1	13	6	4	24 (14%)	0 (0%)	0 (0%)	42 (25%)
3	Total	55	9	64 (38%)	17	43	26	8	94 (56%)	0 (0%)	10 (6%)	168 (100%)
4	Pre-contextual	47	4	51 (31%)	24	21	10	4	59 (36%)	2 (1%)	11 (7%)	123 (75%)
4	Contextual	17	3	20 (12%)	5	5	8	2	20 (12%)	0 (0%)	1 (1%)	41 (25%)
4	Total	64	7	71 (43%)	29	26	18	6	79 (48%)	2 (1%)	12 (7%)	164 (100%)
5	Pre-contextual	34	2	36 (21%)	18	32	7	10	67 (40%)	0 (0%)	12 (7%)	115 (68%)
5	Contextual	14	3	17 (10%)	4	12	9	6	31 (18%)	2 (1%)	3 (2%)	53 (32%)
5	Total	48	5	53 (32%)	22	44	16	16	98 (58%)	2 (1%)	15 (9%)	168 (100%)
Total	Pre-contextual	197	19	216 (26%)	92	149	76	27	344 (41%)	2 (0%)	51 (6%)	613 (73%)
Total	Contextual	70	20	90 (11%)	16	47	35	24	122 (15%)	2 (0%)	8 (1%)	222 (27%)

Total	Total	267	39	306 (37%)	108	196	111	51	466 (56%)	4 (0%)	59 (7%)	835 (100%)

Table 4.41 *Revision-associated cognitive activities by stage and context in L1 Chinese narrative writing (N = 16)*

		Planning			Translation					Others	No recall	Total
Stage	Context	Content	Organisation	Total (%)	Translation in general	Lexical retrieval	Syntactic encoding	Cohesion	Total (%)	Total (%)	Total (%)	Total (%)
1	Pre-contextual	20	4	24 (24%)	19	12	13	7	51 (52%)	0 (0%)	10 (10%)	85 (87%)
1	Contextual	3	1	4 (4%)	4	1	2	2	9 (9%)	0 (0%)	0 (0%)	13 (13%)
1	Total	23	5	28 (29%)	23	13	15	9	60 (61%)	0 (0%)	10 (10%)	98 (100%)
2	Pre-contextual	37	1	38 (28%)	10	25	17	3	55 (40%)	2 (1%)	13 (10%)	108 (79%)
2	Contextual	13	0	13 (10%)	3	1	11	0	15 (11%)	0 (0%)	0 (0%)	28 (21%)
2	Total	50	1	51 (38%)	13	26	28	3	70 (51%)	2 (1%)	13 (10%)	136 (100%)
3	Pre-contextual	51	3	54 (36%)	19	31	15	6	71 (47%)	1 (1%)	6 (4%)	132 (87%)
3	Contextual	8	0	8 (5%)	4	3	1	3	11 (7%)	0 (0%)	1 (1%)	20 (13%)
3	Total	59	3	62 (41%)	23	34	16	9	82 (54%)	1 (1%)	7 (5%)	152 (100%)
4	Pre-contextual	36	1	37 (31%)	25	16	20	6	67 (55%)	0 (0%)	6 (5%)	110 (91%)
4	Contextual	3	0	3 (2%)	4	0	0	4	8 (7%)	0 (0%)	0 (0%)	11 (9%)
4	Total	39	1	40 (33%)	29	16	20	10	75 (62%)	0 (0%)	6 (5%)	121 (100%)
5	Pre-contextual	41	3	44 (32%)	18	18	15	4	55 (40%)	0 (0%)	9 (6%)	108 (78%)
5	Contextual	5	5	10 (7%)	6	2	4	8	20 (14%)	0 (0%)	1 (1%)	31 (22%)
5	Total	46	8	54 (39%)	24	20	19	12	75 (54%)	0 (0%)	10 (7%)	139 (100%)
Total	Pre-contextual	185	12	197 (30%)	91	102	80	26	299 (46%)	3 (0%)	44 (7%)	543 (84%)
Total	Contextual	32	6	38 (6%)	21	7	18	17	63 (10%)	0 (0%)	2 (0%)	103 (16%)

Total	Total	217	18	235 (36%)	112	109	98	43	362 (56%)	3 (0%)	46 (7%)	646 (100%)

Summary: the role of genre in Chinese writing Genre was found to affect writing behaviours both for the whole session and at different stages in L2 Chinese writing. For the whole session, L2 writers tended to be more fluent in argumentative writing, while at the same time they paused more frequently within a word in argumentative writing. L2 writers produced argumentative and narrative essays in different ways across the five stages. For argumentative writing, differences were mainly observed between the middle stage(s) and the last stage for three pause frequency measures (i.e. overall, between Pinyin and character, between clauses) and two revision measures (i.e. pre-contextual, Pinyin). For narrative writing, differences were found between the middle writing stages and Stages 1 and 5 for frequencies of pause overall and between-clause pause, and the numbers of total, pre-contextual and Pinyin revisions. In addition, differences concerning between-Pinyin-and-character pause frequency set apart the middle stages from Stage 5, and those concerning the number of below-clause revisions were observed between Stage 1 and the middle writing stages.

In terms of the distribution of pause-associated cognitive activities, L2 writers mentioned proportionately more translation processes, and fewer planning and monitoring processes in argumentative than narrative writing for the whole session. When stages of writing were considered, genre was found to have little influence on distribution. With regard to revision-associated cognitive activities for the entire writing session, proportionally more references were made to planning processes and fewer to translation processes in argumentative than narrative tasks. Across the five stages, relatively more planning activities were found in the middle stages of argumentative writing, and more translation activities in the initial and final stages. For narrative writing, comparatively more planning-related comments were found at the beginning, while more translation-associated references were made at the middle and the end.

Genre was also found to have an influence on writing behaviours in L1 Chinese. Similar to L2 writers, L1 writers showed greater speed fluency in argumentative than narrative writing for the whole session. L1 argumentative writing also generated more pauses at smaller textual unit boundaries (i.e. between words) than narrative writing. In terms of the effect of genre as a function of stage, similar to L2 writing, L1 narrative writing demonstrated more variation across stages than argumentative writing. The effect of stage was, to some extent, different between L1 and L2 argumentative writing, but largely similar between L1 and L2 narrative writing. In terms of cognitive activities associated with writing behaviours, as evidenced in the verbal protocols, the findings contradicted those for L2 writing. Genre in L1 writing seemed to have little influence on cognitive activities for the whole writing session. The temporal distribution of cognitive activities associated with pauses tended to be affected by genre to a small extent, but a limited genre effect was found for cognitive activities associated with revisions.

4.3 Proficiency in L2 Chinese and writing behaviours

The section reports the results for the relationships between L2 proficiency and

L2 Chinese writing behaviours. The L2 writers' proficiency in Chinese in the current study ranged from 13 to 41 out of 45 (M = 28.25, SD = 7.59), as measured by the cloze test described in Section 3.3.4. The relationships were investigated by running a series of mixed-effect linear regressions with measures of speed fluency, pausing or revision as the dependent variables and L2 proficiency, operationalised as the score on the cloze test, as the fixed effect. For all models, the initial random effects were, again, participant and prompt. Prompt was removed from some models in order to achieve model convergence. No random slope was added to avoid convergence failure. A log transformation was performed to a dependent variable when needed.

Statistically significant relationships were found between L2 proficiency and seven writing behaviour indices (Table 4.42): production rate, frequencies of pauses within a word and between Pinyin and character(s), and frequencies of revisions in total, below the clause level, at the clause level and above and on Chinese characters. See Table 9 in Appendix H for non-statistically significant relationships. The findings indicate that L2 writers of Chinese with higher proficiency tended to be more fluent, as measured by production rate, pause less frequently within a word and between Pinyin and character(s), and make more changes overall to larger textual units and Chinese characters. Except for the amount of revision at the clause level and above, L2 proficiency explained over 10% of the variance in all writing behaviour measures.

Table 4.42 *Significant results from linear mixed-effects regressions examining the relationships between L2 proficiency and writing behaviours (N = 32)*

Dependent variable	Fixed effect	*Est*	*SE*	*t*	*p*
Production rate	Proficiency	.51	.17	3.08	< .01

Model = lmer(production_rate ~ proficiency + (1|participant) + (1|prompt), data, REML=F)					
$R^2_m = .19$, $R^2_c = .80$					
Ww pause frequency	Proficiency	< -.01	< .01	2.48	.02
Model = lmer(ww_pause_frequency ~ proficiency + (1|participant) + (1|prompt), data, REML=F)					
$R^2_m = .10$, $R^2_c = .49$					
Bpc pause frequency	Proficiency	< .01	< .01	2.56	.02
Model = lmer(1/log(bpc_pause_frequency) ~ proficiency + (1|participant) + (1|prompt), data, REML=F)					
$R^2_m = .13$, $R^2_c = .48$					
Total revision	Proficiency	.02	.01	2.23	.03
Model = lmer(log(total_revision) ~ proficiency + (1|participant), data, REML=F)					
$R^2_m = .12$, $R^2_c = .88$					
Below-clause revision	Proficiency	.01	< .01	3.21	< .01
Model = lmer(log(below_clause_revision) ~ proficiency + (1|participant) + (1|prompt), data, REML=F)					
$R^2_m = .19$, $R^2_c = .74$					
Clause-and-above revision	Proficiency	< .01	< .01	2.10	.04
Model = lmer(log(clause_and_above_revision) ~ proficiency + (1|participant), data, REML=F)					
$R^2_m = .08$, $R^2_c = .50$					
Character revision	Proficiency	.02	.01	2.62	.01
Model = lmer(log(character_revision) ~ proficiency + (1|participant), data, REML=F)					
$R^2_m = .16$, $R^2_c = .86$					

Note: Production rate = Number of characters produced divided by active writing time (in minutes), Ww = Within a word, Bpc = Between Pinyin and character(s)

Next, to investigate how stages and genre of writing mediate the relationships between L2 proficiency and writing behaviours, I included stage/ genre and the interaction between the score on the cloze test and stage/ genre as the additional fixed effects of the models. In order to reduce collinearity within the model, the score on the cloze test was centred around its mean by subtracting the average score from that of each individual (Cunnings, 2012). As shown in Table 4.43, significant interactions between L2 proficiency and stage were found for five writing behaviour indices: P-burst length, between-clause pause duration, and frequencies of revision below the clause level, to Pinyin and Chinese characters. In terms of genre, one interaction was identified in the frequency of pause between contextual revisions. The amount of variance explained by the interaction was around 1%. See Tables 10 and 11 in Appendix H for non-significant results.

Table 4.43 *Significant results from linear mixed-effects regressions examining how stage/genre mediate the relationships between L2 proficiency and writing behaviours (N = 32)*

Dependent variable	Predictor: proficiency * stage/genre	*Est*	*SE*	*t*	*p*
P-burst length	Proficiency * 2	-.05	.02	-2.33	.02*
	Proficiency * 3	-.06	.02	-2.38	.02*
	Proficiency * 4	.01	.02	.36	.72
	Proficiency * 5	-.02	.02	-.81	.42
Model = lmer(pburst_length ~ stage*cproficiency + (1|participant) + (1|prompt), data, REML=F)					
Amount of variance explained by the interaction: $R^2_m = .01$, $R^2_c = .02$					
Bc pause duration	Proficiency * 2	-.02	.01	-2.06	.04*
	Proficiency * 3	-.02	.01	-2.14	.03*
	Proficiency * 4	-.01	.01	-.63	.53
	Proficiency * 5	-.02	.01	-2.20	.03*
Model = lmer(log(bc_pause_duration) ~ stage*cproficiency + (1|participant) + (1|prompt), data, REML=F)					
Amount of variance explained by the interaction: $R^2_m = .01$, $R^2_c = .01$					
Below-clause revision	Proficiency * 2	< .01	< .01	.41	.68
	Proficiency * 3	.01	< .01	2.90	<. 01*
	Proficiency * 4	< .01	< .01	.52	.60
	Proficiency * 5	.01	< .01	2.18	.03*
Model = lmer(log(below_clause_revision) ~ stage*cproficiency + (1|participant) + (1|prompt), data, REML=F)					
Amount of variance explained by the interaction: $R^2_m = .01$, $R^2_c = .02$					
Pinyin revision	Proficiency * 2	-.01	< .01	-2.12	.03*
	Proficiency * 3	-.01	< .01	-1.08	.28
	Proficiency * 4	< -.01	< .01	-.07	.94
	Proficiency * 5	< -.01	< .01	-.59	.56
Model = lmer(log(pinyin_revision) ~ stage*cproficiency + (1|participant), data, REML=F)					
Amount of variance explained by the interaction: $R^2_m < .01$, $R^2_c = .02$					
Character revision	Proficiency * 2	< .01	< .01	.95	.34
	Proficiency * 3	.01	< .01	3.39	< .01*
	Proficiency * 4	< .01	< .01	.83	.40
	Proficiency * 5	.01	< .01	2.25	.02*
Model = lmer(log(character_revision) ~ stage*cproficiency + (1|participant), data, REML=F)					
Amount of variance explained by the interaction: $R^2_m = .01$, $R^2_c = .01$					

Bcr pause frequency	Proficiency * narrative	.01	< .01	2.34	.02*

Model = lmer(sqrt(bcr_pause_frequency) ~ stage*cproficiency + (1|participant), data, REML=F)
Amount of variance explained by the interaction: $R^2_m = .01$, $R^2_c = .02$

Note: P-burst length = Number of characters between two pauses, Bc = Between clauses, Bcr = Between contextual revisions, R^2_m = marginal *R*-squared, R^2_c = conditional *R*-squared, * = statistically significant result

I further examined the interactions by running linear mixed-effects regression analyses for individual stage/ genre with L2 proficiency as the only fixed effect. The results of the analyses are shown in Table 4.44. Although interactions were found for six writing behaviour measures, follow-up analyses only confirmed two sets of relationships being mediated by stage. Specifically, L2 proficiency positively related to below-clause revision in Stages 3, 4 and 5, and explained 20%, 8% and 19% of the variance in the amount of below-clause revision in each stage. In addition, more proficient writers revised characters more often in Stages 2, 3 and 5, and the amount of variance attributed to L2 proficiency was 10% for Stage 2, 21% for Stage 3 and 17% for Stage 5. No relationship was found between L2 proficiency and either measure in Stage 1. The results indicate that the relationships between L2 proficiency and the two revision indices were not constant across stages. Higher-proficiency L2 writers tended to revise below the clause level and characters more often after the initial stage than their lower-proficiency counterparts.

Table 4.44 *Results from linear mixed-effects regressions examining the relationships between L2 proficiency and writing behaviours by stage/genre (N = 32)*

Dependent variable	Fixed effect	Stage	*Est*	*SE*	*t*	*p*
P-burst length	Proficiency	1	.04	.03	1.08	.29
	Proficiency	2	-.02	.03	-.56	.58
	Proficiency	3	-.02	.04	-.50	.62
	Proficiency	4	.04	.03	1.57	.13
	Proficiency	5	.02	.03	.57	.57
S1: Model = lmer(pburst_length_s1 ~ proficiency + (1|participant) + (1|prompt), data, REML = F)						
S2: Model = lmer(pburst_length_s2 ~ proficiency + (1|participant), data, REML = F)						
S3: Model = lmer(pburst_length_s3 ~ proficiency + (1|participant) + (1|prompt), data, REML = F)						
S4: Model = lmer(pburst_length_s4 ~ proficiency + (1|participant), data, REML = F)						
S5: Model = lmer(pburst_length_s5 ~ proficiency + (1|participant), data, REML = F)						
Bc pause duration	Proficiency	1	.02	.01	2.01	.05
	Proficiency	2	< -.01	.01	-.36	.72
	Proficiency	3	< -.01	.01	-.63	.53
	Proficiency	4	.01	.01	1.65	.11
	Proficiency	5	< -.01	.01	-.56	.58
S1: Model = lmer(bc_pause_duration_s1 ~ proficiency + (1|participant) + (1|prompt), data, REML = F)						
S2: Model = lmer(bc_pause_duration_s2 ~ proficiency + (1|participant), data, REML = F)						
S3: Model = lmer(bc_pause_duration_s3 ~ proficiency + (1|participant) + (1|prompt), data, REML = F)						
S4: Model = lmer(bc_pause_duration_s4 ~ proficiency + (1|participant) + (1|prompt), data, REML = F)						
S5: Model = lmer(bc_pause_duration_s5 ~ proficiency + (1|participant) + (1|prompt), data, REML = F)						
Below-clause	Proficiency	1	.01	< .01	1.77	.09

revision	Proficiency	2	.01	< .01	1.93	.06
	Proficiency	3	.02	< .01	3.89	< .01*
	Proficiency	4	.01	< .01	2.29	.03*
	Proficiency	5	.02	< .01	4.09	< .01*
S1: Model = lmer(log(below_clause_revision_s1) ~ proficiency + (1|participant) + (1|prompt), data, REML = F)						
S2: Model = lmer(log(below_clause_revision_s2) ~ proficiency + (1|participant), data, REML=F)						
S3: Model = lmer(log(below_clause_revision_s3) ~ proficiency + (1|participant) + (1|prompt), data, REML = F), $R^2_m = .20$, $R^2_c = .53$						
S4: Model = lmer(log(below_clause_revision_s4) ~ proficiency + (1|participant), data, REML = F), $R^2_m = .08$, $R^2_c = .39$						
S5: Model = lmer(log(below_clause_revision_s5) ~ proficiency + (1|participant), data, REML = F), $R^2_m = .19$, $R^2_c = .39$						
Pinyin revision	Proficiency	1	.01	.01	1.52	.14
	Proficiency	2	< .01	.01	.48	.63
	Proficiency	3	.01	.01	.99	.33
	Proficiency	4	.01	.01	1.61	.12
	Proficiency	5	.01	.01	1.44	.16
S1: Model = lmer(log(pinyin_revision_s1) ~ proficiency + (1|participant) + (1|prompt), data, REML = F)						
S2: Model = lmer(log(pinyin_revision_s2) ~ proficiency + (1|participant), data, REML = F)						
S3: Model = lmer(log(pinyin_revision_s3) ~ proficiency + (1|participant) + (1|prompt), data, REML = F)						
S4: Model = lmer(log(pinyin_revision_s4) ~ proficiency + (1|participant), data, REML = F)						
S5: Model = lmer(log(pinyin_revision_s5) ~ proficiency + (1|participant) + (1|prompt), data, REML = F)						
Character revision	Proficiency	1	.01	.01	1.89	.07
	Proficiency	2	.02	.01	2.17	.04*
	Proficiency	3	.03	.01	3.44	< .01*
	Proficiency	4	.02	.01	1.94	.06
	Proficiency	5	.02	.01	2.99	< .01*
S1: Model = lmer(log(character_revision_s1) ~ proficiency + (1|participant), data, REML = F)						
S2: Model = lmer(log(character_revision_s2) ~ proficiency + (1|participant), data, REML = F), $R^2_m = .10$, $R^2_c = .67$						

S3: Model = lmer(log(character_revision_s3) ~ proficiency + (1|participant) + (1|prompt), data, REML = F), R^2_m = .21, R^2_c = .70

S4: Model = lmer(log(character_revision_s4) ~ proficiency + (1|participant), data, REML = F)

S5: Model = lmer(log(character_revision_s5) ~ proficiency + (1|participant) + (1|prompt), data, REML = F), R^2_m = .17, R^2_c = .71

		Genre				
Bcr pause frequency	Proficiency	Argumentative	< .01	< .01	.35	.73
	Proficiency	Narrative	.01	< .01	1.28	.21

argumentative: Model = lmer(log(bcr_pause_frequency_arg) ~ proficiency + (1|participant) + (1|prompt), data, REML = F)

narrative: Model = lmer(log(bcr_pause_frequency_nar) ~ proficiency + (1|participant) + (1|prompt), data, REML = F)

Note: P-burst length = Number of characters between two pauses, Bc = Between clauses, Bcr = Between contextual revisions, R^2_m = marginal R-squared, R^2_c = conditional R-squared, * = statistically significant result

Summary: relationships between L2 Chinese proficiency and writing behaviours The increased L2 Chinese proficiency led to greater speed fluency, as measured by production rate, and more frequent revisions above the word level and to Chinese characters. Writers with higher L2 proficiency also paused less often on a word, and between Pinyin and character(s). The links between L2 proficiency and writing behaviours tended to be modulated by stages of writing. Specifically, more proficient L2 writers revised more often below the clause level and made more changes to Chinese characters than less proficient writers after the initial writing stage. The relationships between L2 proficiency and writing behaviours were similar in argumentative and narrative writing.

4.4 Writing behaviours and text quality in Chinese

The last section of the chapter presents the results from linear mixed-effects regression analyses that examined the extent to which writing behaviours predict text quality in Chinese writing. First, I investigated how writing behaviours predict text quality when taking the writing process as a whole. In this set of analyses, the dependent variable was text quality, operationalised as the average score for the text given by the two raters. The fixed effect was a writing behaviour index with participant and prompt as the random effects. For some analyses, prompt was removed to achieve model convergence. No by-participant random slope was included in the models due to the failure to converge.

The analyses yielded statistically significant results for six writing behaviour indices in L2 Chinese (Table 4.45): production rate, P-burst length, between-word pause frequency, overall pause duration, between-word pause duration and between-clause pause duration. Non-significant results are reported in Table 12 in Appendix H. The findings indicated that greater speed fluency, less frequent pauses between word, and shorter pauses overall, between words and between clauses contributed to better texts in L2 Chinese. In addition, speed fluency measures were found to be stronger predictors of L2 text quality than pausing measures, as production rate and P-burst length explained a larger amount of variance (4% and 6%, respectively) than the four pausing indices (2% for between-word pause frequency, 2% for total pause duration, 1% for between-word pause duration and 1% for between-clause pause duration). In L1 writing, none of the writing behaviour indices was found to predict text quality. See Table 13 in Appendix H for the results for L1 writing.

Table 4.45 *Significant results from linear mixed-effects regressions examining how writing behaviours predict L2 Chinese text quality (N = 32)*

Dependent variable	Predictor	*Est*	*SE*	*t*	*p*
Text quality	Production rate	.02	.01	2.51	.01
Model = lmer(text_score ~ production_rate + (1|participant) + (1|prompt), data, REML = F)					
$R^2_m = .04$, $R^2_c = .83$					
Text quality	P-burst length	.16	.05	3.20	< .01
Model = lmer(text_score ~ pburst_length + (1|participant) + (1|prompt), data, REML = F)					
$R^2_m = .06$, $R^2_c = .86$					
Text quality	Bw pause frequency	-.22	.09	-2.41	.02
Model = lmer(text_score ~ bw_pause_frequency + (1|participant) + (1|prompt), data, REML = F)					
$R^2_m = .02$, $R^2_c = .86$					
Text quality	Total pause duration	-.16	.06	-2.06	.01
Model = lmer(text_score ~ total_pause_duration + (1|participant) + (1|prompt), data, REML = F)					
$R^2_m = .02$, $R^2_c = .88$					
Text quality	Bw pause duration	-.16	.06	-3.06	< .01

Model = lmer(text_score ~bw_pause_duration + (1|participant) + (1|prompt), data, REML = F)					
R^2_m = .01, R^2_c = .87					
Text quality	Bc pause duration	-.06	.02	-2.71	< .01
Model = lmer(text_score ~bw_pause_duration + (1|participant), data, REML = F)					
R^2_m = .01, R^2_c = .86					

Note: Production rate = Number of characters produced divided by active writing time (in minutes), P-burst length = Number of characters between two pauses, Bw = Between words, Bc = Between clauses, R^2_m = marginal *R*-squared, R^2_c = conditional *R*-squared

Next, I explored whether the relationship between writing behaviours and text quality can be mediated by genre. To address this, genre and its interaction with the writing behaviour measures were added as additional fixed effects to the models. No interaction was found between genre and writing behaviours indices for L2 text quality (See Table 14 in Appendix H for the results). One interaction (Table 4.46) was identified between genre and the frequency of pause between contextual revisions to L1 text quality. See Table 15 in Appendix H for non-significant interactions. By performing separate analyses for argumentative and narrative writing with the frequency of pauses between contextual revisions as the single fixed effect, the results showed that L1 writers who paused less often between contextual revisions tended to produce better argumentative essays (5% of the variance explained), whereas no relationship was found in narrative writing.

Table 4.46 *Significant results from linear mixed-effects regressions examining the effects of interaction between genre and writing behaviours on L1 Chinese text quality (N = 32)*

Dependent variable	Predictor	*Est*	*SE*	*t*	*p*
Text quality	Bcr pause frequency * genre (narrative)	.84	.34	2.50	.01*
Model = lmer(text_score ~ bcr_pause_frequency*genre + (1|participant), data, REML = F)					
Amount of variance explained by the interaction: R^2_m = .02, R^2_c = .02					
Individual analysis by genre					

Text quality (argumentative)	Bcr pause frequency	-.65	.32	-2.01	.05*
Model = lmer(text_score ~ bcr_pause_frequency_arg + (1|participant), data, REML = F) R^2_m = .05, R^2_c = .69					
Text quality (narrative)	Bcr pause frequency	.14	.37	.37	.71
Model = lmer(text_score ~ bcr_pause_frequency_nar + (1|participant) + (1|prompt), data, REML = F)					

Note: Bcr = Between contextual revisions, R^2_m = marginal *R*-squared, R^2_c = conditional *R*-squared, * = statistically significant results

In terms of L2 writing, I additionally examined whether L2 proficiency can modulate the relationship between writing behaviours and text quality. To address this, two additional fixed effects, the score on the proficiency test and its interaction with writing behaviour measures, were added to the models. The score on the proficiency test was centred by subtracting the average score from that of each individual so as to reduce the collinearity between the fixed effects (Cunnings, 2012). Tables 4.47 and 16 in Appendix H separately present the significant and non-significant results for the analyses. Interactions were found between L2 proficiency and three behaviour measures: between-sentence pause frequency, between-word pause duration and revision at the word level, each explaining 2%, less than 1% and 3% of the variance in text quality. The results indicated that the three indices predicted text quality in different ways for participants with different L2 proficiency. Specifically, more pauses at the sentence boundary improved text quality for L2 writers who scored below 28 out of 45 on the cloze test, while for those who scored 28 and above, more between-sentence pauses hindered text quality. Longer pauses between words benefitted text scores for L2 writers who achieved over 39 in the test, whereas for those who gained 39 and below, between-word pause duration had a negative effect on text quality. Revision at the word level was positively linked to text quality only when L2 proficiency reached a certain level (32 and above in this case).

Table 4.47 *Significant results from linear mixed-effects regressions examining the effects of interaction between proficiency and writing behaviours on L2 Chinese text quality (N = 32)*

Dependent variable	Predictor	*Est*	*SE*	*t*	*p*
Text quality	Bs pause frequency	.08	.24	.34	.74
	Proficiency	.14	.02	5.58	< .01*
	Bs pause frequency * proficiency	-.08	.03	-2.60	.01*
Model = lmer(text_score ~ bs_pause_frequency*cproficiency + (1|participant) + (1|prompt), data, REML)					
Amount of variance explained by the interaction: $R^2_m = .02$, $R^2_c = .01$					
Text quality	Bw pause duration	-.17	.05	-3.12	< .01*
	Proficiency	.03	.03	.85	.40
	Bw pause duration * proficiency	.02	.01	2.08	.04*
Model = lmer(text_score ~ bw_pause_duration*cproficiency + (1|participant) + (1|prompt), data, REML)					
Amount of variance explained by the interaction: $R^2_m < .01$, $R^2_c = .01$					
Text quality	Word revision	-.29	.24	-1.20	.23
	Proficiency	.03	.03	1.28	.21
	Word revision * proficiency	.08	.03	2.21	.03*
Model = lmer(text_score ~ log(word_revision)*cproficiency + (1|participant) + (1|prompt), data, REML)					
Amount of variance explained by the interaction: $R^2_m = .03$, $R^2_c < .01$					

Note: Bw = Between words, Bs = Between sentences, R^2_m = marginal *R*-squared, R^2_c = conditional *R*-squared, * = statistically significant results

Second, the extent to which writing behaviours predict text quality was looked into when the writing process was divided into five stages. In the series of linear mixed-effects regression models that I constructed to address this, the dependent variables were, again, the average scores of the texts. The fixed effects were an index of writing behaviour, stage of writing and the interaction between them. The random effect was participant, and prompt was included as a second random effect only when the models succeeded in converging. No by-participant random slope was added to avoid a failure to converge. The results (Table 4.48) revealed interactions between stage and four pausing measures in L2 writing: between-word pause frequency, and the duration of pauses within a word, between Pinyin and character(s) and between clauses, and three behaviour measures in L1 writing: between-sentence pause frequency, total pause duration and contextual revision. See Tables 17 and 18 in Appendix H for non-significant results for the interactions.

Table 4.48 *Significant results from linear mixed-effects regressions examining the effects of interaction between writing behaviours and stage on Chinese text quality*

Dependent variable	Predictor: behaviour index * stage	*E*	*SE*	*t*	*p*
L2 Chinese (N = 32)					
Text quality	Bw pause frequency * 2	-.10	.05	-2.19	.03*
	Bw pause frequency * 3	-.08	.05	-1.79	.07
	Bw pause frequency * 4	-.05	.04	-1.13	.26
	Bw pause frequency * 5	-.05	.04	-1.05	.29
Model = lmer(text_score ~ bw_pause_frequency * stage + (1|participant) + (1|prompt), data, REML = F)					
Amount of variance explained by the interaction: $R^2_m < .01$, $R^2_c < .01$					
Text quality	Ww pause duration * 2	-.04	.02	-1.67	.10
	Ww pause duration * 3	-.05	.02	-2.14	.03*
	Ww pause duration * 4	-.02	.02	1.44	.15
	Ww pause duration * 5	-.03	.02	-1.72	.09
Model = lmer(text_score ~ ww_pause_duration * stage + (1|participant) + (1|prompt), data, REML = F)					
Amount of variance explained by the interaction: $R^2_m < .01$, $R^2_c < .01$					
Text quality	Bpc pause duration * 2	.05	.02	2.32	.02*
	Bpc pause duration * 3	-.02	.02	.89	.37
	Bpc pause duration * 4	.02	.03	.53	.60
	Bpc pause duration * 5	< .01	.02	.11	.92
Model = lmer(text_score ~ bpc_pause_duration * stage + (1|participant) + (1|prompt), data, REML = F)					
Amount of variance explained by the interaction: $R^2_m < .01$, $R^2_c < .01$					

Text quality	Bc pause duration * 2	< -.01	.01	-.87	.39
	Bc pause duration * 3	< -.01	.01	-.71	.48
	Bc pause duration * 4	< .01	.01	.23	.82
	Bc pause duration * 5	-.02	.01	-2.04	.04*
Model = lmer(text_score ~ bc_pause_duration * stage + (1|participant) + (1|prompt), data, REML = F)					
Amount of variance explained by the interaction: $R^2_m < .01$, $R^2_c < .01$					
L1 Chinese (N = 32)					
Text quality	Bs pause frequency * 2	.25	.10	2.49	.01*
	Bs pause frequency * 3	.03	.11	.35	.73
	Bs pause frequency * 4	.14	.10	1.37	.17
	Bs pause frequency * 5	.05	.11	.43	.67
Model = lmer(text_score ~ bs_pause_frequency * stage + (1|participant) + (1|prompt), data, REML = F)					
Amount of variance explained by the interaction: $R^2_m < .01$, $R^2_c < .01$					
Text quality	Total pause duration * 2	-.03	.02	-1.18	.24
	Total pause duration * 3	-.02	.03	-.48	.63
	Total pause duration * 4	-.07	.03	-2.18	.03*
	Total pause duration * 5	-.02	.02	-1.19	.24
Model = lmer(text_score ~ total_pause_duration * stage + (1|participant) + (1|prompt), data, REML = F)					
Amount of variance explained by the interaction: $R^2_m < .01$, $R^2_c < .01$					
Text quality	Contextual revision * 2	-.20	.12	-1.67	.10
	Contextual revision * 3	-.14	.12	-1.13	.26
	Contextual revision * 4	-.17	.14	-1.27	.20
	Contextual revision * 5	-.29	.13	-2.30	.02*
Model = lmer(text_score ~ log(contextual_revision) * stage + (1|participant) + (1|prompt), data, REML = F)					
Amount of variance explained by the interaction: $R^2_m < .01$, $R^2_c < .01$					

Note: Ww = Within a word, Bw = Between words, Bpc = Between Pinyin and character(s), Bc = Between clauses, Bs = Between sentences, R^2_m = marginal *R*-squared, R^2_c = conditional *R*-squared, * = statistically significant results

The interactions were examined by running linear mixed-effects regression analyses for each stage with writing behaviour measure as the single fixed effect. The results (Table 4.49) showed that in L2 Chinese writing, longer within-word pauses in Stage 1 and fewer between-word pauses in Stage 2 led to better texts. The duration of pause between Pinyin and character(s) positively correlated with text quality in Stage 2, but the correlation turned negative in Stage 3. In addition, longer between-clause pauses in Stage 5 improved text quality. In each model, the fixed effect, the writing behaviour measure, explained 1% of the variance in text quality. In L1 Chinese writing, better texts were associated with more between-sentence pauses in Stage 2 (4% of variance explained) and shorter pauses overall in Stage 4 (2% of variance explained). L1 writers who made fewer revisions to already-written texts in Stage 5 produced texts of higher quality, and the amount of contextual revision accounted for 2% of the variance in L1 text quality.

Table 4.49 *Results from linear mixed-effects regressions examining the effects of writing behaviours on Chinese text quality by stage*

Dependent variable	Fixed effect	Stage	*Est*	*SE*	*t*	*p*
L2 Chinese (N =32)						
Text quality	Bw pause frequency	1	.05	.05	.95	.35
	Bw pause frequency	2	-.14	.05	-2.73	.01*
	Bw pause frequency	3	-.10	.05	-1.94	.06
	Bw pause frequency	4	-.06	.05	-1.39	.17
	Bw pause frequency	5	-.06	.05	-1.21	.23
S1: Model = lmer(text_score ~ bw_pause_frequency_s1 + (1|participant) + (1|prompt), data, REML = F)						
S2: Model = lmer(text_score ~ bw_pause_frequency_s2 + (1|participant) + (1|prompt), data, REML = F), $R^2_m = .01$, $R^2_c = .87$						
S3: Model = lmer(text_score ~ bw_pause_frequency_s3 + (1|participant), data, REML = F)						
S4: Model = lmer(text_score ~ bw_pause_frequency_s4 + (1|participant) + (1|prompt), data, REML = F)						
S5: Model = lmer(text_score ~ bw_pause_frequency_s5 + (1|participant) + (1|prompt), data, REML = F)						
Text quality	Ww pause duration	1	.23	.11	2.05	.05*
	Ww pause duration	2	-.03	.12	-.27	.79
	Ww pause duration	3	-.02	.03	-.84	.41
	Ww pause duration	4	< .01	.01	.33	.74
	Ww pause duration	5	-.02	.08	-.32	.75
S1: Model = lmer(text_score ~ log(ww_pause_duration_s1) + (1|participant) + (1|prompt), data, REML = F), $R^2_m = .01$, $R^2_c = .87$						
S2: Model = lmer(text_score ~log(ww_pause_duration_s2) + (1|participant) + (1|prompt), data, REML = F)						
S3: Model = lmer(text_score ~ ww_pause_duration_s3 + (1|participant) + (1|prompt), data, REML = F)						
S4: Model = lmer(text_score ~ ww_pause_duration_s4 + (1|participant) + (1|prompt), data, REML = F)						
S5: Model = lmer(text_score ~ log(ww_pause_duration_s5) + (1|participant) + (1|prompt), data, REML = F)						
Text quality	Bpc pause duration	1	< .01	.02	.24	.81
	Bpc pause duration	2	.07	.02	2.76	.01*

	Bpc pause duration	3	-.23	.10	-2.27	.03*
	Bpc pause duration	4	.02	.04	.45	.65
	Bpc pause duration	5	< .01	.02	.06	.95
S1: Model = lmer(text_score ~ bpc_pause_duration_s1 + (1|participant) + (1|prompt), data, REML = F)						
S2: Model = lmer(text_score ~ bpc_pause_duration_s2 + (1|participant) + (1|prompt), data, REML = F), R^2_m = .01, R^2_c = .88						
S3: Model = lmer(text_score ~ log(bpc_pause_duration_s3) + (1|participant) + (1|prompt), data, REML = F), R^2_m = .01, R^2_c = .89						
S4: Model = lmer(text_score ~ bpc_pause_duration_s4 + (1|participant) + (1|prompt), data, REML = F)						
S5: Model = lmer(text_score ~ bpc_pause_duration_s5 + (1|participant) + (1|prompt), data, REML = F)						
Text quality	Bc pause duration	1	< -.01	.01	-.39	.69
	Bc pause duration	2	-.01	.01	-.97	.33
	Bc pause duration	3	-.02	.01	-1.26	.21
	Bc pause duration	4	.01	.08	.15	.89
	Bc pause duration	5	-.17	.06	-2.79	.01*
S1: Model = lmer(text_score ~ bc_pause_duration_s1 + (1|participant) + (1|prompt), data, REML = F)						
S2: Model = lmer(text_score ~ bc_pause_duration_s2 + (1|participant) + (1|prompt), data, REML = F)						
S3: Model = lmer(text_score ~ bc_pause_duration_s3 + (1|participant) + (1|prompt), data, REML = F)						
S4: Model = lmer(text_score ~ log(bc_pause_duration_s4) + (1|participant) + (1|prompt), data, REML = F)						
S5: Model = lmer(text_score ~ log(bc_pause_duration_s5) + (1|participant) + (1|prompt), data, REML = F), R^2_m = .01, R^2_c = .87						
L1 Chinese (N =32)						
Text quality	Bs pause frequency	1	-.11	.09	-1.22	.22
	Bs pause frequency	2	.29	.11	2.80	.01*
	Bs pause frequency	3	-.04	.11	-.40	.69
	Bs pause frequency	4	.14	.12	1.20	.23
	Bs pause frequency	5	-.02	.11	-.17	.87
S1: Model = lmer(text_score ~ bs_pause_frequency_s1 + (1|participant) + (1|prompt), data, REML = F)						
S2: Model = lmer(text_score ~ bs_pause_frequency_s2 + (1|participant) + (1|prompt), data, REML = F), R^2_m = .04, R^2_c = .68						

S3: Model = lmer(text_score ~ bs_pause_frequency_s3 + (1|participant) + (1|prompt), data, REML = F)						
S4: Model = lmer(text_score ~ bs_pause_frequency_s4 + (1|participant) + (1|prompt), data, REML = F)						
S5: Model = lmer(text_score ~ bs_pause_frequency_s5 + (1|participant) + (1|prompt), data, REML = F)						
Text quality	Total pause duration	1	< -.01	.01	-.70	.48
	Total pause duration	2	-.04	.03	-1.49	.14
	Total pause duration	3	-.05	.05	-1.07	.29
	Total pause duration	4	-.10	.04	-2.04	.02*
	Total pause duration	5	-.04	.02	-1.61	.11
S1: Model = lmer(text_score ~ total_pause_duration_s1 + (1|participant) + (1|prompt), data, REML = F)						
S2: Model = lmer(text_score ~ total_pause_duration_s2 + (1|participant) + (1|prompt), data, REML = F)						
S3: Model = lmer(text_score ~ total_pause_duration_s3 + (1|participant) + (1|prompt), data, REML = F)						
S4: Model = lmer(text_score ~ total_pause_duration_s4 + (1|participant) + (1|prompt), data, REML = F), R^2_m = .02, R^2_c = .67						
S5: Model = lmer(text_score ~ total_pause_duration_s5 + (1|participant) + (1|prompt), data, REML = F)						
Text quality	Contextual revision	1	.27	.16	1.73	.09
	Contextual revision	2	-.11	.12	-.86	.39
	Contextual revision	3	.01	.14	.08	.94
	Contextual revision	4	-.04	.16	-.22	.83
	Contextual revision	5	-.23	.11	-2.02	.05*
S1: Model = lmer(text_score ~ log(contextual_revision_s1) + (1|participant) + (1|prompt), data, REML = F)						
S2: Model = lmer(text_score ~ log(contextual_revision_s2) + (1|participant) + (1|prompt), data, REML = F)						
S3: Model = lmer(text_score ~ log(contextual_revision_s3) + (1|participant) + (1|prompt), data, REML = F)						
S4: Model = lmer(text_score ~ log(contextual_revision_s4) + (1|participant) + (1|prompt), data, REML = F)						
S5: Model = lmer(text_score ~ log(contextual_revision_s5) + (1|participant) + (1|prompt), data, REML = F), R^2_m = .02, R^2_c = .64						

Note: Ww = Within a word, Bw = Between words, Bpc = Between Pinyin and character(s), Bc = Between clauses, Bs = Between sentences, R^2_m = marginal *R*-squared, R^2_c = conditional *R*-squared, * = statistically significant results

The relationships between writing behaviours and text quality across different writing stages were also looked into with genre as the mediator. To address this, genre and its interaction with the writing behaviour index was included as the additional fixed effects in the models. As demonstrated in Table 4.50, five interactions between writing behaviour measures and genre were found in L2 Chinese writing, and six in L1 writing. This means that the way in which the temporal distribution of writing behaviours predicts text quality was dependent to the genre of writing. For non-significant results for interactions, see Tables 19 and 20 in Appendix H.

Table 4.50 *Significant results from linear mixed-effects regressions examining the effects of interaction between writing behaviours and genre by stage on Chinese text quality*

Dependent variable	Predictor	*E*	*SE*	*t*	*p*
L2 Chinese (N =32)					
Text quality (S1)	Bs pause frequency	-.18	.15	-1.24	.22
	Genre (narrative)	-.19	.16	-1.13	.26
	Bs pause frequency * genre (narrative)	.46	.19	2.38	.02
Model = lmer(text_score ~ bs_pause_frequency_s1 * genre + (1|participant) + (1|prompt), data, REML = F)					
Amount of variance explained by the interaction: $R^2_m = .01$, $R^2_c = .01$					
Text quality (S1)	Bc pause duration	.01	.01	1.01	.31
	Genre (narrative)	.44	.11	3.82	< .01
	Bc pause duration * genre (narrative)	-.04	.02	-2.81	.01
Model = lmer(text_score ~ bc_pause_duration_s1 * genre + (1|participant), data, REML = F)					
Amount of variance explained by the interaction: $R^2_m = .01$, $R^2_c = .02$					
Text quality (S1)	Bcr pause duration	.10	.03	2.84	.01
	Genre (narrative)	1.11	.27	4.09	< .01
	Bcr pause duration * genre (narrative)	-.13	.04	-3.00	.01
Model = lmer(text_score ~ bcr_pause_duration_s1 * genre + (1|participant), data, REML = F)					
Amount of variance explained by the interaction: $R^2_m = .08$, $R^2_c = .03$					
Text quality (S5)	Total pause duration	-.04	.02	-2.00	.05
	Genre (narrative)	-.05	.12	-.40	.69
	Total pause duration * genre (narrative)	.04	.02	2.15	.03
Model = lmer(text_score ~ total_pause_duration_s5 * genre + (1|participant), data, REML = F)					
Amount of variance explained by the interaction: $R^2_m = .01$, $R^2_c = .01$					

Text quality (S5)	Bpc pause duration	.06	.03	2.01	.05
	Genre (narrative)	.63	.16	3.94	< .01
	Bpc pause duration * genre (narrative)	-.13	.05	-2.81	.01
Model = lmer(text_score ~ bpc_pause_duration_s5 * genre + (1|participant), data, REML = F)					
Amount of variance explained by the interaction: $R^2_m < .01$, $R^2_c = .01$					
L1 Chinese (N =32)					
Text quality (S1)	Bpc pause frequency	-.10	.29	-.36	.72
	Genre (narrative)	.03	.08	.40	.69
	Bpc pause frequency * genre (narrative)	1.20	.40	3.03	< .01
Model = lmer(text_score ~ bpc_pause_frequency_s1 * genre + (1|participant), data, REML=F)					
Amount of variance explained by the interaction: $R^2_m = .03$, $R^2_c = .03$					
Text quality (S1)	Bcr pause frequency	-.27	.24	-1.12	.26
	Genre (narrative)	.06	.09	.73	.49
	Bcr pause frequency * genre (narrative)	.61	.29	2.09	.04
Model = lmer(text_score ~ bcr_pause_frequency_s1 * genre + (1|participant) + (1|prompt), data, REML = F)					
Amount of variance explained by the interaction: $R^2_m = .02$, $R^2_c = .02$					
Text quality (S1)	Character revision	-.12	.15	-.80	.43
	Genre (narrative)	-.23	.21	-1.07	.29
	Character revision * genre (narrative)	.33	.16	2.02	.05
Model = lmer(text_score ~ log(character_revision_s1) * genre + (1|participant), data, REML = F)					
Amount of variance explained by the interaction: $R^2_m = .01$, $R^2_c = .02$					
Text quality (S2)	Bs pause frequency	.08	.13	.61	.54
	Genre (narrative)	-.11	.14	-.81	.42
	Bs pause frequency * genre (narrative)	.39	.17	2.37	.02
Model = lmer(text_score ~ bs_pause_frequency_s2 * genre + (1+genre|participant), data, REML = F)					
Amount of variance explained by the interaction: $R^2_m = .02$, $R^2_c = .02$					

Text quality (S4)	Bcr pause frequency	-.41	.22	-1.85	.07
	Genre (narrative)	.01	.08	.17	.87
	Bcr pause frequency * genre (narrative)	.88	.28	3.13	< .01
Model = lmer(text_score ~ bcr_pause_frequency_s4 * genre + (1|participant), data, REML = F)					
Amount of variance explained by the interaction: $R^2_m = .03$, $R^2_c = .04$					
Text quality (S5)	Bc pause frequency	.12	.10	1.18	.24
	Genre (narrative)	.36	.11	3.14	< .01
	Bc pause frequency * genre (narrative)	-.29	.14	-2.07	.04
Model = lmer(text_score ~ bc_pause_frequency_s5 * genre + (1+genre|participant), data, REML = F)					
Amount of variance explained by the interaction: $R^2_m = .01$, $R^2_c = .02$					

Note: Bpc = Between Pinyin and character(s), Bc = Between clauses, Bs = Between sentences, Bcr = Between contextual revisions, R^2_m = marginal *R*-squared, R^2_c = conditional *R*-squared

To examine how genre mediates the relationship between behaviours and text quality across stages, linear mixed-effects regression analyses were performed on the stage in which interaction was found to be significant. For all models, the writing behaviour indices reported in Table 4.50 were the fixed effects with the score of the text as the dependent variable. The results are presented in Table 4.51.

In Stage 1, L2 writers who paused more often between sentences (explaining 5% of the variance) produced better narrative essays, whereas those who paused longer between clauses (contributing to 6% of the variance) tended to create poorer narrative texts. It seems that longer pauses between contextual revisions improved argumentative text quality but inhibited narrative text quality in L2 writing. However, the results should be interpreted with caution, as almost half of the L2 writers (13 in argumentative writing and 15 in narrative writing) did not pause between contextual revisions in Stage 1, resulting in 42 out of 64 missing values for this index in argumentative writing and 41 out of 64 in narrative writing. In Stage 5, the quality of L2 argumentative essays benefitted from longer pauses between Pinyin and character(s) (accounting for 5% of the variance), while it was the other way round for narrative essays (explaining 2% of the variance). No significant relationship was found between total pause duration in Stage 5 and text quality in either argumentative or narrative writing, although the previous analysis yielded a statistically significant interaction.

Table 4.51 *Results from linear mixed-effects regressions examining the effects of writing behaviours on text quality where the interaction was significant by genre*

	Argumentative					Narrative				
Dependent variable	Fixed effect	*E*	*SE*	*t*	*p*	Fixed effect	*E*	*SE*	*t*	*p*
L2 Chinese (N = 32)										
Text quality (S1)	Bs pause frequency	-.40	.21	-1.88	.07	Bs pause frequency	.58	.20	2.97	<.01*
	Model = lmer(text_score_arg ~ bs_pause_frequency_s1 + (1|participant) + (1|prompt), data, REML = F)					Model = lmer(text_score_nar ~ bs_pause_frequency_s1 + (1|participant), data, REML = F), $R^2_m = .05$, $R^2_c = .87$				
Text quality (S1)	Bc pause duration	.02	.01	1.67	.11	Bc pause duration	-.05	.02	-3.58	<.01*
	Model = lmer(text_score_arg ~ bc_pause_duration_s1 + (1|participant), data, REML = F)					Model = lmer(text_score_nar ~ bc_pause_duration_s1 + (1|participant), data, REML = F), $R^2_m = .06$, $R^2_c = .89$				
Text quality (S1)	Bcr pause duration	.10	.03	3.72	.03*	Bcr pause duration	-.07	.02	-3.12	<.01*
	Model = lmer(text_score_arg ~ bcr_pause_duration_s1 + (1|participant), data, REML = F), $R^2_m = .07$, $R^2_c = .96$					Model = lmer(text_score_nar ~ bcr_pause_duration_s1 + (1|participant) + (1|prompt), data, REML = F), $R^2_m = .25$, $R^2_c = .95$				
Text quality (S5)	Total pause duration	-.02	.02	-.91	.37	Total pause duration	.01	.02	.42	.68
	Model = lmer(text_score_arg ~ total_pause_duration_s5 + (1|participant), data, REML = F)					Model = lmer(text_score_nar ~ total_pause_duration_s5 + (1|participant) + (1|prompt), data, REML = F)				
Text quality (S5)	Bpc pause duration	.10	.04	2.49	.02*	Bpc pause duration	-.08	.03	-2.60	.02*

	Model = lmer(text_score_arg ~ bpc_pause_duration_s5 + (1|participant), data, REML = F), $R^2_m = .05$, $R^2_c = .87$					Model = lmer(text_score_nar ~ bpc_pause_duration_s5 + (1|participant), data, REML = F), $R^2_m = .02$, $R^2_c = .94$				
L1 Chinese (N = 32)										
Text quality (S1)	Bpc pause frequency	.20	.34	.58	.56	Bpc pause frequency	1.08	.37	2.95	< .01*
	Model = lmer(text_score_arg ~ bpc_pause_frequency_s1 + (1|participant), data, REML = F)					Model = lmer(text_score_nar ~ bpc_pause_frequency_s1 + (1|participant), data, REML = F), $R^2_m = .08$, $R^2_c = .66$				
Text quality (S1)	Bcr pause frequency	-.55	.25	-2.19	.03*	Bcr pause frequency	.16	.26	.60	.55
	Model = lmer(text_score_arg ~ bcr_pause_duration_s1 + (1|participant), data, REML = F), $R^2_m = .05$, $R^2_c = .71$					Model = lmer(text_score_nar ~ bcr_pause_duration_s1 + (1|participant) + (1|prompt), data, REML = F)				
Text quality (S1)	Character revision	-.16	.18	-.86	.39	Character revision	.25	.15	1.70	.10
	Model = lmer(text_score_arg ~ log(character_revision_s1) + (1|participant), data, REML = F)					Model = lmer(text_score_nar ~ log(character_revision_s1) + (1|participant), data, REML = F)				
Text quality (S2)	Bs pause frequency	-.11	.15	-.70	.49	Bs pause frequency	.48	.16	2.99	< .01*
	Model = lmer(text_score_arg ~ bs_pause_frequency_s2 + (1|participant), data, REML = F)					Model = lmer(text_score_nar ~ bs_pause_frequency_s2 + (1|participant), data, REML = F), $R^2_m = .09$, $R^2_c = .70$				
Text quality (S4)	Bcr pause frequency	.01	.17	.07	.95	Bcr pause frequency	-.23	.25	-.92	.36
	Model = lmer(text_score_arg ~ bcr_pause_frequency_s4 + (1|participant), data, REML = F)					Model = lmer(text_score_nar ~ bcr_pause_duration_s4 + (1|participant) + (1|prompt), data, REML = F)				
Text quality (S5)	Bc pause frequency	-.05	.12	-.44	.67	Bc pause frequency	.22	.15	1.45	.15
	Model = lmer(text_score_arg ~ bc_pause_frequency_s5 +					Model = lmer(text_score_nar ~ bc_pause_frequency_s5 +				

(1|participant), data, REML = F)	(1|participant), data, REML = F)

Note: Bpc = Between Pinyin and character(s), Bc = Between clauses, Bs = Between sentences, Bcr = Between contextual revisions, R^2_m = marginal *R*-squared, R^2_c = conditional *R*-squared, * = statistically significant results

In L1 writing, more pauses between Pinyin and character(s) in Stage 1 were beneficial to narrative text quality, while more pauses between contextual revisions were harmful to argumentative text quality. L1 writers who paused more often at the sentence boundary in Stage 2 tended to produce better narrative essays. The amount of variance explained by the writing behaviour index in the three models was 8%, 5% and 9%, respectively. The number of character revisions in Stage 1, the frequency of pause between contextual revisions in Stage 4 and the frequency of pause between clauses in Stage 5 were not predictors of text quality in either genre, although significant interactions were revealed in the previous set of regression analyses.

Lastly, I explored the extent to which Chinese proficiency modulated the relationships between behaviours and text quality in L2 writing. I carried out the same set of analyses as I did with genre as the mediator, but I replaced genre with the centred score of the cloze test. The analyses yielded five significant interactions between writing behaviour measures and L2 proficiency (Table 4.52), with the interaction explaining no more than 3% of the variance in text quality in each model. See Table 21 in Appendix H for non-significant results.

Table 4.52 *Significant results from linear mixed-effects regressions examining the effects of interaction between writing behaviours and proficiency by stage on L2 Chinese text quality (N = 32)*

Dependent variable	Predictor	*E*	*SE*	*t*	*p*
Text quality (S1)	Word-level revision	-.05	.16	-.35	.73
	Proficiency	.05	.02	2.75	.01
	Word-level revision * proficiency	.06	.02	2.81	.01
Model = lmer(text_score ~ log(word_revision_s1) * cproficiency + (1|participant), data, REML = F)					

Amount of variance explained by the interaction: $R^2_m = .03$, $R^2_c = .01$					
Text quality (S3)	Bs pause duration	.01	< .01	1.53	.13
	Proficiency	.07	.02	4.84	< .01
	Bs pause duration * proficiency	< .01	< .01	2.08	.04
Model = lmer(text_score ~ bs_pause_duration_s3 * cproficiency + (1|participant), data, REML = F)					
Amount of variance explained by the interaction: $R^2_m = .01$, $R^2_c = .01$					
Text quality (S4)	Below-clause revision	.05	.17	.28	.78
	Proficiency	.06	.02	3.28	<.01
	Below-clause revision * proficiency	.05	.02	2.07	.04
Model = lmer(text_score ~ log(below_clause_revision_s4) * cproficiency + (1|participant) + (1|prompt), data, REML = F)					
Amount of variance explained by the interaction: $R^2_m = .02$, $R^2_c = .01$					
Text quality (S4)	Clause-and-above revision	-.24	.25	-.93	.36
	Proficiency	.07	.01	5.08	<.01
	Clause-and-above revision * proficiency	.09	.04	2.21	.03
Model = lmer(text_score ~ log(clause_and_above_revision_s4) * cproficiency + (1|participant), data, REML = F)					
Amount of variance explained by the interaction: $R^2_m = .01$, $R^2_c = .01$					
Text quality (S5)	Total pause frequency	-.05	.04	-1.33	.19
	Proficiency	.13	.02	5.50	< .01
	Total pause frequency * proficiency	-.01	.01	-2.46	.02
Model = lmer(text_score ~ total_pause_frequency_s5 * cproficiency + (1|participant), data, REML = F)					
Amount of variance explained by the interaction: $R^2_m < .01$, $R^2_c = .01$					

Note: Bs = Between sentences, R^2_m = marginal *R*-squared, R^2_c = conditional *R*-squared

Specifically, more word-level revisions in Stage 1 resulted in better texts for L2 writers scoring above 28 out of 45 on the cloze test. Longer between-sentence pauses in Stage 3 positively correlated with text quality only when L2 writers gained over 22 on the test. In Stage 4, more below-clause revisions had a positive effect on text quality for those who reached over 28, and more frequent revisions at the clause level and above led to better essays under the condition that the writers scored over 30. In Stage 5, more pauses overall were beneficial to text quality for those who gained 32 or below.

Summary: relationships between writing behaviours and text quality Six writing behaviour indices, i.e. production rate, P-burst length, between-word pause frequency, total pause duration, between-word pause duration and between-clause pause duration, were identified as predictors of L2 Chinese text quality. Specifically, better L2 texts were linked to greater speed fluency, fewer between-word pauses, and shorter pauses overall, between words and between clauses. In addition, between-sentence pause frequency, between-word pause duration and the amount of word-level revision were found to predict text quality in different ways for more and less proficient L2 writers. Different from L2 writing, no writing behaviour index was found to predict L1 Chinese text quality. The genre of writing tended to mediate the relationship between the frequency of pause between contextual revisions and L1 text quality.

When the whole writing session was divided into five stages, the way in which writing behaviours predict text quality changed across stages. In L2 writing, more within-word pauses in Stage 1 enhanced text quality, while more between-word pauses in Stage 2 inhibited text quality. Longer pauses between Pinyin and character(s) were beneficial to the quality of the text in Stage 2 but became harmful in Stage 3. The duration of between-clause pauses in Stage 5 negatively correlated with text quality. In L1 writing, more between-sentence pauses in Stage 2 improved text quality, while longer pauses overall in Stage 4 reduced text quality. In addition, there was a negative correlation between contextual revision in Stage 5 and text quality.

Genre was found to modulate the relationships between writing behaviours and text quality across the five stages. L2 narrative text quality benefitted from more between-sentence pauses and shorter between-clause pauses in Stage 1, and shorter pauses between Pinyin and character(s) in Stage 5. Better L2 argumentative essays were associated with longer pauses between Pinyin and character(s) in Stage 5. As for L1 writing, pause between Pinyin and character(s) in Stage 1 positively correlated with narrative text quality, while pause between contextual revisions in Stage 1 negatively correlated with argumentative text quality. In addition, more pauses at the sentence boundary in Stage 2 led to narrative texts of higher quality.

Finally, the strength of the relationship in L2 writing could be mediated by writers' proficiency in L2 Chinese. That is, more word-level revisions in Stage 1, longer between-sentence pauses in Stage 3, more revisions above the word level in Stage 4 and fewer pauses overall in Stage 5 were likely to benefit text quality only when writers' L2 proficiency reached a certain level.

Chapter 5 Discussion

Chapter 5 discusses the results obtained in the current study in relation to the research questions outlined in Chapter 2. First, the findings for the role of writing stages in the cognitive processes of L2 Chinese writing (i.e. Research Question 1) are discussed. This is followed by a discussion of the effect of genre on the cognitive processes in L2 Chinese writing (i.e. Research Question 2), and the relationships between L2 proficiency and the behaviours in which L2 Chinese writers engage (i.e. Research Question 3). The final section of this chapter discusses the findings on the way in which writing behaviours predict L2 Chinese text quality (i.e. Research Question 4). The discussion of the findings concerning the cognitive processes in L1 Chinese writing in relation to stages and genre, and the links between writing behaviours and L1 Chinese text quality (i.e. Research Question 5), is incorporated into corresponding parts when the findings obtained for L2 Chinese are discussed.

5.1 Stages of writing and cognitive processes in Chinese writing

This section discusses the findings for the extent to which stages affect cognitive processes in L2 Chinese writing, as reflected in the behaviours captured by keystroke logging and the associated cognitive activities reported in the stimulated recall interviews. It additionally considers the role of stages in the cognitive processes involved in L1 Chinese writing.

Linear mixed-effects regressions and qualitative analyses revealed an effect of stages in the cognitive processes for both L1 and L2 Chinese writing. In general, the majority of the differences across stages were found between the beginning and/or end, and the middle of writing. According to Kellogg's (1996) model, cognitive activities involved in text production, such as planning, translation and monitoring, draws on a limited pool of cognitive rescources. Thus, the variances in the distribution of writing behaviours and associated cognitive activities across the five stages tend to indicate different allocation of attentional resources at the beginning, middle and end of text production in Chinese.

The results also well illustrate the dynamic nature of writing, that is, different writing stages tend to feature different characteristics of the cognitive processes in which the writers engage. Writing behaviours, such as speed fluency, pausing and revision, do not have equal probabilities of being activated at any time during text production, and the cognitive activities associated with these behaviours are not equally likely to be seen in the various stages of writing either. The findings constitute supportive evidence for the temporal model of writing proposed by Rijlaarsdam and Van Den Bergh (1996), which points out that certain writing activities might have higher probabilities of occurrence at certain moments due to the combined effects of internal (i.e. writer's default parameter settings) and external factors (i.e. changing task environment). It indicates that text production is not 'a purely recursive process' (Roca de Larios et al., 2008, p. 43). In other words, writing neither consists of a set of pre-established activities which progress in a linear way, nor should it be perceived as a process in which the writer allocates similar amounts of time to activities across different stages (Roca de Larios et al., 2008). The temporal model of writing has been empirically tested in a few earlier writing studies (see Breetvelt et al., 1994; Kellogg, 1988; Tillema et al., 2011 for L1 writing; see Barkaoui, 2015, 2016, 2019; Khuder & Harwood, 2015; Manchón et al., 2009; Tillema, 2012; Van Weijen, 2008; Xu, 2018; Xu & Qi, 2017 for L2 writing), and the results of the current study further confirm that writing is a dynamic process by expanding on the previous findings obtained for alphabetic language writers in the context of non-alphabetic language writing. In the rest of this section, specific trends in relation to speed fluency, pausing, revision and the cognitive activities associated with behaviours are discussed.

5.1.1 Stages of writing and speed fluency

In terms of the effect of stages on speed fluency, L2 writers of Chinese appeared to be more fluent, as measured by production rate and P-burst length, at the beginning of writing but less so at the end. A similar finding was obtained for L1 Chinese writers, in that Stages 1 and 5 were characterised by the highest and lowest production rates, respectively. These findings are largely in line with Van Waes and Leijten (2015), in which the variability of fluency was observed throughout the text production process, with the highest production rate in the initial stages.

One possible reason for the comparably greater production rate in Stage 1 may be associated with the relatively larger amount of time spent pausing at the beginning of writing. Before starting to type text, the writers might have to spend some time reading the writing prompt, generating ideas, considering the general structure of the text and transforming their ideas into linguistic forms for the first few utterances. As a result, the actual time they spent typing the text could be very limited at the beginning. In addition, as the first several utterances of the text tended to be pre-planned and rehearsed during the pausing period, the writers were quite unlikely to pause and/or revise often when producing this part of the text. Therefore, the limited time spent typing the text, together with a relatively smooth execution process (i.e. typing the text) without being frequently interrupted by pauses and/or revisions jointly contributed to the greater production rate (and longer P-bursts in the case of L2 Chinese writing) in the initial stage of writing. With regard to the reduced fluency at the end of writing, we suspect that in Stage 5, most writers might have finished writing the text and were focused more on re-reading and monitoring, which was reflected in a noticeable increase in the proportion of monitoring activities in the verbal protocols in the final stage. Consequently, relatively fewer characters were produced in this stage as compared to the middle stages, leading to a lower production rate (and shorter P-bursts in the case of L2 Chinese writing) in the last writing stage.

5.1.2 Stages of writing, pausing behaviours and associated cognitive activities

Turning to pausing behaviours in L2 Chinese writing, the differences mainly set apart the middle stages from the initial and/or final stages. Overall, L2 writers of Chinese paused less frequently in Stages 1 than in Stages 2, 3 and 4, replicating the findings of Barkaoui (2019) and Xu and Qi (2017). I additionally found fewer but longer pauses in Stage 5 than in the middle stages, which was not reported by either Barkaoui (2019) or Xu and Qi (2017). When pauses were analysed in terms of location, the effect of stages seemed to be different for pauses at smaller and larger textual unit boundaries. In terms of frequency, pauses between smaller textual units occurred less often only in Stage 5, while pauses between larger textual units occurred less frequently in both Stages 1 and 5. In terms of duration, stages of writing appeared to have no influence on how long L2 writers paused at smaller textual unit boundaries. However, longer between-sentence pauses were found in Stage 1 than in the middle writing stages. It is worth noticing that L1 Chinese writing exhibited very similar distribution patterns in terms of the frequency and duration of pauses overall and at various locations across the five stages.

The different effect of stages on pauses at various locations may be attributed to the differential cognitive activities that underlie these pauses (Schilperoord, 1996). First of all, as evidenced in the stimulated recall comments reported by L2 writers of Chinese, pauses at larger textual unit boundaries, especially those between sentences, were primarily linked to planning processes in writing. Therefore, the less frequent occurrences of pauses overall, between clauses and between sentences, as well as longer between-sentence pauses in Stage 1, may serve as an indicator of more attention being devoted to planning at the beginning of L2 Chinese writing. This explanation is supported by the thoughts that L2 writers reported in stimulated recall when pausing in Stage 1, as shown in Examples 1 and 2. Thus, the findings here align with the conclusion reached by researchers studying alphabetic language writers, i.e. the initial stages of writing are dominated by planning activities (Barkaoui, 2015; Khuder & Harwood, 2015; Roca de Larios et al., 2008; Tillema, 2012; Van Weijen, 2008).

Example 1:

Participant 1: I was trying to decide what story to write about. Err…I was thinking of, so like write about playing the piano. I was also thinking of other stories to write, backing up three or four. (a between-sentence pause of 86.52 seconds)

Example 2:

Participant 8: I was just thinking in general whether mobile phones are good or not, and then I was thinking what side to take. (a between-sentence pause of 9.91 seconds)

Secondly, as L2 writers of Chinese made more references to translation processes when pausing between lower-level textual units, the drop in the frequency of pauses between Pinyin and character(s), between words, and between clauses in Stage 5 may indicate the end of the drafting period, in which the writers were preoccupied with text formulation, and mark the start of a systematic reviewing period. This received further confirmation from the increased number of pauses between contextual revisions in the final writing stage, which was mainly associated with re-reading for monitoring, as revealed in the verbal protocol data (Examples 3 and 4).

Example 3:

Participant 6: I decided that I didn't like the sentence structure. That's why I added a 来 (in order to). And then I was just reading along the whole thing. (re-reading for monitoring after a contextual revision to syntactic structure)

Example 4:

Participant 18: Because I thought 发生 (take place) was the wrong word, I wanted to say 'grow up' into that type of person. After I made the change, I continued reading through the text. (re-reading for monitoring after a contextual revision to vocabulary use)

Interestingly, although similarities between L1 and L2 Chinese writing were observed in the behavioural data, the cognitive activities in which L1 and L2 Chinese writers engaged when pausing were, to some extent, different. Unlike in L1 writing, in which planning took up around 45% of the time before coming to a noticeable decrease in the last stage, the proportion of planning activities in L2 writing dropped considerably after Stage 1. The difference may be explained in light of Kellogg's (1996) model of writing. As anticipated by the model, due to attentional limitations, some cognitive activities would be postponed or even sacrificed to accommodate the immediate call for other activities. Compared to L1 writers, L2 writers with limited proficiency in Chinese are more likely to encounter linguistic problems during text composition (Murphy & Roca de Larios, 2010; Tillema, 2012). As the writers in the current study were required to produce a complete essay in 30 minutes, L2 writers might naturally and unconsciously direct more attention to translation processes, such as retrieving, encoding and/or (internally) revising lower-level textual units, in order to overcome language difficulties and produce a text meaningful to the reader within the time limit. Given limited capacity of writers' working memory (Kellogg, 1996), increased effort involved in translation operations would result in less cognitive resources available for higher-order activities, such as generating ideas. Consequently, when L2 writers had a rough idea of what to write and how to organise the content, planning activities reduced considerably after the beginning of writing. In contrast, L1 writers who are mostly free from linguistic barriers would experience much less pressure on linguistic encoding. Thus, the decreased cognitive demands on translation processes allows more attentional resources to be allocated to planning processes, which is manifested in L1 writers' constant engagement in planning before the final writing stage.

Another possible explanation for the difference between L1 and L2 writing in terms of pause-related cognitive activities may be that L2 writers consciously adopted a decomposition strategy to prevent cognitive overload (Broekkamp & Van Den Bergh, 1996), as activities involved in text production are carried out within a limited pool of cognitive resources (Kellogg, 1996). A decomposition strategy involves breaking down the text production process into smaller components, and it requires writers to concentrate on one or a limited number of components at a time in order to minimise the chance of attention conflicts. In other words, L2 writers may intentionally (but not completely) segregate the processes of planning, translation and monitoring, so as to allow a certain process to be maximally activated and spread over a certain stage with minimal interruption from other processes. Guided by this strategy, it is possible that L2 writers intentionally separated the processes by carrying out most of the planning activities in Stage 1 while primarily focusing on translation in subsequent stages so as to avoid being overwhelmed by constant switches between the two activities and/or simultaneous activations of the two. L1 writers, however, are highly unlikely to experience attention conflicts among ideas, organisation and language, as all the topics for writing in the current study were about daily issues that the writers were supposed to be familiar with. Therefore, the chances are low for L1 writers to utilise a decomposition strategy to segregate planning and translation processes, as done by L2 writers.

5.1.3 Stages of writing, revision behaviours and associated cognitive activities

As regards revision behaviours in L2 Chinese writing, the majority of the differences were, once again, found between Stage 1 and/or Stage 5 and the middle writing stages. The initial stage of writing featured fewer revisions in total, which is consistent with the results of Barkaoui (2016) and Roca de Larios et al. (2008). In addition, L1 writers in the current study also made fewer revisions at the beginning, suggesting that such a pattern is not unique in L2 writing. Similar to that for pausing behaviours, fewer revisions in Stage 1 may be explained by a dominant focus on planning at the beginning, which resulted in less text production and thus fewer revisions. This was supported by some evidence from the verbal protocols, in which participants made proportionally more references to ideas and organisation, despite a smaller number of revisions in total in Stage 1 than in other stages.

I also found that the total revision amount dropped at the end of writing, which runs counter to Roca de Larios et al. (2008), who reported a gradual increase in revisions after the initial stage. This discrepancy may be explained by the different writing modes that the participants used for task performance, i.e. typing in the current study versus handwriting in Roca de Larios et al. (2008). Researchers in the field of L1 writing (e.g. Van Waes & Schellens, 2003) have found that writing processes can be affected by the mode of writing. With regard to revision behaviours, L1 writers, when writing with paper and pen, prefer to revise the text systematically in the later stages. However, when the writing task is performed on a computer, L1 writers are more likely to carry out revisions evenly throughout the drafting period. The findings for L1 computer writers in the previous literature also largely parallel the observation for L1 Chinese writers. Assuming that the same holds for L2 writers, it can be speculated that the decrease in total revision frequency in Stage 5 may be due to the preference that the writers have when producing essays on a computer.

A second explanation for the reduction of revisions in the final stage could be that L2 writers of Chinese were reluctant to revise extensively when reading through their own texts. As shown in the data, pre-contextual revisions, which occur at the point of inscription, reduced considerably in Stage 5, signalling the end of the drafting period in which L2 writers mostly engaged in text production. Increased attention was then allocated to re-reading for monitoring, which is evidenced in the pausing and stimulated recall data, as mentioned earlier. However, the monitoring activity did not seem to be very productive (Khuder & Harwood, 2015). In other words, although re-reading the texts did indeed lead to more contextual revisions in Stage 5, L2 writers did not revise already written text on a large scale (contextual revisions only accounting for 24% of total revisions in the final stage). This observation is in line with the speed fluency and stimulated recall data. In Stage 5, as compared to previous stages, L2 writers produced shorter P-bursts but reported a considerably larger proportion of revisions associated with translation.

The reluctance to do extensive revision towards the end of writing may be associated with difficulty in character recognition (Gunderson et al., 2011). As a result, L2 writers of Chinese may be extremely slow in reading texts written in Chinese characters, even though the texts were written by them. Under time pressure, they might have prioritised correcting a few obvious language mistakes, which was less time-consuming and (they believed) would improve the form and mechanics of the text, so as to be on the 'safe side' (Crawford et al., 2008). This is well demonstrated by two episodes (Examples 5 and 6) from the verbal protocols by Participants 2 and 13 at the end of the drafting period. It seems that due to the time limit for task completion, both writers had a clear goal of what to revise in the systematic reviewing period, i.e. focusing on minor changes without affecting the content of the text. Indeed, 17 out of 20 revisions that Participant 2 carried out when re-reading the text were language-oriented, which simply involved adding, deleting or replacing a few characters, whereas Participant 13 only changed one character while reading through the essay.

Example 5:

Participant 2: This was the last sentence I wrote. I was then reading the text through, and this was the point when you said ten minutes left. You will see I only made small changes here and there, mainly to the language.

Example 6:

Participant 13: I read back a bit, and there was not much time left, because for me to check properly, I would need more than the last several minutes. So I was just looking for the obvious mistakes that I always made for the characters.

Another intriguing finding that emerged from the data is that although L2 writers made proportionally more revisions associated with translation processes in Stage 5, L1 writers revised the language more often in Stage 1. More translation-related revisions in Stage 5 of L2 writing may be associated with the writers' previous learning and assessment experience, which made them believe that linguistic accuracy was valued most by the potential readers of their L2 texts (G. Porte, 1996). With such a belief in mind, L2 writers might have focused predominantly on language use when re-reading and editing previously-written texts at the end of writing in order to improve accuracy. This finding is, to some extent, consistent with those of Barkaoui (2016), i.e. that the majority of contextual revisions are language-oriented and occur in the final stage of writing.

In terms of L1 writing, a higher percentage of revisions to language in Stage 1 might be attributed to a belief that L1 writers hold for a good Chinese text. It is traditionally believed that good Chinese writing features the head of a phoenix (Tao, 2012), meaning that the beginning of an essay should be appealing. One way to achieve this is to manipulate the language in the opening paragraph. As demonstrated in Examples 7–9, it seems that when writing the first paragraph of the essay, L1 writers were indeed aware of the 'phoenix head' metaphor and revised the language intentionally for a more impressive beginning. For instance, Participant 4 made two word-level revisions because he wanted to start the essay with a parallel structure containing three similes embedded with idioms; Participant 6 changed a phrase in the first paragraph from four characters to three for a neater style; Participant 20 replaced a normal adjective with an idiom (*天真无邪*, innocent and pure) for a more attractive opening sentence.

Example 7:

Participant 4: I decided to write three clauses to describe childhood using similes. So here, I deleted *往事* (things that happened in the past) while I kept *童年是* (childhood is). (a word-level revision)

Then to make the second clause run in parallel with the first one, I started with *童年也是* (childhood is also). I wanted to use seashells to describe childhood, but it was too plain to just write seashells, so I added an idiom *五彩缤纷* (colourful) to emphasise how beautiful childhood is. (a word-level revision)

Example 8:

Participant 6: I wanted to use *上网冲浪* (surf the Internet) initially, but then it didn't sound right. The previous phrase was in three Chinese characters, and if here I used a four-character phrase, they wouldn't match. (a below clause revision)

Example 9:

Participant 20: I knew people always say childhood is *重要的* (important) or *美好的* (beautiful), but I didn't want to use either of the adjectives because they were clichés. I wanted to use a better adjective to describe childhood. (a word-level revision)

5.1.4 Summary of the section

Taking the behavioural and verbal protocol data together, L2 writers of Chinese were found to allot more attention to planning at the beginning of writing, engage more in translation in the middle, and carry out more monitoring activities at the end. These trends are consistent with some patterns reported in previous studies on L2 English writers (Barkaoui, 2016, 2019; Roca de Larios et al., 2008; Tillema, 2012; Van Weijen, 2008). However, running counter to the findings of some earlier studies (e.g. Roca de Larios et al., 2008), fewer revisions were observed towards the end of L2 Chinese writing. This discrepancy may be attributed to the mode of writing and the unique orthographic system of the Chinese language.

In addition, the data also revealed some differences between L1 and L2 Chinese writers in terms of the cognitive activities associated with writing behaviours. The differences may be explained by the number of language resources available for writing. A limited language proficiency reduced the extent to which L2 writers engaged in higher-order planning processes after the initial writing stage, while sufficient linguistic resources allowed L1 writers to constantly allocate a certain amount of attention to planning throughout text production. Writers believed that a good piece of writing could also result in different cognitive processes being involved in text composition. L2 writers prioritised linguistic accuracy of the text, leading to a greater proportion of language-oriented revisions in the final writing stage, whereas L1 writers, who were aware of the importance of a well-written opening in a Chinese essay, carried out more language revisions at the beginning.

5.2 Genre of writing and cognitive processes in Chinese writing

In this section, I discuss the effect of genre (i.e. argumentative versus narrative) on the cognitive processes in which L2 writers of Chinese engage. In addition, the findings on L1 writers are combined in the discussion for a fuller picture of the genre effect on the writing processes in Chinese.

5.2.1 Genre and cognitive processes: the whole writing session

Linear mixed-effects regression analyses revealed a genre effect on three behavioural indices, i.e. production rate, P-burst length and within-word pause frequency, in L2 Chinese writing. More specifically, L2 writers of Chinese produced more characters per minute and longer P-bursts, and they paused more often on a word when writing argumentative than narrative essays. No genre difference was found in the rest of the pausing and revision measures. Largely similar findings were obtained from L1 Chinese writing, in that L1 writers showed a greater production rate and made more frequent pauses at word boundaries when producing argumentative rather than narrative essays.

Our results for pausing and revision measures are partially in line with some patterns observed in previous studies. Similar to the findings for L1 Chinese writers, Medimorec and Risko (2017) reported that L1 writers of English paused more frequently between words in argumentative than narrative writing, indicating prolonged lower-order processes, such as choosing less frequently used lexical items, when writers were performing an argumentative writing task. It appears that a deeper engagement in lower-level processes was also required in argumentative writing in L2

Chinese, reflected in more frequent within-word pauses in the argumentative than the narrative task. The effect of genre on lower-level processes of writing may be because of the different communicative goals of argumentative and narrative texts (Biber & Conrad, 2009), argumentative pieces generally feature greater linguistic complexity than narrative pieces (Beers & Nagy, 2009; Biber & Conrad, 2009; Lu, 2011; Way et al., 2000; Yoon & Polio, 2016). In order to meet this genre requirement for argumentative writing, writers may allocate more attentional resources to lower-level writing processes, which are normally expressed as pauses at smaller textual unit boundaries, such as within a word or between words. In addition, no genre effect was detected in the frequency of total and contextual revisions in L2 Chinese writing, mirroring what Thorson (2000) found for L2 writers of German. A lack of genre effect on revision frequencies was also found in L1 Chinese writing, contrary to the result of Thorson (2000) for L1 English writing where genre had an impact on total revision amount. However, I suspected that the different amount of total revision in L1 writing reported by Thorson (2000) might not be attributable to genre alone. For example, the amount of time available for task completion may be one possible factor that contributed to the difference. Unlike in the current study where the same time limit was set for each writing task, the participants in Thorson (2000) were given 30 minutes to write a letter, but 40 minutes to complete a newspaper article. The extra 10 minutes might increase the chances of revisions when the participants were drafting the newspaper article.

It is also worth noting that, slightly different from L1 writers, L2 writers of

Chinese paused more often within a word rather than between words when performing argumentative rather than narrative writing tasks. A plausible reason for this small difference may be related to a stylistic feature of formal (i.e. written) Chinese language. Generally speaking, formal Chinese is characterised by the use of a set of special monosyllabic words (i.e. words from ancient Chinese that contain one Chinese character) in a disyllabic template and more frequent use of disyllabic words (i.e. words that are made up with two Chinese characters) as compared to informal (i.e. spoken) Chinese (Feng, 2006). Given that the argumentative genre is considered more 'formal' than the narrative genre, writers may intentionally use more words in disyllabic form when producing argumentative essays to make the text appear more formal. In fact, 54% of the words that the L2 writers used in the argumentative texts were in disyllabic form, compared to 45% in narrative writing in the current study. As these disyllabic words are more abstract and less frequently used than normal monosyllabic words, it is possible that when composing argumentative essays, L2 writers of Chinese had to pause more often within these disyllabic words (i.e. between two characters) to recall the second halves of words, since these words may be stored as individual characters rather than as whole units in their long-term memory. As a consequence, the retrieval of these disyllabic words might not be fully automated, leading to an increase in the pause frequency within a word (i.e. between two characters) in argumentative writing. In contrast, L1 writers of Chinese, whose retrieval process for these disyllabic words is as fully automated as that for other words, are less likely to pause more frequently within a word in an argumentative than a narrative task.

One surprising finding that emerged from the data was that both L1 and L2 writers of Chinese showed greater speed fluency in argumentative than narrative writing, although narrative writing is supposed to be less cognitively demanding, as this genre requires more knowledge-telling and fewer knowledge-transforming strategies (Bereiter & Scardamalia, 1987). One explanation for this result may be that even though the participants, as adults, had already established a well-developed schema for narratives (Bouwer et al., 2014), they had not had as many opportunities in real life to write narrative essays as opposed to argumentative essays. A lack of practice can, to some extent, defamiliarise adult writers with personal narratives, resulting in a lower production rate as compared to argumentative writing. I additionally found that L2 writers produced P-bursts of greater length in the argumentative than the narrative task. This may be related to L2 writers' greater reliance on formulaic sequences that are commonly used in argumentative essays. For example, some formulaic sequences for expressing ideas and providing evidence, such as *我觉得/认为* (I think), *比如说* (for instance), *对…来说* (in terms of), *在…看来* (in one's opinion) and *总的来说* (to sum up), were found to be used more than twice as often in argumentative essays (9.17 occurrences per 1,000 characters) than in narratives (4.27 occurrences per 1,000 characters).

Analyses of stimulated recall comments showed that L2 writers of Chinese used proportionally more pauses for translation in argumentative than in narrative writing, while they allocated relatively more pauses to planning and monitoring in narrative than in argumentative writing. These results support the findings from behavioural data that argumentative writing in L2 places a greater demand on lower-level processes that are typically associated with language use rather than narrative writing. Surprisingly, however, when it came to revision, L2 writers reported relatively more revisions triggered by planning processes when performing an argumentative rather than a narrative task, while they made proportionately fewer references to revision for translation in the argumentative than the narrative task. I suspect that this 'unexpected' finding may have to do with the stimulus that we used for eliciting L2 writers' cognitive activities underlying writing behaviours. The stimulus in the current study was a recording of the writer's writing performance at the character level. In other words, due to technical limitations, the writers were unable to see how characters were typed via the Pinyin input method before they appeared on screen. The lack of the stimulus at the Pinyin level is very likely to reduce the chances for writers to report their thoughts associated with revisions at lower linguistic unit levels, which potentially entail changes made to Pinyin. Since revisions at lower linguistic unit levels tend to be linked more often with translation processes, such as lexical retrieval or morphosyntactic encoding (Révész et al., 2017, 2019), it is possible that L2 writers in the present study might actually make more translation-related revisions when producing argumentative than narrative essays, but failed to report this owing to the lack of stimulus. Different from what we found for L2 writers, the verbal protocol data elicited from L1 Chinese writers did not demonstrate much difference between argumentative and narrative writing in terms of the cognitive activities associated with behaviours. These findings, however, are consistent with some previous work on L1 writers of English in that writers tend to adopt similar strategies (i.e. planning content and organisation, reading back for monitoring) when producing essays in different genres (e.g. Haas, 1989; Van Waes & Schellens, 2003).

5.2.2 Genre and cognitive processes: across the five writing stages

The results obtained from linear mixed-effects regression and post hoc analyses showed that L2 Chinese writers' behaviours were considerably more varied across the five stages in narrative (36 differences) than argumentative writing (six differences). For both genres, stages of writing influenced the frequencies of pauses overall, pre-contextual revision and revision to Pinyin. For narrative writing, stages of writing additionally had an impact on the frequencies of pauses between Pinyin and character(s) and between clauses, as well as the amount of revision in total, below the word level, below the clause level, occurring at the point of inscription and made to Pinyin. The differences observed in argumentative writing set Stage 5 apart from the middle writing stage(s), while those in narrative writing were mostly seen between Stage 1 and/or Stage 5 and the middle stages. The differences between argumentative and narrative writing in the distribution of writing behaviours across the five stages lend some support to an implication inferred from Rijlaarsdam and Van Den Bergh's (1996) theoretical framework that the time-related features of writing are expected to be modulated by external factors, such as the genre of writing.

For argumentative writing, Stage 5 was characterised by less frequent total and between-clause pauses, and fewer pre-contextual and Pinyin revisions than in the middle stage(s). According to stimulated recall comments in relation to pausing behaviours, there was a considerable decrease in terms of the percentage of comments describing translation processes in the final stage, while the proportion of monitoring-related comments displayed a large increase. Triangulating the two data sources, it seems that L2 writers almost completed drafting the essay before Stage 5, and then devoted more time re-reading and monitoring in the final stage. These results are consistent with some observations from earlier studies. Similar to the findings of the current study, one result of Barkaoui (2016) showed that pre-contextual revisions decreased in the final stage of writing. Previous literature also shows that more formulation activities were carried out in the middle of writing (Roca de Larios et al., 2008; Van Weijen, 2008), and more monitoring activities occurred at the end (Gánem-Gutiérrez & Gilmore, 2018; Khuder & Harwood, 2015; Tillema, 2012).

For narrative writing, Stage 1 (and Stage 5) featured fewer pauses and revisions than the middle. The percentage of stimulated recall comments mentioning planning processes dropped drastically after Stage 1, while the proportion of those associated with monitoring started to increase after Stage 2. Data from both sources indicated that when writing narrative essays, L2 writers may try hard to decide what story to write at the beginning, while being concerned much less with the content in subsequent stages once the decision was made. Although the majority of monitoring activities were carried out in Stage 5, L2 writers tended to allocate a certain amount of time to reading throughout the text for monitoring in Stages 3 and 4. This may be due to the relatively lower cognitive load on translation processes in narrative writing, which frees cognitive resources for higher-order activities.

It is worth highlighting that in this study, no difference was observed between Stage 1 and the middle stages in L2 writers' behaviours during the performance of argumentative tasks. The stimulated recall comments associated with pausing and revision also revealed relatively small changes in the percentage of planning activities from Stage 1 to Stage 4 (i.e. 28%, 26%, 25% and 24% for the first four stages, respectively). These findings do not fully align with the patterns reported by previous studies that also elicited participants' writing processes with argumentative tasks. They seem to uniformly suggest that significantly more planning occurred in the first than the sequential stages of argumentative writing, with evidence from either behavioural or verbal protocol data (Barkaoui, 2015, 2019; Khuder & Harwood, 2015; Roca de Larios et al., 2008; Tillema, 2012; Van Weijen, 2008; Xu & Qi, 2017), while we found more planning at the beginning of L2 narrative writing only. The similar amount of planning between the initial and middle stages in argumentative writing in the current study may be explained by the extent to which the writers were familiar with the genre. As mentioned earlier, adult L2 writers were perhaps more familiar with argumentative than narrative writing; thus, it might have been less time-consuming for the participants to generate arguments than to think of a personal story.

In addition, the writing prompts of this study may be another possible explanation for the lack of difference between stages. In the argumentative writing prompts, two different views on a controversial topic were given, and the writers were asked to discuss both views and provide their own opinions. Such prompts may limit the degree of freedom in terms of the content and structure of the argumentative text. Given the relatively fixed text format, L2 writers might not have had to engage extensively in planning activities in the initial stage when writing argumentative essays. Instead, they might have spent some time thinking about what ideas to include in the first section (e.g. a discussion of the first view), draft it and then move on planning the second section, as these sections are relatively independent.

For example, when commenting on the processes involved in writing an argumentative essay on the extensive use of mobile phones, Participant 22 mentioned that, at the beginning of writing, she did not spend much time planning in detail for the whole text, because the writing prompt had provided clear guidance on the main ideas to include in each part of the essay (Example 10). The participant also reported later that, during text production, she would pause to generate ideas every time when moving to a new paragraph. The three episodes in Example 11 demonstrate the participant's thoughts before drafting the second, third and last paragraphs of the essay. Taking Examples 10 and 11 together, it seems that because of the presentation of the argumentative writing prompt, Participant 22 tended to carry out planning activities paragraph by paragraph, rather than making a full plan at the very start. This resulted in a similar amount of engagement in planning (around 25%) from Stage 1 to Stage 4 in her production of the argumentative essay.

Example 10:

I was planning what I was going to say in general in the essay. It was quite easy because what you have to do were the title, the beginning, the for and against, and then the summary. The prompt had already made it quite obvious. Then I was thinking how to write a short opening paragraph.

Example 11:

#1: I was thinking what I was going to put in the next paragraph. What was the first advantage of having a phone? What examples shall I use to support it?

#2: It was the third paragraph. I thought I was running out of time, so I decided to go for the biggest problem of mobile phones. I thought it should be that people were addicted to phones.

#3: I was thinking what to write in the next paragraph, basically about my opinion towards the massive use of mobile phones. I wanted to make a point that the benefits of the phone outweigh the drawbacks.

However, when it comes to narrative writing, the prompts asked the writers to tell a story with details included, such as when and where the event took place, who was involved and how they felt about the event. It could take some time for the participant to decide what event to write about and recall the details before drafting the text, as illustrated in Example 12. Sometimes, the participants would abandon the story that they initially planned to write shortly after typing the first few utterances. Then, they had to engage in another round of generating ideas (Example 13), which additionally increased the percentage of planning time in Stage 1.

Example 12:

Participant 19: I definitely started to gather my thoughts a bit first since it was talking about a specific experience that I had. It took me a while to pick one, the one I felt like I would write about. Even when I settled on the experience I wanted to write about, I was still thinking, ok, now where I shall begin and what I wanted to include for a bit.

Example 13:

Participant 17: I was trying to think of something that I could describe as a first-time experience. I struggled a bit for something that would be suitable. My thoughts went from the first cigarette, the first time I had sex, the first time I had alcohol, but I couldn't remember when it was. Nothing seemed suitable. And then, eventually, I decided that I would write the first time I studied Taichi. (typed the first sentence)

I was questioning whether this was a good topic. (after typing the first sentence)

Then I deleted all of this 'cause I decided it was not a good topic. (deleted the first sentence)

I didn't like the topic, but what else could I talk about? I was trying to think of that. I just came up with a potential new topic. And I was wondering whether it was good and how I would develop the ideas. (the second round of content planning)

It is also worthwhile discussing the different effects of genre on L1 and L2 writing as a function of stages. Although narrative writing (36 differences) in L1 Chinese also revealed more diverse behaviours across stages than argumentative writing (seven differences), the indices that were affected by stages were not the same as those in L2 writing. For both genres, an effect of stages was found on between-sentence pause duration and total revision frequency. For narrative writing, stage effects, in addition, were observed in the frequencies of pause overall and between clauses, total pause duration, and the number of below-word, pre-contextual, Pinyin and character revisions. The majority of the differences were observed between the initial and sequential stages for both genres, while narrative writing also showed differences between Stage 5 and the middle stages for most revision indices.

The results indicated that although behaving similarly when performing narrative essays, L1 and L2 writers tended to differ in their behaviours in argumentative writing. It seems that, in contrast to L2 writers, L1 writers engaged in planning to a greater extent at the beginning of both argumentative and narrative writing, reflected as longer between-sentence pauses and relatively fewer total revisions in Stage 1. Probably, L1 writers tend to employ similar global strategies, such as generating and organising ideas, regardless of the genre of the text that they are composing (e.g. Haas, 1989; Van Waes & Schellens, 2003). However, L2 writers may experience greater differences in cognitive loads when writing essays with different communicative purposes. Therefore, it is likely that L2 writers would adjust the way in which they adopt global strategies when writing texts in different genres to meet the rhetorical requirements without overloading working memory, especially when there is a time limit on text production. This interpretation, to some extent, received confirmation from the verbal protocol data. It seems that the distribution of stimulated recall comments that describe planning, translation and monitoring processes across the five stages differed little between argumentative and narrative writing in L1, whereas in

L2 writing, a clear difference was observed between genres, as mentioned earlier in this section.

5.2.3 Summary of the section

In summary, for the whole session, L2 writers of Chinese were more fluent but paused more often within a word when producing argumentative than narrative essays. Less fluent narrative writing may be attributed to adult L2 writers' lack of practice in narratives in real life. More frequent within-word pauses may be explained by a deeper involvement in lower-order processes to fulfil the specific discourse purposes of argumentative essays which require the use of more advanced vocabulary. L1 writers were also found to be more fluent in argumentative writing. However, instead of pausing more frequently within a word, they made more pauses at word boundaries. This difference may be associated with a stylistic property of Chinese written language which features more frequent use of disyllabic words.

When stage was considered, the results revealed greater differences in the processes when writers were performing argumentative and narrative writing tasks in Chinese. This, again, highlights the need to examine writing, especially in the field of L2, from a temporal perspective to deepen the understanding of the intrinsic dynamism of the writing process, which has long been overlooked (Gánem-Gutiérrez & Gilmore, 2018; Manchón et al., 2009). Specifically, in argumentative writing, L2 Chinese writers tended to finish drafting before the final stage and to engage in re-reading for monitoring more towards the end, which echoes the findings of earlier studies on L2 writers of English (e.g. Barkaoui, 2016, 2019). In narrative writing, growing attention was paid to monitoring after the middle writing stage, possibly because of a lower degree of engagement in translation processes in writing narratives, leaving more cognitive resources available for other activities.

Additionally, L2 writers tended to carry out more planning in the initial stage when performing the narrative rather than the argumentative writing task. This may be due to the fact that adult L2 writers were more familiar with argumentative writing, thus spending less time at the beginning planning what to include in the essay. The differences in terms of the presentation and requirements of the writing prompts may be another reason for L2 argumentative writing processes not featuring more planning in the initial stage. Contrasting with L2 writing, more planning activities were found at the beginning of both argumentative and narrative writing in L1 Chinese. The L1-L2 difference may be a result of, unlike L2 writers, L1 writers being likely to experience similar cognitive loads when writing essays with different rhetorical purposes. Consequently, the global strategies (e.g. generating ideas, considering the structure) that L1 writers adopted to produce argumentative texts tended to be very much like those they utilised in narrative writing.

5.3 L2 proficiency and behaviours in Chinese writing

In this section, I move on to discuss the relationships between L2 proficiency and writing behaviours.

5.3.1 L2 proficiency and writing behaviours: the whole session

First, linear mixed-effects regression models suggested significant relationships between L2 proficiency and seven writing behaviour measures, including production rate, frequencies of pauses within a word and between Pinyin and character(s), and frequencies of revisions in total, below the clause level, at the clause level and above and to Chinese characters. These distinctions can be attributed to the competition between higher- (e.g. planning, monitoring) and lower-order processes (e.g. translation) for limited cognitive resources during writing (Kellogg, 1996). More proficient L2 writers, who are expected to have greater automaticity in translation, are less likely to be constrained by linguistic encoding processes, while less proficient writers may experience an increased demand on translation due to less sufficient linguistic resources for writing.

As manefisted in the results, writers of L2 Chinese exhibited greater speed fluency, expressed in terms of more Chinese characters produced per minute, as their proficiency increased. This mirrors previous research findings in English L2 writing (Barkaoui, 2019; Chenoweth & Hayes, 2001; Tiryakioglu et al., 2019). Employing either process- or product-based fluency measures, these studies also found that writers with higher L2 proficiency were more fluent than those with lower proficiency.

In terms of pausing behaviours, higher proficiency in L2 Chinese was found to be related to fewer pauses at lower linguistic unit boundaries. Earlier work suggested that L2 writers with lower proficiency were more likely to encounter language difficulties and struggle with linguistic problems that are often linked to vocabulary use during text production (e.g. Murphy & Roca de Larios, 2010; Tillema, 2012). To tackle this crucial problem, less proficient L2 writers may allocate more cognitive resources to finding suitable lexical items than more proficient ones, leading to more frequent pauses within a word, which is assumed to be potentially associated with lower-order cognitive processes, such as lexical retrieval (e.g. Révész et al., 2019).

It is noteworthy that, as expected, increased proficiency in L2 Chinese led to a decrease in the frequency of pauses made between Pinyin words and character(s). The negative correlation between L2 proficiency and pause frequency between Pinyin and character(s) tends to indicate that transcription activities, which are expected to place few demands on working memory (Kellogg, 1996), are likely to be non-automated in writing in Chinese. As mentioned in Chapter 3, typing Chinese on a computer involves two steps. The user has to first enter the Pinyin of the wanted Chinese character(s) and then select the character(s) from a list containing characters that share the same pronunciation and are sometimes relatively similar in shape. Due to this complicated transcription process, L2 writers, especially those whose L1 orthographic system is completely different from logographic Chinese, may have to stop frequently after typing Pinyin to select the character(s) they need because of the partial establishment of sound (Pinyin)-form (Chinese character) connections in Chinese orthography. Higher L2 proficiency in Chinese reduces the chances of L2 writers pausing between Pinyin words and character(s), suggesting a more automated process for mapping sound with form as L2 proficiency increases. Data obtained from L1 writers of Chinese lend further support to this argument. Although L1 writers also paused between Pinyin words and character(s) during task performance, the numbers of such pauses (.15 per minute) they made were considerably fewer than those of L2 writers (.49 per minute), with statistical significance (E = -.32, SE = .06, t = -4.89, p < .01, R^2 = .25).

However, the finding of negative relationships between L2 proficiency and certain pause measures contradicts Barkaoui (2019), where no difference was found between high and low proficiency groups in the frequency of pauses at various locations. This discrepancy in the findings might be attributed to the way in which L2 proficiency was operationalised. In Barkaoui (2019), each participant was allocated to a low or a high proficiency group depending on whether s/he was a pre- or post-admission student of a university in Canada. As the university normally imposes a minimum requirement for English proficiency for both pre- and post-admission students (usually IELTS 5.5, equivalent to lower B2 in CEFR, for the former and IELTS 6.5, equivalent to higher B2 in CEFR, for the latter), it is possible that the gap in L2 proficiency between the two groups was not big enough to reflect any differences in pausing behaviours. The level of L2 proficiency in the current study was established by a cloze test. The scores on the test ranged from 13 to 41 out of 45, with a mean of 28.25 and a standard deviation of 7.59. As the test consisted of 15 items from TOCFL Bands A, B and C (equivalent to Levels A, B and C in CEFR) respectively, the range of the scores indicated a relatively larger variance in terms of writers' L2 proficiency in this study than that in Barkaoui (2019). Perhaps, this may explain why I found relationships between L2 proficiency and certain pause frequency measures, while Barkaoui (2019) did not.

With regard to revisions, the first finding was that, overall, more proficient L2 writers of Chinese revised more often than less proficient ones. This echoes Gánem-Gutiérrez and Gilmore's (2018) finding that L2 proficiency positively correlated to the number of episodes that participants devoted to revising. When revisions were classified in terms of linguistic domain, I additionally found that L2 writers of higher proficiency made more revisions beyond the word level than those of lower proficiency. Taken together, the results show that higher L2 proficiency leads to more frequent and extensive revisions during the process of writing. As pointed out in previous literature (e.g. Barkaoui, 2007; Xu, 2018), the ability to revise indicates how effectively writers reallocate their attention in order to reshape their ideas, reconstruct the meaning and restructure utterances. As L2 proficiency develops, L2 writers are supposed to have more linguistic resources available for task completion. This enables them to produce alternatives to a greater extent, leading to more reshaping, reconstructing and restructuring, as they are less likely to 'get trapped in their initial anchorings' (Roca de Larios et al., 1999, p. 17), a typical result due to a lack of L2 proficiency. On the other hand, more proficient L2 writers tend to have fewer linguistic problems to deal with during text production than less proficient writers; consequently, more attention can be diverted from overcoming linguistic obstacles to other higher-order writing activities, which may be reflected in revisions to larger units of text (Whalen & Ménard, 1995).

Notably, more and less proficient L2 writers of Chinese also differed in their revision behaviours in terms of levels of transcription. That is, writers made similar numbers of Pinyin revisions regardless of their L2 proficiency, while more proficient writers outnumbered less proficient ones in the numbers of revisions made to Chinese characters. This difference might be related to the unique transcription process and orthographic system of Chinese. According to most classic models of writing (e.g. Flower & Hayes, 1981; Kellogg, 1996), transcribing a text is generally considered an automated process for adult writers, thus not competing for attentional resources with other activities during text production. However, in non-alphabetic language writing, this assumption may be questionable. It seems that the transcription process may not be fully automated and could be burdensome, even for adult writers (Ellis & Yuan, 2004; Hayes, 2012). First, it may not be an effortless task for L2 writers of Chinese to spell what they want to write in an accurate way in Pinyin (in letters though), as several sounds in Pinyin correspond to written forms in a slightly different way to those in their L1. Second, the non-transparent link between Chinese characters and their pronunciation (Kang, 2011) requires extra effort for L2 writers to find target characters due to the lack of full acquisition in sound and form mapping in Chinese. L2 writers with relatively lower proficiency tend to make and correct Pinyin mistakes more frequently when typing than those with higher proficiency. When there is a time limit for a task, it is likely that less proficient writers are not so willing to make changes once they have managed to get characters to appear on the screen because of time pressure. Consequently, writers with lower proficiency in L2 Chinese made comparatively fewer changes to characters than their more proficient counterparts. The development of L2 proficiency is supposed to speed up the process of linking pronunciation to characters, making it less onerous for L2 writers to carry out revisions to Chinese characters. As a result, more proficient L2 writers of Chinese revised characters more frequently than less proficient ones. This interpretation additionally received some support from the data obtained from L1 writers. L1 writers of Chinese, who were unlikely to experience cognitive tension caused by sound-form mapping, revised Chinese characters (3.25 per minute) more frequently than Pinyin (2.70 per minute). Inferential statistics further confirmed that the difference was significant (E = .13, SE = .03, t = 4.25, p < .01, R^2 = .05).

In addition, I found the effect of L2 proficiency was not dependent on the genre of writing. In other words, whether performing the argumentative or narrative writing tasks, more proficient L2 writers of Chinese were always more fluent, made fewer within-word and between-Pinyin-and-character(s) pauses, and carried out more revisions overall, to larger textual units and to Chinese characters than less proficient ones. Although the effect of genre as a modulator was limited, the findings still shed some useful light. First, it appears that the areas of difficulty in L2 writing, especially for writers with relatively lower proficiency, tend be similar irrespective of genre. The results showed that whether participants were producing argumentative or narrative essays, insufficient vocabulary and orthographic knowledge in L2 Chinese might be the greatest challenge that L2 writers, less proficient ones in particular, were faced with. Second, it seems that certain writing behaviours may serve as universal indicators in the profile of successful L2 writing, as proficiency-related differences in these behaviours were found to be comparable between argumentative and narrative writing.

5.3.2 L2 proficiency, writing behaviours: across the five writing stages

It is also important to look at how the relationships between L2 proficiency and writing behaviours were modulated by stages of writing. The results of linear mixed-effect regression analyses showed correlations between L2 proficiency and the temporal distribution of two revision behaviour indices, i.e. revision below the clause level and to Chinese characters, across the five writing stages. Post hoc analyses revealed no relationship between proficiency and either measure in Stage 1, while a relationship between L2 proficiency and below-clause revision was identified in Stages 3, 4 and 5, and one between proficiency and character revision in Stages 2, 3 and 5. The findings seems to suggest that L2 proficiency could, to some extent, mediate the temporal characteristics of writing, supporting Rijlaarsdam and Van Den Bergh's (1996) claim that the variance in the temporal distribution of writing activities could be explained by internal, writer-related factors. In addition, the results indicated that L2 proficiency may play a greater role after the initial writing stage when language-specific activities (e.g. formulation) surpass content-related activities (e.g. planning) and become the dominant process during text production.

On the other hand, I should acknowledge that the relationships between L2 proficiency and the temporal distribution of writing behaviours are relatively weak, as only two out of 25 relationships investigated in the current study was found to be significant. This seems to imply that writers with different levels of L2 proficiency may engage in largely similar cognitive activities throughout the process of performing writing tasks in L2 Chinese. This result is consistent with some of the findings reported in earlier studies for L2 writers of English. Similar to the present study, Barkaoui (2016) discovered that the temporal distribution of the total revision amount was not affected by L2 proficiency. Likewise, Gánem-Gutiérrez and Gillmore (2018) reported no relationship between L2 proficiency and writing processes across the five stages of writing. In Tillema (2012), the distribution of most cognitive activities across writing stages was largely the same between more and less proficient writers. L2 proficiency was only found to correlate with the temporal distribution of process planning (i.e. monitoring and metacomments), an activity that had a very low probability of occurrence during L2 writing processes in the study.

However, the weak relationship between L2 proficiency and the temporal dimension of writing processes needs to be interpreted with some caution. As Tillema (2012) argues, although L2 writers with higher and lower proficiency did not differentiate quantitatively in terms of the temporal distribution of writing activities, there might be qualitative differences when more and less proficient L2 writers were carrying out these activities. In the same vein, Gánem-Gutiérrez and Gillmore (2018) did indeed observe some quality differences regarding viewing behaviours when two participants of different L2 proficiency were consulting external recourses in a certain writing stage. In line with their argument, when examining the data in more detail, we found that more and less proficient L2 writers differed in terms of what they actually did during text production, even when the distribution of their writing behaviours was largely similar across the five stages.

Let us consider how Participants 15 and 31 engaged in translation processes during writing as examples to illustrate the contrasts. Participant 15, a more proficient writer, scored 39 out of 45 on the cloze test, while Participant 31, a less proficient writer, achieved 13 on the test. Both participants paused more frequently in Stages 2, 3 and 4, and less often in Stages 1 and 5 when performing the narrative writing task on a memorable childhood event. The verbal protocols showed that over half of the pauses they made in the middle writing stages were translation-related. However, differences emerged when we examined these stimulated recall comments at a more micro-level. The more proficient writer (Participant 15) tended to pause often between words in order to make decisions on phrases to use in the text, because he usually had more than one linguistic option available in mind, as demonstrated in Examples 14 and 15. His internal debate on the choice of lexical items eventually turned into more complex or accurate utterances that benefitted the overall text quality.

Example 14:

Researcher: What were you thinking when you were pausing here?

Participant 15: I was thinking, shall I use *理解* (understand) or *了解* (understand) or *意识到* (realise)?

Example 15:

Researcher: What were you thinking at that moment?

Participant 15: I was thinking whether to say *明明是有区别的* (…is obviously different) or *区别是明显的* (the difference is obvious).

In addition, Participant 15 was aware of the pragmatic aspect of the language that he was to write in the text, instead of focusing only on whether or not the phrase/sentence was grammatically correct. In Example 16, the participant showed his intention to make the story more appealing to the potential reader by looking for an interesting way to express the intended meaning. In Example 17, he was considering the style of the next sentence to match the given context (i.e. in a conversation).

Example 16:

Researcher: What were you thinking when you were stopping here?

Participant 15: The next sentence was *用外语刺激我的脑子* (use a foreign language to stimulate my brain), so I was thinking if I wanted to say that or say like *不再学一个外语* (no need to learn a foreign language). I was thinking which one sounded more interesting.

Example 17:

Researcher: What made you pause at that moment?

Participant 15: I was wondering whether I should wrote it in a spoken style, *你哪国人呢* (where are you from), *你哪国人* (where are you from) or *你来自哪个国家* (which country do you come from). I wanted to write it in a colloquial way 'cause it was a question.

In contrast, the less proficient participant (Participant 31) paused frequently to search for the Chinese translation of an English word due to his limited vocabulary size, as shown in Examples 18 and 19. In the first example, Participant 31 tried hard to find the Chinese equivalent for 'village' or 'town' but ended up putting the English name of his hometown in the text. In the second example, the participant struggled with the word 'tiny' in Chinese, which led to him pausing at the word boundary three times during the production of a very short clause. In the end, the participant succeeded in finding the correct translation for 'tiny'. However, he did not realise that the clause sounded unnatural, as he was heavily involved with translating the word 'tiny' into Chinese, which left him with little attention available to check other aspects of the utterance.

Example 18:

Researcher: What were you thinking when you were pausing here?

Participant 31: I was thinking about how to say 'village' or 'town'. I couldn't think of how to say it in Chinese.

Example 19:

Researcher: What were you thinking then?

Participant 31: I wanted to describe, and I wanted to say the word 'tiny', but I couldn't think how to say it.

Researcher: What were you thinking here?

Participant 31: Trying to think of something that sounded correct to say 'tiny'.

Researcher: What about then?

Participant 31: I was still thinking about how to say 'tiny'.

This nuanced level of examination of the verbal protocols clearly demonstrated that, although the stimulated recall comments above were all referred to as translation-related pauses between words in the middle stages of writing, they were not the same in nature. Because of the huge gap in L2 proficiency between the two participants, the translation processes were triggered by radically different reasons and, consequently, led to more or less successful writing outcomes.

5.3.3 Summary of the section

To conclude, it was found that, for the whole session, writers with higher proficiency in L2 Chinese were likely to be more fluent, paused less frequently within a word, and made more revisions in total and at the phrasal level and above than those with lower L2 proficiency. The findings indicate possible differences between more and less proficient L2 Chinese writers in terms of cognitive activities during the process of task performance. In addition, the results confirmed the prediction that L2 proficiency had a relatively strong relationship with the writing behaviours associated with the unique orthographic system of Chinese (i.e. the frequency of pausing between Pinyin and character(s), the amount of revision to Chinese characters). This finding indicates non-automated transcription processes in Chinese writing, suggesting that when modelling writing in non-alphabetic languages, such as Chinese, one adaptation needs to be made to the classic writing models (e.g. Flower & Hayes, 1981; Kellogg, 1996), i.e. placing an equal emphasis on transcription to other processes involved in text production.

L2 proficiency was found to correlate with two writing behaviour indices only when the stage of writing was taken into account. The weak relationship between proficiency and the temporal distribution of writing behaviours may be related to a methodological issue. That is, results from quantitative analyses alone may fail to disentangle the intertwined internal and external factors that may influence the temporal nature of L2 writing. As I demonstrated in the previous section, qualitative analyses with a more detailed investigation of writers' cognitive activities underlying their behaviours can potentially shed some light on the relationship between L2 proficiency and the temporal dimension of L2 writing processes.

5.4 Writing behaviours and text quality in Chinese writing

This section is dedicated to discussing the extent to which writing behaviours predict text quality in L2 Chinese writing. I additionally combine the results obtained from L1 Chinese writing to interpret the findings gained from L2 Chinese writers.

5.4.1 Writing behaviours and text quality: the whole session

The results of linear mixed-effects regression analyses suggested that text quality in L2 Chinese was predicted by six writing behaviours: production rate, length of P-burst, between-word pause frequency, and the duration of pauses overall, between words and between clauses. First, greater speed fluency, as characterised by a higher rate of production and longer P-bursts, led to better texts. The findings partially corroborate those of Spelman Miller et al. (2008), who also observed that burst duration positively correlated with L2 text quality. There is one interesting discrepancy between the findings of the current study and Spelman Miller et al.

(2008). That is, Spelman Miller et al. (2008) found that fluency during bursts (time on text production between pauses and/or revisions) rather than production rate (number of characters typed per minute) was a predictor of L2 text quality. However, this apparent difference does, in fact, indicate certain consistencies in the findings of the two studies. It is very likely that the production rate and fluency during bursts closely correlated, given that the two measures were both generated based on time, text length and revisions. Spelman Miller et al. (2008) included fluency during bursts when building regression models to reduce collinearity, whereas in the present study, I used production rate for the sake of convenience in calculation. As the two measures were found to be predictors of L2 text quality in both studies, respectively, and the correlation between them tended to be high, I argue that fluency measures calculated based on time and the amount of production, with revised characters/ words included, might effectively predict L2 text quality, no matter how the measures are manipulated in the study.

I additionally found that less frequent pauses at word boundaries, as well as shorter pauses overall, between words and between clauses tended to result in more successful outcomes in L2 Chinese writing. The results are inconsistent with Spelman Miller et al.'s (2008) finding that no relationship was identified between any pause measures and L2 text quality. This difference may be explained by the way in which pauses were analysed. First, pause duration was represented by the median length of pauses in the current study, but by the mean in Spelman Miller et al. (2008). Median length may be a more reliable representation of pause duration than the mean as the distribution of pause length is highly skewed (Baaijen et al., 2012). Therefore, the finding that pause duration was a weak predictor of L2 text quality in Spelman Miller et al. (2008) may be associated with how the measure was operationalised in the study, which could possibly have been biased due to the heterogeneous distribution of pause length.

Secondly, in the present study, pauses were analysed in terms of location, whereas all pauses were treated the same in Spelman Miller et al. (2008), that is, the researchers did not differentiate between pauses occurring at different linguistic unit boundaries. Previous literature in the fields of both L1 and L2 writing (e.g. Barkaoui, 2019; Medimorec & Risko, 2017; Révész et al., 2019; Van Waes & Leijten, 2015) has tested and proved that pauses at various locations are not equal in nature and may indicate different levels of processing during writing. Thus, an additional reason for the lack of relationship between pauses and L2 text quality in Spelman Miller et al. (2008) may be the relatively unsophisticated way in which pauses were examined. In the current study, between-word pause frequency and duration, as well as between-clause pause duration, emerged as predictors of L2 text quality in a more fine-grained approach that took into account pause location. It is also worth noticing that these predictors were all measures associated with pauses between smaller textual units. As mentioned in Chapter 4, it was evidenced in the verbal protocols that such pauses were primarily linked to translation processes. This may imply that, in general, language constraints tend to be one of the top issues that hinder L2 text quality.

Another noteworthy finding of the study is that L2 Chinese writing quality was not predicted by revision behaviours. This echoes previous research in L2 English writing that revision is a weak indicator of text quality (Chambers, 2011; Spelman Miller et al., 2008; Stevenson et al., 2006). I additionally discovered that the trend of no relationship between revision measures and text quality is not unique in L2 Chinese writing; the limited predictive power of revision behaviours for text quality was also observed in L1 Chinese writing. It is possible that the effect of revision to text quality is very subject to individual variation. In other words, the answer to whether more or fewer revisions are beneficial may vary considerably across writers or even within the same writer when s/he is engaging in different writing tasks. For instance, some participants may prefer to carry out external revisions extensively to make the text as accurate as possible, while others prefer to prepare and rehearse the utterances carefully in their mind before starting to type, so as to avoid too many revisions. As the backgrounds of the participants in our study were fairly diverse, it is not surprising that we did not find any relationship between revision behaviours and text quality.

Interestingly, three writing behaviour indices, i.e. between-sentence pause frequency, between-word pause duration and amount of word-level revision, were found to predict the quality of the texts produced by more and less proficient L2 writers in different ways. Specifically, texts created by less proficient L2 writers were likely to benefit from more pauses at sentence boundaries, while more proficient L2 writers tended to produce texts of better quality if they paused longer between words and revised at the word level more frequently. The findings indicate that less proficient L2 writers may have to balance their attentional resources between local and global aspects during text production, as more between-sentence pauses, which are mainly planning-related as previously revealed in the stimulated recall data, are supposed to result in better writing outcomes. Equipped with sufficient linguistic resources in L2, more proficient writers have the potential to generate utterances that best fit the context, even within the time limit. Consequently, careful selection of the language to use may improve the overall text quality.

Apart from the results for the research question, the study yielded an additional finding related to writing outcomes that is worth discussing. The mean length of the texts produced by L2 writers of Chinese in this study reached 541.28 (*SD* = 199.86) characters. The most proficient L2 writer produced four texts with an average length of 569.50 characters, and the least proficient writer 273.25. I found these figures quite surprising, as much shorter texts were produced by writers in earlier work that investigated writing outcomes in L2 Chinese. For example, in Wu (2016a, 2016b, 2017), advanced L2 writers of Chinese produced texts of approximately 200 characters, on average, in 40 minutes; in Fei (2015), the length of essays by writers at the B2 level according to the CEFR, in 30 minutes, ranged from 186 to 392 characters. Although the comparatively shorter texts in previous literature could be because the participants were intentionally making their essays concise, a minimum of 400 characters is in general needed to provide a full presentation of the writer's understanding of the topic (Fei, 2015). Thus, I speculate that the comparatively shorter texts might have to do with the mode of transcription in earlier studies, i.e.

handwriting, which is more time-consuming and involves more cognitive resources than using the Pinyin input method. Traditionally, the Pinyin input method is not welcomed in most L2 Chinese classrooms, as many teachers fear that using a word processor can harm the acquisition of writing in Chinese (Kang, 2011). This may be partially true; however, it should not be ignored that L2 writers in the current study created much longer texts using the Pinyin input method than what earlier researchers would have expected. It might be that typing Chinese characters using Pinyin occupies fewer attentional resources than writing characters stroke by stroke. The attention freed from struggling to put strokes together could be allocated to real 'writing' activities that involve knowledge-telling, knowledge-transforming (Bereiter & Scardamalia, 1987) and/or knowledge-constituting (Galbraith, 1999) processes, which would potentially improve text quality. Thus, I argue that the Pinyin input method may not be so detrimental to L2 writers of Chinese. If used properly, it has the potential to develop writers' ability to produce essays in L2 Chinese by directing more cognitive resources to activities, such as retrieving ideas, considering rhetorical requirements and developing an understanding of the topic, rather than draining their attention by recalling the exact shape of a certain Chinese character.

5.4.2 Writing behaviours and text quality: across the five writing stages

In terms of the extent to which writing behaviours predict L2 text quality across the five stages, longer within-word pauses in Stage 1 were found to lead to higher text quality in L2 Chinese writing; text quality benefitted from fewer between-word pauses and longer pauses between Pinyin and character(s) in Stage 2; shorter between-Pinyin-and-character(s) pauses in Stage 3 resulted in better texts; shorter pauses at clause boundaries improved text quality in Stage 5. The findings clearly demonstrate that the relationship between writing behaviours and L2 text quality was not constant throughout the text production process. That is, first of all, a certain writing behaviour can positively affect L2 text quality in one writing stage, but may become a negative predictor in subsequent stages. In addition, L2 text quality may correlate with certain writing behaviours in one writing stage, while it may be predicted by different behaviours in other stages. These findings are in line with those of Tillema (2012), Van Weijen (2008), Xu and Qi (2017), and Xu (2018), in which the researchers also observed that variances in L2 text quality were linked to variations in the temporal distribution of writing behaviours or cognitive activities during text production. Although disparities remain in terms of the specific way in which a writing activity predicts L2 text quality, the findings, obtained from both alphabetic and non-alphabetic language writers, seem to consistently lend empirical support to an implication inferred based on Rijlaarsdam and Van Den Bergh's (1996) temporal model of writing; that is, text quality relates to the moment at which a cognitive activity takes place during writing.

I additionally observed that the contribution of writing behaviours to text quality across stages tended to vary between L1 and L2 Chinese writing. Different from that in L2 writing, more between-sentence pauses in Stage 2 and shorter pauses overall in Stage 4 enhanced text quality in L1 Chinese writing. This finding resembles the trends reported in Tillema (2012) and Van Weijen (2008), i.e. that the temporal distribution of writing activities affects L1 and L2 text quality in different ways. However, the precise details of the relationships between processes and products across L1 and L2 writing may not be directly comparable between the studies due to methodological differences. First, in the present study, writing activities were captured via an unobtrusive method and were not mediated by think-alouds, as Tillema (2012) and Van Weijen (2008) did. This discrepancy may also be explained by the different types of writing tasks used in the studies: four argumentative writing tasks in Tillema (2012) and Van Weijen (2008), and two argumentative and two narrative writing tasks in the current study.

Another interesting observation is that the strength of the relationships between writing behaviours and text quality across the five stages tended to be modulated by genre and L2 proficiency (in the case of L2 writing). The findings support an insight derived from Rijlaarsdam and Van Den Bergh's (1996) temporal model of writing that the time-based features of writing, i.e. the link between the temporal distribution of writing activities and text quality, are constrained by internal (e.g. L2 proficiency) and external (e.g. genre) factors.

In terms of how genre mediated the extent to which behaviours predicted text quality in L2 Chinese, writers who paused less frequently between clauses in Stage 1 produced argumentative essays of higher quality, while those who paused more often at sentence boundaries in the same stage wrote better narrative texts. In Stage 5, the duration of pauses between Pinyin and character(s) positively correlated with L2 argumentative text quality, while the same indices negatively correlated with L2 narrative text quality. In L1 Chinese writing, fewer pauses between contextual revisions in Stage 1 led to better argumentative essays, whereas the quality of narrative writing was predicted by more pauses between Pinyin and character(s) in Stage 1 and more between-sentence pauses in Stage 2. The findings seem to suggest that when producing texts of different genres, writers are probably required to adopt different strategies at the beginning (and end) of writing to fulfil specific rhetorical purposes, whereas what happens in the middle stages, where language-specific activities constitute a dominant process, tend to have similar effects on text quality regardless of genre.

Turning to the effect of L2 proficiency on the links between behaviours and text quality, it seems that more proficient L2 writers produced better texts if they revised more often at the word level in Stage 1, and revisions beyond the word level in Stage 4. More overall pauses in Stage 5 are likely to improve text quality for less proficient L2 writers. The results also showed that longer between-sentence pauses in Stage 3 benefitted text quality for those who obtained over 22 in the cloze test. However, 27 out of 32 L2 writers in the current were above that point. This seems to indicate that the duration of pauses at sentence boundaries positively affected text quality for the majority of the participants in the data set, and only those who had very limited proficiency in L2 Chinese (around A2 according to the CEFR in this study) produced poorer texts with longer between-sentence pauses. These findings are reminiscent of some earlier work (e.g. Cummins, 1980; Sasaki & Hirose, 1996; Schoonen et al.,

2003) which suggested a certain level of L2 proficiency needs to be obtained before writers can execute writing activities in an efficient way. The results of the present study seem to further indicate that such a threshold may be dynamic throughout the writing process where certain stages tend to be dominated by certain writing activities. For example, in Stage 1 where planning processes are activated to a greater extent, more frequent word-level revision is associated with better-written texts when writers have reached a certain level of L2 proficiency; however, the threshold disappears after the initial stage when extensive drafting kicks in. In other words, the proficiency threshold seems to have a greater influence on revision at the word level, which is primarily translation-related, in planning- than translation-oriented stages. This implies that L2 writers who have surpassed the proficiency threshold could shift their attention between the dominant activity and other activities in a given stage more efficiently than those who have not yet reached this level. Given that the dominant activities change throughout text production, the threshold for L2 proficiency for the effective application of a writing activity will not be the same across writing stages.

5.4.3 Summary of the section

In short, the results have shown that text quality in L2 Chinese writing was predicted by speed fluency and measures related to pauses at lower linguistic unit boundaries. This indicates that sufficient and accessible language resources during writing may be a crucial factor for creating good L2 writing pieces. L2 proficiency was found to modulate the relationship between text quality and three writing behaviour indices, i.e. frequency of pauses between sentences, duration of pauses between words and amount of revision at the word level. Thanks to relatively adequate linguistic resources for writing, more proficient L2 writers tended to improve text quality by engaging more in word-level behaviours which mainly have to do with lexical selection. On the other hand, less proficient L2 writers may be able to produce better essays if more attention is allocated to higher-order processing that underlies the pauses at sentence boundaries.

In addition, the relationships between writing behaviours and text quality in both L1 and L2 Chinese writing were found to be mediated by the stage of writing, and these relationships were also modulated by genre and L2 proficiency (in the case of L2 writing). The findings lend support to several assumptions deduced from the temporal model of writing (Rijlaarsdam & Van Den Bergh, 1996) in the context of non-alphabetic language writing. First, text quality in Chinese writing is dependent on the moment at which the writer carries out a writing behaviour during the writing process. Secondly, the link between text quality and the temporal distribution of behaviours in Chinese writing is constrained by factors associated with the writer and the context, such as linguistic proficiency in the target language and the genre of writing.

Chapter 6 Conclusions and implications

In the final chapter, I first synthesise the findings for each of the five research questions in the current study. The chapter continues with a discussion of the theoretical, methodological and pedagogical implications of the results obtained. The last section of the chapter deals with the limitations of the study and some possible directions for further research.

6.1 Summary of the main findings

This section provides a brief summary of each research question and its corresponding findings. First, I summarise the results for the separate and joint effects of stage and genre on the cognitive processes in which L2 writers of Chinese engage. Next, the main findings of the relationships between L2 proficiency and writing behaviours in L2 Chinese, and the links between writing behaviours and L2 Chinese text quality, are presented. The section ends with a summary of the trends found in L1 Chinese writing regarding the influences of stage and genre on cognitive writing processes and the relation to text quality.

Research question 1: *To what extent do stages of writing affect cognitive processes in L2 Chinese writing?* I discovered an effect of writing stage on the cognitive processes in L2 Chinese writing. Specifically, the initial and/or final stages of writing were found to differentiate from the middle stages in terms of writing behaviours and associated cognitive activities. The findings parallel the observations obtained from L2 writers of English and confirm a dynamic L2 writing process in a non-alphabetic language.

Research question 2: *To what extent does genre affect the cognitive processes in L2 Chinese writing?* It was found that L2 writers of Chinese engaged in different writing behaviours and cognitive activities when writing argumentative and narrative essays throughout the whole of, and across the five stages of, the writing process. I argue that the genre differences reflected in the process of L2 writing may be a result of different levels of cognitive load that writers are experiencing during writing, which may be related to the specific communicative purposes of the text, writers' familiarity with a given genre and the unique stylistic properties of the language of writing.

Research question 3: *To what extent are L2 proficiency and the behaviours in L2 Chinese writing related? To what extent do the stage and genre of writing mediate these relationships?* The results revealed that L2 proficiency in Chinese was positively related to speed fluency and the amount of revision of larger textual unit levels, but correlated negatively to the frequency/ duration of pauses at smaller linguistic unit boundaries in writing. Notably, more and less proficient L2 writers behaved differently in terms of the pausing and revision indices associated with the unique orthographic system of Chinese. This highlights the importance of modelling non-alphabetic language writing with an explicit emphasis on the transcription process. Turning to the second half of the research question, I observed that the relationships between L2 proficiency and writing behaviours were mediated by stage, but to a relatively small extent. However, it is speculated that stronger relationships between L2 proficiency and the temporal dimension of text composition may be revealed if writing processes are analysed at a more nuanced level. Genre was not found to modulate the relationship between L2 proficiency and writing behaviours.

Research question 4: *To what extent do writing behaviours predict L2 Chinese text quality?* The findings showed that speed fluency and pauses at smaller textual unit boundaries were predictors of text quality in L2 Chinese. In addition, L2 text quality tended to be predicted by pauses between smaller textual units in different ways across the five stages of writing. Genre and L2 proficiency were found to be mediators of the links between writing behaviours and L2 text quality. The findings constitute supportive evidence for the hypotheses inferred based on the temporal model of writing (Rijlaarsdam & Van Den Bergh, 1996): text quality is related to the stage in which a certain writing activity occurs, and the temporal feature of process-product links is influenced by the internal and external factors of writing.

Research question 5: *Do these relationships differ in L1 Chinese writing?* Similar to L2 Chinese writing, an effect of stage was found in the cognitive processes in L1 Chinese writing, in spite of some slight differences with regard to certain pausing indices and associated cognitive activities. Unlike L2 writers, L1 writers of Chinese behaved differently when performing writing tasks in different genres, but they engaged in largely similar cognitive activities when producing argumentative and narrative pieces. Different from L2 writing, text quality in L1 Chinese was predicted by pauses between larger textual units and the amount of contextual revision in certain stages of writing. The strength of the relationships was dependent on genre, with minor differences to that in L2 writing.

6.2 Theoretical implications

The findings obtained in this study have several important implications for theory. To begin with, an effect of writing stages on cognitive processes, albeit small, emerged in L2 Chinese writing, which is consistent with the results of the limited number of studies that have investigated writing processes in L2 English with a time variable included (e.g. Gánem-Gutiérrez & Gilmore, 2018; Van Weijen, 2008). In addition, the time-related characteristics of L2 Chinese writing, such as the temporal distribution of writing activities and its relation to text quality, were found to be modulated by individual and contextual factors, such as L2 proficiency and genre. The results echo some trends reported in previous empirical works on L2 writers of English (e.g. Barkaoui, 2019; Manchón et al., 2009; Tillema, 2012). These temporal features, although not exactly the same as those in L2 writing, were also observed in L1 Chinese writing. These findings jointly offer some support, as discussed in Chapter 5, for the temporal model of writing and the assumptions derived from that model (Rijlaarsdam & Van Den Bergh, 1996) by empirically testing and confirming them in the context of writing in Chinese. This suggests that the temporal model of writing and perhaps other cognitive models initially developed to investigate L1 writing in an alphabetic language can serve as suitable starting points for theory-building in non-alphabetic language writing.

The second theoretical implication of this study concerns the process of transcription, a component that has long been neglected in the majority of cognitive models of writing (e.g. Flower & Hayes, 1981; Scardamalia & Bereiter, 1987).

Although Hayes (2012) has explicitly called for attention to be paid to the process of transcription by adding the component of transcriber to his latest model of writing, it has remained a tacit agreement that transcription in adults' writing process is thoroughly automated, and thus it is not supposed to have any significant influence on other cognitive activities while writing. The only exception, where transcription was found to compete with other activities for cognitive resources, was when researchers deliberately made the mode of transcription unfamiliar to adult writers. For example, in Bourdin and Fayol (1994), the transcription process was found to be burdensome when the participants were asked to produce texts using upper case. Nevertheless, constructing writing models based on such observations carries potential risks to validity, as writing only in capitals does not normally take place in the real world. As a result, it can be questioned whether or not to include transcription as an essential component when conceptualising adults' writing processes, because transcription is highly likely to be effortless when adult writers are performing a real writing task. However, several findings in the current study when the transcription mode was not manipulated seem to pose a challenge to the widely shared belief in a fully automated transcription process in adults' writing. In other words, I found that even when the writing task was carried out with the most popular transcription mode among L1 and L2 writers of Chinese (i.e. the Pinyin input method), the process of transcription still required extra cognitive effort. This has been clearly demonstrated in the differences between more and less proficient L2 writers as well as L1 and L2 writers of Chinese in terms of certain pausing and revision behaviours that are associated with the logographic Chinese writing system and the dual-step procedure of typing Chinese. On the basis of these findings, it is speculated that a certain number of cognitive resources may have to be allocated to the process of transcription when adult writers, and L2 writers in particular, are producing essays in Chinese. This calls for putting a clear emphasis on the transcription process when modelling writing processes in a non-alphabetic language.

6.3 Methodological implications

Apart from the implications for theory, the present study also has some implications for research methodology. First, unlike the majority of L2 writing research in which writers' text production processes were interpreted based on a single data set, the current study has employed a combination of data collection techniques, i.e. keystroke-logging and stimulated recall interviews. It was anticipated that triangulating data obtained from different sources would address the limitations of the use of a single method as references made based on data triangulation tend to be more reliable and valid than conclusions drawn from one data set alone. Therefore, we expected that the results of this study would provide more complete and specific information on the cognitive processes in which L1 and L2 writers engage. Although the findings of the current study have turned out to be largely similar (though with several slight differences) to those of earlier work, this study, which has incorporated both behavioural and verbal protocol data, is more likely to paint an accurate picture of writers' mental representations when they are producing text. This highlights the benefits of adopting a mixed-methods approach in writing studies for a more comprehensive understanding of the issue under exploration.

The second methodological implication of this study is related to the way in which the temporal dimension of writing has been analysed. Although our findings have confirmed a dynamic process in both L1 and L2 writing of Chinese, the fact should not be neglected that the influence of time during writing, operationalised as five equal stages in the present study, on cognitive processes is relatively limited. On the other hand, a snapshot of the data from two individual writers that I presented in Section 5.3.2 revealed big qualitative differences in terms of the distribution of pauses at word boundaries. This raises an interesting question: in how much detail should behaviours and cognitive activities be examined in order to uncover the dynamic nature of the writing process in relation to individual and contextual factors? At present, it might be quite difficult to untangle such a web given the relatively small number of studies that have investigated the temporal dimension of writing. Nevertheless, it seems that an examination at the micro-level, which has been briefly touched upon in the previous section, may offer some information about the dynamism of writing from a different angle as compared to a macro-level investigation. As a matter of fact, some researchers (e.g. Gánem-Gutiérrez & Gilmore, 2018; Van Weijen, 2008) have indeed observed greater differences in the cognitive activities involved in writing between individuals with higher and lower L2 proficiency when temporal analyses of writing process were carried out at a more nuanced level. There is no doubt that exploring writing processes in relation to the time variable is of great value for a more thorough understanding of how writers produce essays in real time (Roca de Larios et al., 2008). However, it is of equal importance to carry out more fine-grained qualitative data analysis at various levels (e.g. Chan, 2017; Larsen-Freeman & Cameron, 2008) to investigate the relationships between individual and contextual factors and the temporal dimension of cognitive writing processes.

6.4 Pedagogical implications

We should consider whether or not the findings obtained from the current study, which was conducted in a laboratory setting, are directly applicable to real-world writing situations. However, it is worth noting, with caution, some pedagogical implications from the results presented in this thesis.

First of all, information about the cognitive processes involved in writing and process-product relationships may shed some useful light on identifying good writing strategies for efficient text production. For example, the findings of the current study showed that cognitive activities did not occur randomly across the five stages of the writing process. L2 writers may benefit from learning strategies that can help to distribute cognitive resources in a more effective way, such as defining clear goals for certain stages. The results also indicated that genre and writer's L2 proficiency may mediate the relationships between writing behaviours and text quality. This suggests a potential advantage in providing L2 writers with strategies that are tailored to the genre of the text and their L2 proficiency.

The additional finding of comparatively longer texts by L2 writers using the Pinyin input method than those written with pen and paper in earlier studies indicates the potential value of using a word processor when writing Chinese essays. This is certainly not to deny the importance of learning to write Chinese characters in strokes. However, to produce a good essay in Chinese is not merely about writing each character in the text correctly; it is also associated with a wider range of issues. In order to develop well-rounded writing abilities, L2 writers of Chinese should be given opportunities to practise various strategies during writing. However, the prevailing use of handwriting in Chinese L2 writing classrooms seems to deprive writers of these opportunities, as character writing is one of the greatest challenges faced by learners of Chinese (Gunderson et al., 2011) and a big obstacle during their writing task performance. Use of the Pinyin input method can, to some extent, lighten the load that character writing places on L2 writers' working memory. In addition, L2 writers might be more motivated to write, as the Pinyin input method makes writing less challenging. As a result, L2 writers of Chinese are likely to focus on and interact with the actual task of writing to a greater extent, which could be beneficial to the development of their overall writing abilities. This calls for a balanced use of the Pinyin input method and handwriting in L2 Chinese writing classes.

6.5 Limitations and directions for future research

It is important to acknowledge the limitations of the current study, which, on the other hand, point to several areas for future research. The first major weakness resides in the partial triangulation of the data. Although I have combined keystroke-logging and retrospective verbal reports, and triangulated the data obtained from two sources, the triangulation was not complete due to a technical limitation of *Translog* 2.0, the keystroke-logging software used in this study. That is, the recording used for eliciting writers' cognitive activities associated with pausing and revision behaviours during writing was a replay of their writing performance at the character level. In other words, the writers were unable to see in the recording what they had typed and deleted in Pinyin before the Pinyin was converted into Chinese character(s). Thus, the validity of the data obtained via stimulated recall may be questionable. For example, it is possible that a higher proportion of pauses and revisions associated with translation processes would be found if the participants had access to stimuli at the Pinyin level, such as pause within a word, pause between Pinyin and character(s), and revision below the word level. Future studies could address this shortcoming by using a screen recorder to capture the complete writing process in Chinese writing (i.e. at both Pinyin and character levels) as a more effective stimulus for the interviews.

A second limitation of this study also concerns the tool for data collection. Although an attempt was made to tap into the cognitive processes during writing with the joint use of keystroke-logging and stimulated recall interviews, our findings provided limited information about the viewing behaviours in which writers are involved. As the process of text composition features the changing task environment that is shaped by the text-produced-so-far (Flower & Hayes, 1981; Galbraith, 1999), our understanding of writing processes can never be complete without exploring the role that visual input plays and the interaction between reading and other activities during writing. Reading through already-written text can serve different purposes in different writing stages. For example, rereading is more likely to be associated with planning what to write in the earlier stages of writing, but it tends to be linked to evaluating whether or not the text has met the task requirements in later stages. Consequently, rereading may not always be followed by the same writing activity; it might trigger formulation at the beginning, while activating revision towards the end of writing.

One way to address this limitation is to incorporate an eye-tracking technique. This technique allows researchers to capture writers' real-time eye movements, which are assumed to reflect how attentional resources are directed when they are performing writing tasks (Duchowski, 2007). Eye-tracking has been used in a few studies investigating L1 writing (e.g. Alamargot et al., 2011; Torrance et al., 2016) and is just starting to gain in popularity in the field of L2 writing (e.g. Gánem-Gutiérrez & Gilmore, 2018; Révész et al., 2019). Undoubtedly, more studies with a combination of research methods that include eye-tracking are called for to shed more light on writing processes. Research questions can be formulated in terms of the extent to which writers engage in re-reading activities during writing. For example, it may be worth investigating at which text location writers are more likely to look back, and whether the distances between lookback fixations and the point of inscription differ at various text locations. This could help to build a more thorough understanding of writing fluency, as where the writer looks during pauses (e.g. within the linguistic unit that has just been formulated versus the paragraph prior the point of inscription) is seen as an indicative of different levels of processing (Chukharev-Hudilainen et al., 2019; Torrance et al., 2016). In terms of studies on revision behaviours, it would be interesting to explore at which part of the already-written text the writer has looked before a revision is triggered. This could offer useful insights to writers' revision strategies, especially by comparing task performances under different conditions (e.g. a test versus a non-test condition). Eye-tracking technology also potentially permits new research opportunities in L1 and L2 writing in Chinese. For instance, eye-tracking could enable researchers to compare the way in which writers of L1 and L2 Chinese interact with the list providing homophones for them to select the desired Chinese character(s) after typing the Pinyin. Differences between L1 and L2 writers' approaches to selecting characters may reveal specific struggling points in L2 Chinese writing in terms of sound-form mapping and character recognition. Additionally, information gathered by eye-tracking technology about how L2 writers interact with the list could advance our knowledge on whether and how L2 writers of Chinese can notice the gap between their interlanguage and target forms (Williams, 2012) in a writing task for the development of L2 linguistic knowledge.

Third, text quality in the current study was only operationalised as the mean score given by two independent raters using a holistic rating scale. Although both raters were experienced L2 writing instructors and had received proper training before the rating was conducted, an overall score carries a certain degree of subjectivity, and thus runs the risk of being biased. Therefore, based on the findings obtained in this study, only limited conclusions can be drawn about the extent to which writing behaviours predict the quality of texts. In a follow-up study, the link between writing behaviours and other aspects of text quality assessed via objective linguistic measures, such as linguistic complexity and accuracy, could be investigated. A combination of holistic rating and objective linguistic measures may enable us to gain more profound insights into the relationships between writing processes and text quality. Apart from examining these relationships via a cross-sectional approach, it would also be interesting to conduct studies in a longitudinal design to look at how writing processes and process-product relations change over time. Longitudinal studies may yield more fruitful results if writers' individual differences, such as working memory, aptitude, motivation and self-efficacy, to name but a few, are taken into consideration.

A fourth shortcoming of the study is that we did not control for the participants' keyboarding skills. Previous studies on alphabetic language writers have revealed a potential effect of writers' keyboarding skills on patterns of pausing and revision behaviours during text production (e.g. Alves et al., 2007 for L1 writing; Barkaoui, 2016, 2019 for L2 writing). Assuming that the same applies to non-alphabetic language writers, not controlling for keyboarding skills could potentially confound some results of the current study (e.g. the relationships between L2 proficiency and writing behaviours, potential behavioural differences between L1 and L2 writers), despite all the participants’ claims that they were familiar and comfortable with writing essays in Chinese on a computer using the Pinyin input method. To mitigate this problem, future work could recruit participants with similar keyboarding skills by, for example, administering a typing test during the recruiting process.

A related limitation concerns the choice for adopting a two-second pause threshold. As mentioned in Chapter 3, the decision was made mainly due to a practical reason that identifying and coding pauses had to be conducted manually in this study and thus, it is impossible to employ shorter pause thresholds given the limited time and human resources. Nevertheless, this choice can be challenged on several grounds. First, the two-second pause threshold was developed mainly based on observations obtained from alphabetic language writers. Therefore, it is debatable whether the same threshold should apply to writers of a non-alphabetic language to set apart cognitive activities during writing and execution of the simple mechanics of typing (i.e. the transition between from key to another). In addition, recent studies in the fields of L1 and L2 writing questioned the appropriateness of using the two-second threshold to exclude motor activities that are irrelevant to cognitive activities. Some researchers (e.g. Chukharev-Hudilainen, 2014; Olive & Kellogg, 2002) have argued that the duration of two seconds may be too long to provide adequate information about lower-order processing, such as linguistic encoding processes, thus suggesting utilising lower pause thresholds or multiple pause thresholds for pauses at different text locations. For example, future studies might consider employing various pause thresholds (e.g. 200, 500, 1000 and 2000 milliseconds) and compare the results gained from multiple thresholds to find out where differences emerge. Another avenue for further research is to establish suitable thresholds for pauses at various text locations for non-alphabetic language writing. It might be conducive to start with a copy task, which involves motor-level planning only (Conijn et al., 2019). A copy task could help to identify a range of inter-word keystroke intervals and then, based on this range, the pause threshold in non-alphabetic language writing might be determined.

Another limitation associated with pause analysis is the way in which pause duration was operationalised (i.e. median length of pause). The major reason was due to the relatively skewed distribution of the duration data of each task; in this case, median is assumed to be a more valid representation of the construct than mean. However, this decision could potentially reduce the comparability between the findings of the current thesis and those of previous studies that used the mean value. Additionally, as mentioned in Chapter 3, to further improve the normality in distribution, I trimmed outliers to values of three standard deviations from the mean for each writing behaviour index. This resulted in a replacement of around 5% of the duration data, with a higher percentage of those concerning pauses between sentences and between contextual revisions. This could potentially harm the validity and reliability of the results given that a small proportion of the data have been artificially manipulated. One possible solution to the issue is to operationalise pause duration with multiple measures (e.g. mean, median, logarithmic mean) without deleting/trimming the data and compare how different the results would be if different measures are utilised.

In addition, the results may be difficult to generalise due to the potential bias regarding the participants and the writing tasks. First, in order to control for L1 orthographic influences, all L2 writers in this study were L1 users of an alphabetic language (mainly English). The patterns observed here may not be generalisable to other populations, such as L2 writers of Chinese whose L1 is also a non-alphabetic language (e.g. Japanese). It is very likely that the similarities between the orthographic systems would increase fluency or reduce pauses and revisions that are associated with transcription processes. It might, therefore, be worthwhile in future to investigate the writing performance of L2 writers of Chinese with an L1 background other than English or an alphabetic language. Another bias concerning the participants is that the writers who took part in the present study were all relatively young and might be fairly confident with their L1 or L2 writing skills as they volunteered for the research. In other words, the sample may not well represent the traits of the population of writers of Chinese. However, as in most studies of L2 acquisition, it is logistically impossible to perform random sampling. Thus, we shall keep this bias in mind when interpreting the findings of the current study. Furthermore, the tasks used to elicit participants' writing performance could limit the generalisability of our findings, especially to real-life writing tasks. Similar to the majority of (quasi-)experimental studies on L1 and L2 writing, participants in this study were asked to complete each essay within a time limit and not allowed to use any external materials. Although it is normal to ask writers to perform timed writing tasks without referring to other resources in a writing classroom or a test environment, it seems essential for future studies to consider the use of additional materials, such as a dictionary or the Internet, during writing, given the increasingly common practice of using external resources, online information in particular, when people are producing essays in the real world (Gánem-Gutiérrez & Gilmore, 2018; Leijten et al., 2014).

Finally, in future studies of L2 writing, it would also be interesting to investigate the potential pedagogical value of keystroke-logging techniques. As proposed by a few researchers (e.g. Lindgren & Sullivan, 2003), the recorded writing process may serve as a stimulus for L2 writers to reflect on and evaluate what they did and thought during writing. Indeed, in the current study, some L2 writers of Chinese realised that they were interrupted too often by language issues while watching the video of them writing the essay in their stimulated recall interview. They also claimed that they did not notice this when producing the essay. In future, researchers could design studies to compare the quality of texts produced by L2 writers before and after reviewing and reflecting on the writing processes captured by a keystroke-logging program. The recording could either be of the writer's own process of text composition or of how other writers (e.g. more proficient L2 learners, L1 writers) produce an essay. If improvement in (certain aspects of) text quality was found, it might be a promising practice to introduce a keystroke-logging technique into L2 writing classrooms. It would enable L2 writing instructors to provide process-based feedback and model good writing by demonstrating both successful writing outcomes and the process through which a good writing piece is developed.

References

Abdel Latif, M. M. M. (2008). A State-of-the-Art Review of the Real-Time Computer-Aided Study of the Writing Process. *International Journal of English Studies*, *8*(1), 29–50.

Abdel Latif, M. M. M. (2009). Toward a New Process-Based Indicator for Measuring Writing Fluency: Evidence from L2 Writers' Think-Aloud Protocols. *Canadian Modern Language Review*, *65*(4), 531–558.

Abdel Latif, M. M. M. (2013). What Do We Mean by Writing Fluency and How Can It Be Validly Measured? *Applied Linguistics*, *34*(1), 99–105.

Abdel Latif, M. M. M. (2019). Using think-aloud protocols and interviews in investigating writers' composing processes: Combining concurrent and retrospective data. *International Journal of Research & Method in Education*, *42*(2), 111–123.

Abdi Tabari, M. (2016). The effects of planning time on complexity, accuracy, fluency, and lexical variety in L2 descriptive writing. *Asian-Pacific Journal of Second and Foreign Language Education*, *1*(1), 1–15.

Aitken, K. G. (1977). Using Cloze Procedure as an Overall Language Proficiency Test. *TESOL Quarterly*, *11*, 59–67.

Alamargot, D., Caporossi, G., Chesnet, D., & Ros, C. (2011). What makes a skilled writer? Working memory and audience awareness during text composition. *Learning and Individual Differences*, *21*(5), 505–516.

Alderson, C. J. (1979). The Cloze Procedure and Proficiency in English as a Foreign

Language. *Teachers of English to Speakers of Other Languages Quarterly*, *13*(2), 219–27.

Aldridge, M., Fontaine, L., Bowen, N., & Smith, T. (2018). A new perspective on word association: How keystroke logging informs strength of word association. *WORD*, *64*(4), 218–234.

Alves, R. A., Castro, S. L., De Sousa, L., & Strömqvist, S. (2007). Influence of keyboarding skill on pauseexecution cycles in written composition. In M. Torrance, L. Van Waes, & D. Galbraith (Eds.), *Writing and cognition: Research and applications* (pp. 55–65). Elsevier.

An, F. (2015). Analysis of Fluency, Grammatical Complexity and Accuracy of CSL Writing: A Study Based on T-unit Analysis. *Yuyan Jiaoxue yu Yanjiu (Language Teaching and Linguistic Studies)*, *03*, 11–20.

Baaijen, V. M., Galbraith, D., & De Glopper, K. (2012). Keystroke Analysis: Reflections on Procedures and Measures. *Written Communication*, *29*(3), 246–277.

Bachman, L. F. (1986). Performance on Cloze Tests with Fixed-Ratio and Rational Deletions. *TESOL Quarterly*, *19*(3), 535–56.

Baddeley, A. (1986). *Working Memory*. Clarendon.

Baddeley, A. (1992). Working Memory. *Science*, *255*(5044), 556–559.

Baddeley, A. (2010). Working memory. *Current Biology*, *20*(4), 136–140.

Baddeley, A., & Hitch, G. J. (1974). Working Memory. In G. A. Bower (Ed.), *Recent Advances in Learning and Motivation: Vol. Volume 8* (pp. 47–89). Academic

Press.

Barkaoui, K. (2007). Revision in Second Language Writing: What Teachers Need to Know. *TESL Canada Journal*, *25*(1), 81–92.

Barkaoui, K. (2014). Examining the Impact of L2 Proficiency and Keyboarding Skills on Scores on TOEFL-iBT Writing Tasks. *Language Testing*, *31*(2), 241–259.

Barkaoui, K. (2015). Test Takers' Writing Activities During the TOEFL iBT® Writing Tasks: A Stimulated Recall Study. *ETS Research Report Series*, *2015*(1), 1–42.

Barkaoui, K. (2016). What and When Second-Language Learners Revise When Responding to Timed Writing Tasks on the Computer: The Roles of Task Type, Second Language Proficiency, and Keyboarding Skills. *Modern Language Journal*, *100*(1), 320–340.

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

Beauvais, C., Olive, T., & Passerault, J.-M. (2011). Why are some texts good and others not? Relationship between text quality and management of the writing processes. *Journal of Educational Psychology*, *103*(2), 415–428.

Beauvais, L., Favart, M., Passerault, J.-M., & Beauvais, C. (2014). Temporal Management of the Writing Process: Effects of Genre and Organizing Constraints in Grades 5, 7, and 9. *Written Communication*, *31*(3), 251–279.

Beers, S. F., & Nagy, W. E. (2009). Syntactic Complexity as a Predictor of Adolescent Writing Quality: Which Measures? Which Genre? *Reading and*

Writing: An Interdisciplinary Journal, *22*(2), 185–200.

Beers, S. F., & Nagy, W. E. (2011). Writing Development in Four Genres from Grades Three to Seven: Syntactic Complexity and Genre Differentiation. *Reading and Writing: An Interdisciplinary Journal*, *24*(2), 183–202.

Bereiter, C., & Scardamalia, M. (1987). *The Psychology of Written Composition*. Lawrence Erlbaum Associates.

Biber, D., & Conrad, S. (2009). *Register, genre, and style*. Cambridge University Press.

Bosher, S. (1998). The Composing Processes of Three Southeast Asian Writers at the Post-Secondary Level: An Exploratory Study. *Journal of Second Language Writing*, *7*(2), 205–241.

Bourdin, B., & Fayol, M. (1994). Is Written Language Production more Difficult than Oral Language Production? A Working Memory Approach. *International Journal of Psychology*, *29*(5), 591–620.

Bourdin, B., & Fayol, M. (1996). Mode effects in a sentence production span task. *Cahiers De Psychologie Cognitive – Current Psychology of Cognition*, *15*, 245–264.

Bouwer, R., Béguin, A., Sanders, T., & Van Den Bergh, H. (2014). Effect of genre on the generalizability of writing scores. *Language Testing*, *32*(1), 83–100.

Braaksma, M. A. H., Rijlaarsdam, G., Van Den Bergh, H., & Van Hout-Wolters, B. H. A. M. (2004). Observational Learning and Its Effects on the Orchestration of Writing Processes. *Cognition and Instruction*, *22*(1), 1–36.

Breetvelt, I., Van Den Bergh, H., & Rijlaarsdam, G. (1994). Relations Between Writing Processes and Text Quality: When and How? *Cognition and Instruction*, *12*(2), 103–123.

Bridwell, L., Sirc, G., & Brooke, R. (1985). Revising and computing: Case studies of student writers. In S. W. Freedman (Ed.), *The acquisition of written language: Response and revision/ Sarah Warshauer Freedman, editor* (pp. 172–194). Ablex PubCorp.

Bridwell-Bowles, L., Johnson, P., & Brehe, S. (1987). Composing and computers: Case studies of experienced writers. In A. Matsuhashi (Ed.), *Writing in real time: Modeling production processes* (pp. 81–107). Ablex PubCorp.

Broekkamp, H., & Van Den Bergh, H. (1996). Attention strategies in revising a foreign language text. In H. Van Den Bergh & G. Rijlaarsdam (Eds.), *Theories, Models and Writing research* (pp. 170–181). Amsterdam University Press.

Brooks, L., & Swain, M. (2009). Language in collaborative writing: Creation of and response to expertise. In A. Mackey & C. Polio (Eds.), *Multiple perspectives on interaction: Second language research in honor of Susan M. Gass* (pp. 58–89). Routledge.

Brown, J. S., Mcdonald, J. L., Brown, T. L., & Carr, T. H. (1988). Adapting to Processing Demands in Discourse Production: The Case of Handwriting. *Journal of Experimental Psychology: Human Perception and Performance*, *14*(1), 45–59.

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.

Burtis, P. J., Bereiter, C., Scardamalia, M., & Tetroe, J. (1983). The Development of Planning in Writing. In G. Wells & M. Kroll (Eds.), *Explorations in the Development of Writing* (pp. 153–174). Longman.

Carl, M. (2012). *Translog-II*. 13th International Conference on Intelligent Text Processing and Computational Linguistics.

Chambers, L. (2011). Composition and revision in computer-based written assessment. *Research Notes*, *43*(1), 25–31.

Chan, S. (2017). Using keystroke logging to understand writers' processes on a reading-into-writing test. *Language Testing in Asia*, *10*(7), 1–27.

Chen, X. (2003). On writing course in teaching Chinese as a second language. *Language Teaching and Linguistic Studies*, *5*, 59–63.

Chenoweth, N. A., & Hayes, J. R. (2001). Fluency in Writing: Generating Text in L1 and L2. *Written Communication*, *18*(1), 80–98.

Chenoweth, N. A., & Hayes, J. R. (2003). The Inner Voice in Writing. *Written Communication*, *20*(1), 99–118.

Chenu, F., Pellegrino, F., Jisa, H., & Fayol, M. (2014). Interword and intraword pause threshold in the writing of texts by children and adolescents: A methodological approach. *Frontiers in Psychology*, *5*, 1–7.

Chukharev-Hudilainen, E. (2014). Pauses in spontaneous written communication: A keystroke logging study. *Journal of Writing Research*, *6*, 61–84.

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

Crawford, L., Lloyd, S., & Knoth, K. (2008). Analysis of Student Revisions on a State Writing Test. *Assessment for Effective Intervention*, *33*(2), 108–119.

Cumming, A. (1989). Writing Expertise and Second-Language Proficiency. *Language Learning: A Journal of Applied Linguistics*, *39*(1), 81–141.

Cumming, A. (1995). Fostering writing expertise in ESL composition instruction: Modeling and evaluation. In D. D. Belcher & G. Braine (Eds.), *Academic Writing in a Second Language: Essays on Research and Pedagogy* (pp. 375–397). Greenwood Publishing Group.

Cummins, J. (1980). The Cross-Lingual Dimensions of Language Proficiency: Implications for Bilingual Education and the Optimal Age Issue. *TESOL Quarterly*, *14*(2), 175–87.

Cunnings, I. (2012). An overview of mixed-effects statistical models for second language researchers. *Second Language Research*, *28*(3), 369– 382.

Danzak, R. L. (2011). The Integration of Lexical, Syntactic, and Discourse Features in Bilingual Adolescents' Writing: An Exploratory Approach. *Language, Speech, and Hearing Services in Schools*, *42*(4), 491–505.

De Silva, R., & Graham, S. (2015). The effects of strategy instruction on writing strategy use for students of different proficiency levels. *System*, *53*, 47–59.

De Smet, M. J. R., Brand-Gruwel, S., Leijten, M., & Kirschner, P. A. (2014). Electronic outlining as a writing strategy: Effects on students' writing products, mental effort and writing process. *Computers & Education*, *78*, 352–366.

Deane, P. (2013). On the relation between automated essay scoring and modern views of the writing construct. *Assessing Writing*, *18*(1), 7–24.

Deane, P. (2014). Using Writing Process and Product Features to Assess Writing Quality and Explore How Those Features Relate to Other Literacy Tasks. *ETS Research Report Series*, *2014*(1), 1–23.

Duchowski, A. T. (2007). *Eye Tracking Methodology*. Springer London.

Ellis, R. (2005). *Planning and task performance in a second language*. John Benjamins.

Ellis, R., & Yuan, F. (2004). The Effects of Planning on Fluency, Complexity, and Accuracy in Second Language Narrative Writing. *Studies in Second Language Acquisition*, *26*(1), 59–84.

Ericsson, A. K., & Simon, H. A. (1993). *Protocol Analysis: Verbal Reports as Data (Revised Edition)*. The MIT Press.

Faigley, L., & Witte, S. (1981). Analyzing Revision. *College Composition and Communication*, *32*(4), 400–14.

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191.

Fei, F. (2015). *Formulaic language use in L2 Chinese: The role of pre-writing planning*. Michigan State University.

Feng, S. (2006). On the properties and pedagogy of written Chinese. *Shijie Hanyu*

Jiaoxue (Chinese Teaching in the World), *04*, 98–106, 148.

Flinn, J. Z. (1987a). Case studies of revision aided by keystroke recording and replaying software. *Computers and Composition*, *5*(1), 31–44.

Flinn, J. Z. (1987b). Programming software to trace the composing process. *Computers and Composition*, *5*(1), 45–49.

Flower, L., & Hayes, J. R. (1981). A Cognitive Process Theory of Writing. *College Composition and Communication*, *32*(4), 365–387.

Fotos, S. S. (1991). The Cloze Test as an Integrative Measure of EFL Proficiency: A Substitute for Essays on College Entrance Examinations? *Language Learning*, *41*(3), 313–336.

Gaillard, S., & Tremblay, A. (2016). Linguistic Proficiency Assessment in Second Language Acquisition Research: The Elicited Imitation Task. *Language Learning*, *66*(2), 419–447.

Galbraith, D. (1992). Conditions for discovery through writing. *Instructional Science*, *21*(1–3), 45–71.

Galbraith, D. (1999). Writing as a Knowledge-Constituting Process. In M. Torrance & D. Galbraith (Eds.), *Knowing What to Write* (pp. 139–160). Amsterdam University Press.

Galbraith, D. (2009a). Cognitive models of writing.pdf. *German as a Foreign Language*, *2*, 7–22.

Galbraith, D. (2009b). Writing as discovery. *British Journal of Educational Psychology*, *2*(6), 5–26.

Galbraith, D., Torrance, M., & Hallam, J. (2006). Effects of writing on conceptual coherence. *Proceedings of the 28th Annual Conference of the Cognitive Science Society*, 1340–1345.

Gánem-Gutiérrez, G. A., & Gilmore, A. (2018). Tracking the Real-Time Evolution of a Writing Event: Second Language Writers at Different Proficiency Levels. *Language Learning*, *68*(2), 469–506.

Gass, S. M., & Mackey, A. (2016). *Stimulated Recall Methodology in Applied Linguistics and L2 Research*. Routledge.

Gellert, A. S., & Elbro, C. (2013). Cloze Tests May be Quick, but Are They Dirty? Development and Preliminary Validation of a Cloze Test of Reading Comprehension. *Journal of Psychoeducational Assessment*, *31*(1), 16–28.

Ghavamnia, M., Tavakoli, M., & Esteki, M. (2013). The Effect of Pre-Task and Online Planning Conditions on Complexity, Accuracy, and Fluency on EFL Learners' Written Production. *Porta Linguarum*, *20*, 31–43.

Glynn, S. M., Britton, B. K., Muth, K. D., & Dogan, N. (1982). Writing and revising persuasive documents: Cognitive demands. *Journal of Educational Psychology*, *74*(4), 557–567.

Grabe, W. P. (2002). Genre in the Classroom: Multiple Perspectives. In A. M. Johns (Ed.), *Genre in the Classroom: Multiple Perspectives* (pp. 249–267). Lawrence Erlbaum Associates.

Gunderson, L., Odo, D. M., & D'Sliva, R. (2011). Second Language Literacy. In E. Hinkel (Ed.), *Handbook of Research in Second Language Teaching and*

Learning (Vol. 2, pp. 472–487). Routledge.

Haas, C. (1989). How the writing medium shapes the writing process: Effects of word processing on planning. *Research in the Teaching of English, 23*, 181–207.

Hanania, E., & Shikhani, M. (1986). Interrelationships Among Three Tests of Language Proficiency: Standardized ESL, Cloze, and Writing. *TESOL Quarterly*, *20*(1), 97–109.

Hayes, J. R. (1996). A New Framework for Understanding Cognitive and Affect in Writing. In C. M. Levy & S. Ransdell (Eds.), *The science of writing: Theories, methods, individual differences, and applications* (pp. 1–28). Lawrence Erlbaum Associates.

Hayes, J. R. (2012). Modeling and Remodeling Writing. *Written Communication*, *29*(3), 369–388.

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

Hayes, J. R., & Flower, L. S. (1986). Writing Research and the Writer. *American Psychologist*, *41*(10), 1106–1113.

Hayes, J. R., Flower, L., Schriver, K., Stratman, J., & Carey, L. (1987). Cognitive Processes in Revision. In S. Rosenberg (Ed.), *Advances in Applied Psycholinguistics: Volume 2, Reading, Writing, and Language Learning* (pp. 176–240). Cambridge University Press.

Hirose, K., & Sasaki, M. (1994). Explanatory Variables for Japanese Students'

Expository Writing in English: An Exploratory Study. *Journal of Second Language Writing*, *3*(3), 203–229.

Hsu, H.-C. (2012). Investigating the Effects of Planning on L2 Text Chat Performance. *CALICO Journal*, *29*(4), 619–638.

Hu, G., & Lu, S. (2016). Effects of Task Complexity on the Linguistic Performance of CSL Writing Tasks. *Guoji Hanyu Jiaoxue Yanjiu (Journal of International Chinese Teaching)*, *01*, 80–86.

Hyland, K. (2003). Genre-based pedagogies: A social response to process. *Journal of Second Language Writing*, *12*(1), 17–29.

Hyland, K. (2016). Methods and methodologies in second language writing research. *System*, *59*, 116–125.

Janssen, D., Van Waes, L., & Van Den Bergh, H. (1996). Effects of Thinking Aloud on Writing Processes. In C. M. Levy & S. Ransdell (Eds.), *The science of writing: Theories, methods, individual differences, and applications* (pp. 233–250). Lawrence Erlbaum Associates.

Jeong, H. (2017). Narrative and expository genre effects on students, raters, and performance criteria. *Assessing Writing*, *31*, 113–125.

Jiang, W. (2013). Measurements of Development in L2 Written Production: The Case of L2 Chinese. *Applied Linguistics*, *34*(1), 1–24.

Jin, H. (2007). *Syntactical Maturity in Second Language Writing: A Case of Chinese as a Foreign Language (CFL)*. *1*(42), 27–53.

Jourdenais, R. (2001). Cognition, instruction and protocol analysis. In P. Robinson

(Ed.), *Cognition and Second Language Instruction* (pp. 354–376). Cambridge University Press.

Kang, H. (2011). *Computer-based Writing and Paper-based Writing: A Study of Beginning-level and Intermediate-level Chinese Learners' Writing*. The Ohio State University.

Katzenberger, I. (2005). The Super-Structure of Written Expository Texts—A Developmental Perspective. In D. D. Ravid & H. B.-Z. Shyldkrot (Eds.), *Perspectives on Language and Language Development* (pp. 327–336). Springer US.

Kellogg, R. T. (1988). Attentional Overload and Writing Performance: Effects of Rough Draft and Outline Strategies. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*(2), 355–365.

Kellogg, R. T. (1990). Effectiveness of prewriting strategies as a function of task demands. *American Journal of Psychology*, *103*(3), 327–342.

Kellogg, R. T. (1996). A Model of Working Memory in Writing. In C. M. Levy & S. Ransdell (Eds.), *The science of writing: Theories, methods, individual differences, and applications* (pp. 57–72). Lawrence Erlbaum Associates.

Kellogg, R. T. (1999). Components of Working Memory and Text Production. In M. Torrance & G. C. Jeffery (Eds.), *The cognitive demands of writing: Processing capacity and working memory effects in text production* (pp. 42–61). Amsterdam University Press.

Kellogg, R. T. (2001). Competition for working memory among writing processes.

American Journal of Psychology, *114*(2), 175–191.

Kellogg, R. T., Olive, T., & Piolat, A. (2007). Verbal and Visual Working Memory in Written Sentence Production. In M. Torrance, L. Van Waes, & D. Galbraith (Eds.), *Writing and Cognition: Research and Application* (pp. 97–108). Elsevier.

Khuder, B., & Harwood, N. (2015). L2 writing in test and non-test situations: Process and product. *Journal of Writing Research*, *6*(3), 233–278.

Knoch, U., Fairbairn, J., Myford, C., & Huisman, A. (2018). Evaluating the relative effectiveness of online and face-to-face training for new writing raters. *Language Testing and Assessment*, *7*(1), 61–86.

Kormos, J. (2012). The Role of Individual Differences in L2 Writing. *Journal of Second Language Writing*, *21*(4), 390–403.

Krapels, A. R. (1990). An overview of second language writing process research. In B. Kroll (Ed.), *Second Language Writing* (pp. 37–56). Cambridge University Press.

Larsen-Freeman, D., & Cameron, L. (2008). *Complex systems and applied linguistics*. Oxford University Press.

Leijten, M., Van Horenbeeck, E., & Van Waes, L. (2019). Analysing Keystroke Logging Data from a Linguistic Perspective. In E. Lindgren & K. P. H. Sullivan (Eds.), *Observing Writing: Insights from Keystroke Logging and Handwriting* (pp. 72–95). Brill.

Leijten, M., & Van Waes, L. (2006). Inputlog: New perspectives on the logging of on-line writing processes in a Windows environment. In K. P. H. Sullivan & E. Lindgren (Eds.), *Computer Key-Stroke Logging and Writing: Methods and*

Applications (pp. 73–94). Elsevier.

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.

Leijten, M., Van Waes, L., Schrijver, I., Bernolet, S., & Vangehuchten, L. (2019). Mapping master's students' use of external sources in source-based writing in L1 and L2. *Studies in Second Language Acquisition*, *41*, 555–582.

Leijten, M., Van Waes, L., Schriver, K., & Hayes, J. R. (2014). Writing in the workplace: Constructing documents using multiple digital sources. *Journal of Writing Research*, *5*(3), 285–337.

Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. The MIT Press.

Levy, C. M., & Marek, P. (1999). Testing components of Kellogg's multicomponent model of working memory in writing: The role of phonological loop. In M. Torrance & G. C. Jeffery (Eds.), *The cognitive demands of writing: Processing capacity and working memory effects in text production* (pp. 25–41). Amsterdam University Press.

Levy, C. M., & Ransdell, S. (1995). Is writing as difficult as it seems? *Memory & Cognition*, *23*(6), 767–779.

Li, S. (2015). On the Research of Individual Differences in the L2 Writing Development. *Yuyan Jiaoxue yu Yanjiu (Language Teaching and Linguistic Studies)*, *06*, 27–34.

Liao, J. (2018). The impact of face-to-face oral discussion and online text-chat on L2

Chinese writing. *Journal of Second Language Writing*, *41*, 27–40.

Linck, J. A., & Cunnings, I. (2015). The Utility and Application of Mixed-Effects Models in Second Language Research. *Language Learning*, *65*(1), 185–207.

Lindgren, E. (2005). *Writing and revising: Didactic and Methodological Implications of Keystroke Logging*. Umeå University.

Lindgren, E., Leijten, M., & Van Waes, L. (2011). Adapting to the reader during writing. *Written Language & Literacy*, *14*(2), 188–223.

Lindgren, E., & Sullivan, K. P. H. (2003). Stimulated Recall as a Trigger for Increasing Noticing and Language Awareness in the L2 Writing Classroom: A Case Study of Two Young Female Writers. *Language Awareness*, *12*(3–4), 172–186.

Lindgren, E., & Sullivan, K. P. H. (2006). Analysing Online Revision. In K. P. H. Sullivan & E. Lindgren (Eds.), *Computer Key-Stroke Logging and Writing* (pp. 157–188). Elsevier.

Liu, H., & Ling, W. (2012). Comparison and Analysis between Domestic and Foreign Studies on Second Language Writing. *Yunnan Shifan Daxue Xuebao (Duiwai Hanyu Jiaoxue yu Yanjiu Ban) (Journal of Yunnan Normal University (Teaching and Research on Chinese as a Foreign Language)), 10(3),* 29–40.

Liu, S., & Cao, Q. (2015). On the Effects of Five Different Writing Tasks on Students' Writing. *TCSOL Studies*, *4*, 77–88.

Liu, X., & Furneaux, C. (2014). A multidimensional comparison of discourse organization in English and Chinese university students' argumentative writing:

English and Chinese university students' argumentative writing. *International Journal of Applied Linguistics*, *24*(1), 74–96.

Liu, Y. (2007). The Form Errors of the General Article. *Chinese Language Learning*, *5*, 78–82.

Lu, X. (2011). A Corpus-Based Evaluation of Syntactic Complexity Measures as Indices of College-Level ESL Writers' Language Development. *TESOL Quarterly: A Journal for Teachers of English to Speakers of Other Languages and of Standard English as a Second Dialect*, *45*(1), 36–62.

Lutz, J. A. (1987). A Study of Professional and Experienced Writers Revising and Editing at the Computer and with Pen and Paper. *Research in the Teaching of English*, *21*(4), 398–421.

Ma, X. (2014). Jiyu Yukuai Lilun de Duiwai Hanyu Xiezuo Jiaoxue Moshi Shiyan Yanjiu (Investigating CSL Writing Instruction Based on Lexical Approach). *Duiwai Hanyu Yanjiu (Studies on Chinese as a Foreign Language)*, *01*, 176–184.

Madigan, R. J., Johnson, S. E., & Linton, P. W. (1994). *Working memory capacity and the writing process*. American Psychological Society, Washington D. C.

Manchón, R. M., & Roca de Larios, J. (2007). On the Temporal Nature of Planning in L1 and L2 Composing. *Language Learning*, *57*(4), 549–593.

Manchón, R. M., Roca de Larios, J., & Murphy, L. (2009). The Temporal Dimension and Problem-solving Nature of Foreign Language Composing Processes. In *Writing in Foreign Language Contexts* (pp. 102–129). MPG Books Group.

Matsuhashi, A. (1981). Pausing and Planning: The Tempo of Written Discourse Production. *Research in the Teaching of English*, *15*(2), 113–134.

Matsuhashi, A. (1987). Revising the plan and altering the text. In A. Matsuhashi (Ed.), *Writing in real time: Modelling production processes* (pp. 197–223). Alex Publishing Corporation.

Mazgutova, D. (2015). *Linguistic and Cognitive Development of L2 Writing During an Intensive English for Academic Purposes (EAP) Programme*. Lancaster University.

Mccutchen, D., & Perfetti, C. A. (1982). Coherence and connectedness in the development of discourse production. *Text - Interdisciplinary Journal for the Study of Discourse*, *2*(1), 113–140.

Medimorec, S., & Risko, E. F. (2017). Pauses in written composition: On the importance of where writers pause. *Reading and Writing; Dordrecht*, *30*(6), 1267–1285.

Mo, D. (2016). A study on American and European CSL students formal style development in Chinese written register. *Language Teaching and Linguistic Studies*, *181*(5), 20–29.

Murphy, L., & Roca de Larios, J. (2010). Searching for words: One strategic use of the mother tongue by advanced Spanish EFL writers. *Journal of Second Language Writing*, *19*(2), 61–81.

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.

Olive, T., & Kellogg, R. T. (2002). Concurrent activation of high-and low-level production processes in written composition. *Memory & Cognition, 30*, 594–600.

Ong, J. (2014). How do Planning Time and Task Conditions Affect Metacognitive Processes of L2 Writers? *Journal of Second Language Writing*, *23*(Mar), 17–30.

Ong, J., & Zhang, L. J. (2010). Effects of Task Complexity on the Fluency and Lexical Complexity in EFL Students' Argumentative Writing. *Journal of Second Language Writing*, *19*(4), 218–233.

Ong, J., & Zhang, L. J. (2013). Effects of the Manipulation of Cognitive Processes on EFL Writers' Text Quality. *TESOL Quarterly*, *47*(2), 375–396.

Ortega, L. (2015). Syntactic complexity in L2 writing: Progress and expansion. *Journal of Second Language Writing*, *29*, 82–94.

Passerault, J.-M., & Dinet, J. (2000). The role of visuospatial sketchpad in the written production of descriptive and argumentative texts. *Current Psychology Letters: Behaviour, Brain & Cognition*, *3*, 31–42.

Piri, F., Barati, H., & Ketabi, S. (2015). The Effects of Pre-Task, On-Line, and Both Pre-Task and On-Line Planning on Fluency, Complexity, and Accuracy—The Case of Iranian EFL Learners' Written Production. *English Language Teaching*, *5*(6), 158–167.

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

Polio, C. (2012). How to Research Second Language Writing. In A. Mackey & S. M.

Gass (Eds.), *Research Methods in Second Language Acquisition A Practical Guide* (pp. 139–157). Blackwell Publishing Ltd.

Polio, C., & Friedman, D. A. (2017). *Understanding, Evaluating, and Conducting Second Language Writing Research*. Routledge.

Porte, G. (1996). When writing fails: How academic context and past learning experiences shape revision. *System*, *24*(1), 107–116.

Porte, G. K. (1997). The Etiology of Poor Second Language Writing: The Influence of Perceived Teacher Preferences on Second Language Revision Strategies. *Journal of Second Language Writing*, *6*(1), 61–78.

Raimes, A. (1985). What Unskilled ESL Students Do as They Write: A Classroom Study of Composing. *TESOL Quarterly*, *19*(2), 229–258.

Raimes, A. (1987). Language Proficiency, Writing Ability, and Composing Strategies: A Study of ESL College Student Writers. *Language Learning*, *37*(3), 439–468.

Ransdell, S. (1995). Generating Thinking-Aloud Protocols: Impact on the Narrative Writing of College Students. *The American Journal of Psychology*, *108*(1), 89–98.

Ransdell, S., Levy, C. M., & Kellogg, R. T. (2002). The structure of writing processes as revealed by secondary task demands. *L1-Educational Studies in Language and Literature*, *2*(2), 141–163.

Ravid, D. D. (2005). Emergence of Linguistic Complexity in Later Language Development: Evidence from Expository Text Construction. In D. D. Ravid & H. B.-Z. Shyldkrot (Eds.), *Perspectives on Language and Language Development* (pp. 337–355). Springer US.

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. (2019). Introduction to the special issue. *Studies in Second Language Acquisition*, *41*(3), 491–501.

Révész, A., Michel, M., & Lee, M. (2019). Exploring second language writers' pausing and revision behaviors: A mixed methods study. *Studies in Second Language Acquisition*, *41*(3), 605–631.

Rijlaarsdam, G., & Van Den Bergh, G. (1996). The Dynamic of Composing—An Agenda for Research into an Interactive Compensatory Model of Writing: Many Questions, Some Answers. In C. M. Levy & S. Ransdell (Eds.), *The science of writing: Theories, methods, individual differences, and applications* (pp. 107–126). Lawrence Erlbaum Associates.

Rijlaarsdam, G., & Van Den Bergh, H. (2008). Writing Process Theory: A Functional Dynamic Approach. In C. A. MacArthur, S. Graham, & J. Fitzgerald (Eds.), *Handbook of Writing Research* (pp. 41–53). The Guilford Press.

Rijlaarsdam, G., Van den Bergh, H., Couzijn, M., Janssen, T., Braaksma, M., Tillema, M., Van Steendam, E., & Raedts, M. (2012). Writing. In K. R. Harris, S. Graham, T. Urdan, A. G. Bus, S. Major, & H. L. Swanson (Eds.), *APA educational psychology handbook: Vol. Vol 3: Application to learning and teaching* (pp. 189–227). American Psychological Association.

Roca de Larios, J., Manchón, R. M., & Murphy, L. (2006). Generating Text in Native and Foreign Language Writing: A Temporal Analysis of ProblemSolving Formulation Processes. *The Modern Language Journal*, *90*(1), 100–114.

Roca de Larios, J., Manchón, R. M., Murphy, L., & Marín, J. (2008). The Foreign Language Writer's Strategic Behaviour in the Allocation of Time to Writing Processes. *Journal of Second Language Writing*, *17*(1), 30–47.

Roca de Larios, J., Marín, J., & Murphy, L. (2001). A temporal analysis of formulation processes in L1 and L2 writing. *Language Learning*, *51*(3), 497–538.

Roca de Larios, J., Murphy, L., & Manchón, R. M. (1999). The Use of Restructuring Strategies in EFL Writing: A Study of Spanish Learners of English as a Foreign Language. *Journal of Second Language Writing*, *8*(1), 13–44.

Roca de Larios, J., Murphy, L., & Marín, J. (2002). A Critical Examination of L2 Writing Process Research. In S. Ransdell & M. L. Barbier (Eds.), *New directions for research in L2 writing* (pp. 1–29). Dordrecht, The Netherlands: Kluwer.

Roca de Larios, J., Nicolás-Conesa, F., Coyle, Y., & Matsuda, P. K. (2016). Focus on writers: Processes and strategies. In M. Manchon (Ed.), *Handbook of Second and Foreign Language Writing* (pp. 267–286). Walter de Gruyter, Inc. Boston/Berlin.

Rohman, D. G. (1965). Pre-Writing: The Stage of Discovery in the Writing Process. *College Composition and Communication*, *16*(2), 106–12.

Sasaki, M. (2004). A Multiple-Data Analysis of the 3.5-Year Development of EFL Student Writers. *Language Learning*, *54*(3), 525–582.

Sasaki, M., & Hirose, K. (1996). Explanatory Variables for EFL Students' Expository Writing. *Language Learning*, *46*(1), 137–168.

Scardamalia, M., & Bereiter, C. (1987). Knowlege Telling and Knowledge Transforming in Written Composition. In S. Rosenberg (Ed.), *Advances in Applied Psycholinguistics: Volume 2 Reading, Writing and Language Learning* (pp. 142–175). Cambridge University Press.

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

Schoonen, R., van Gelderen, A., de Glopper, K., Hulstijn, J., Simis, A., Snellings, P., & Stevenson, M. (2003). First language and second language writing: The role of linguistic knowledge, speed of processing, and metacognitive knowledge. *Language Learning*, *53*(4), 165–202.

Schrijver, I., Van Vaerenbergh, L., Leijten, M., & Van Waes, L. (2016). The impact of writing training on transediting in translation, analyzed from a product and process perspective. *Perspectives: Studies in Translatology*, *24*(2), 218–234.

Skehan, P., Bei, X., Li, Q., & Wang, Z. (2012). The Task Is Not Enough: Processing Approaches to Task-Based Performance. *Language Teaching Research*, *16*(2), 170–187.

Skehan, P., & Foster, P. (2005). Strategic and on-line planning: The influence of surprise information and task time on second language performance. In R. Ellis (Ed.), *Planning and Task Performance in a Second Language* (pp. 193–216). John Benjamins.

Skibniewski, L. (1988). The writing processes of foreign language learners in their native and foreign languages. *Studia Anglica Posnaniensia*, *21*, 177–186.

Sonbul, S., & Schmitt, N. (2013). Explicit and Implicit Lexical Knowledge: Acquisition of Collocations Under Different Input Conditions. *Language Learning*, *63*(1), 121–159.

Spelman Miller, K. (2005). Second language writing research and pedagogy: A role for computer logging? *Computers and Composition*, *22*(3), 297–317.

Spelman Miller, K., Lindgren, E., & Sullivan, K. P. H. (2008). The Psycholinguistic Dimension in Second Language Writing: Opportunities for Research and Pedagogy Using Computer Keystroke Logging. *TESOL Quarterly*, *42*(3), 433–454.

Stevenson, M., Schoonen, R., & De Glopper, K. (2006). Revising in two languages: A multi-dimensional comparison of online writing revisions in L1 and FL. *Journal of Second Language Writing*, *15*(3), 201–233.

Tao, Z. (2012). *Nancun chuo geng lu (Records of discontinuing farming)*. Shanghai Guji Chubanshe (Shanghai Ancient Works Publishing House).

Tardy, C. M. (2011). Editorial: The history and future of genre in second language writing. *Journal of Second Language Writing*, *20*(1), 1–5.

Thorson, H. (2000). Using the Computer To Compare Foreign and Native Language Writing Processes: A Statistical and Case Study Approach. *Modern Language Journal*, *84*(2), 155–170.

Tillema, M. (2012). *Writing in first and second language: Empirical studies on text* *quality and writing processes*. LOT.

Tillema, M., Van Den Bergh, H., Rijlaarsdam, G., & Sanders, T. (2011). Relating Self Reports of Writing Behaviour and Online Task Execution Using a Temporal Model. *Metacognition and Learning*, *6*(3), 229–253.

Tirkkonen-Condit, S. (2006). Think-Aloud Protocols. In K. Brown (Ed.), *Encyclopedia of Language and Linguistics* (Vol. 14, pp. 678–686). Elsevier Ltd.

Tiryakioglu, G., Peters, E., & Verschaffel, L. (2019). The Effect of L2 Proficiency Level on Composing Processes of EFL Learners: Data from Keystroke Loggings, Think Alouds and Questionnaires. In E. Lindgren & K. P. H. Sullivan (Eds.), *Observing Writing: Insights from Keystroke Logging and Handwriting* (Vol. 38, pp. 212–235). Brill.

Tong, B. (1996). Cong Kouyu Dao Shumian Yu----Zhongji Hanyu Jiaoxue Keti Zhiyi (From Spoken to Written Language: An issue in L2 Chinese Teaching to Intermediate learners). *Shijie Hanyu Jiaoxue (Chinese Teaching in the World)*, *4*, 94–99.

Torkildsen, J. von K., Morken, F., Helland, W. A., & Helland, T. (2016). The dynamics of narrative writing in primary grade children: Writing process factors predict story quality. *Reading and Writing*, *29*(3), 529–554.

Torrance, M., Johansson, R., Johansson, V., & Wengelin, Å. (2016). Reading during the composition of multi-sentence texts: An eye-movement study. *Psychological Research*, *80*(5), 729–743.

Valfredini, A. (2015). Studying the Process of Writing in a Foreign Language: An

Overview of the Methods. *Journal of Language Teaching and Research*, *6*(5), 907–912.

Van Den Bergh, H., & Rijlaarsdam, G. (1996). The Dynamic of Composing: Modeling Writing Process Data. In C. M. Levy & S. Ransdell (Eds.), *The science of writing: Theories, methods, individual differences, and applications* (pp. 207–233). Lawrence Erlbaum Associates.

Van Den Bergh, H., & Rijlaarsdam, G. (1999). The Dynamics of Idea Generation During Writing: An On-line Study. In M. Torrance & D. Galbraith (Eds.), *Knowing What to Write* (pp. 99–120). Amsterdam University Press.

Van Den Bergh, H., & Rijlaarsdam, G. (2001). Changes in Cognitive Activities During the Writing Process and Relationships with Text Quality. *Educational Psychology*, *21*(4), 373–385.

Van Waes, L., & Leijten, M. (2006). Logging Writing Processes with Inputlog. In L. Van Waes, M. Leijten, & C. M. Neuwirth (Eds.), *Writing and Digital Media* (pp. 158–166). Elsevier.

Van Waes, L., & Leijten, M. (2015). Fluency in Writing: A Multidimensional Perspective on Writing Fluency Applied to L1 and L2. *Computers and Composition*, *38*, 79–95. Scopus.

Van Waes, L., Leijten, M., Lindgren, E., & Wengelin, Å. (2015). Keystroke logging in writing research: Analyzing online writing processes. In C. A. MacArthur, S. Graham, & J. Fitzgerald (Eds.), *Handbook of Writing Research (Second Edition)* (pp. 410–427). Guilford Publications.

Van Waes, L., Leijten, M., & Van Weijen, D. (2009). Keystroke Logging in Writing Research Observing Writing Processes with Inputlog. *GFL - German as a Foreign Language, 2*, 41–64.

Van Waes, L., & Schellens, P. J. (2003). Writing profiles: The effect of the writing mode on pausing and revision patterns of experienced writers. *Journal of Pragmatics*, *35*(6), 829–853.

Van Waes, L., Van Weijen, D., & Leijten, M. (2014). Learning to write in an online writing center: : The effect of learning styles on the writing process. *Computers and Education*, *73*, 60–71.

Van Weijen, D. (2008). *Writing processes, text quality, and task effects: Empirical studies in first and second language writing.* LOT.

Vandermeulen, N. (2018). *The design and effect of two types of writing process feedback*. Sig Writing Conference 2018, Antwerp.

Victori, M. (1999). An Analysis of Writing Knowledge in EFL Composing: A Case Study of Two Effective and Two Less Effective Writers. *System: An International Journal of Educational Technology and Applied Linguistics, 27*(4), 537–55.

Wang, L., & Liu, Y. (2015). Acquisition of the Chinese theme-rheme constructions by native English speakers. *Language Teaching and Linguistic Studies, 5*, 33–41.

Wang, W., & Wen, Q. (2002). L1 use in the L2 composing process: An exploratory study of 16 Chinese EFL writers. *Journal of Second Language Writing*, *11*(3), 225–246.

Wang, Z., Guo, Y., Zheng, S., Xu, W., Liu, L., Liu, Z., & Cui, X. (2018). Users' location analysis based on Chinese mobile social media. *Concurrency and Computation: Practice and Experience,* e4669.

Way, D. P., Joiner, E. G., & Seaman, M. A. (2000). Writing in the Secondary Foreign Language Classroom: The Effects of Prompts and Tasks on Novice Learners of French. *Modern Language Journal*, *84*(2), 171–184.

Weigle, S. C. (2013). English language learners and automated scoring of essays: Critical considerations. *Assessing Writing*, *18*(1), 85–99.

Wengelin, Å. (2006). Examining pauses in writing: Theory, methods and empirical data. In K. P. Sullivan & E. Lindgren (Eds.), *Computer keystroke logging and writing* (pp. 106–130). Elsevier.

Whalen, K. (1993). A strategic approach to the development of second language written production processes. *Proceedings of the International Conference of Applied Linguistics*, 604–617.

Whalen, K., & Ménard, N. (1995). L1 and L2 Writers' Strategic and Linguistic Knowledge: A Model of Multiple-Level Discourse Processing. *Language Learning*, *45*(3), 381–418.

Williams, J. (2012). The potential role(s) of writing in second language development. *Journal of Second Language Writing*, *21*(4), 321–331.

Wolfe-Quintero, K., Inagaki, S., & Kim, H.-Y. (1998). *Second language development in writing: Measures of fluency, accuracy, and complexity*. University of Hawaii Press.

Wong, Y. K. (2018). Exploring the Reading-Writing Relationship in Young Chinese Language Learners' Sentence Writing. *Reading and Writing: An Interdisciplinary Journal*, *31*(4), 945–964.

Wu, J. (2016a). Research on Lexical richness development in CSL writing by English native speakers. *Chinese Teaching In The World*, *30*(1), 129–142.

Wu, J. (2016b). The grammatical complexity in English native speakers' Chinese writing. *Language Teaching and Linguistic Studies*, *4*, 27–35.

Wu, J. (2017). A case study: Dynamic development of English native speaker's CSL writing. *Modern Foreign Languages*, *40*(2), 1–11.

Xiao, J., Liu, B., & Wang, X. (2007). Exploiting pinyin constraints in pinyin-to-character conversion task: A class-based maximum entropy markov model approach. *International Journal of Computational Linguistics & Chinese Language Processing, Volume 12, Number 3, September 2007: Special Issue on Invited Papers from ISCSLP 2006*, *12*(3), 325–348.

Xing, J. Z. (2016). Analysis of Ellipsis Errors in Chinese as a Foreign Language. *Shijie Hanyu Jiaoxue (Chinese Teaching in the World)*, *30*(4), 531–549.

Xu, C. (2018). Understanding online revisions in L2 writing: A computer keystroke-log perspective. *System*, *78*, 104–114.

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

Xue, F. (2013). On Factors Influencing the Design of Second Language Writing Test

Tasks. *Yuyan Jiaoxue yu Yanjiu (Language Teaching and Linguistic Studies), 04*, 1–7.

Yang, C., Hu, G., & Zhang, L. J. (2014). Reactivity of concurrent verbal reporting in second language writing. *Journal of Second Language Writing*, *24*(Jun), 51–70.

Yang, C., Zhang, L. J., & Parr, J. M. (2019). The reactivity of think-alouds in writing research: Quantitative and qualitative evidence from writing in English as a foreign language. *Reading and Writing*, *33*, 451–483.

Yanguas, I., & Lado, B. (2012). Is Thinking Aloud Reactive When Writing in the Heritage Language? *Foreign Language Annals*, *45*(3), 380–399.

Yoon, H.-J., & Polio, C. (2016). The Linguistic Development of Students of English as a Second Language in Two Written Genres. *TESOL Quarterly*.

Yuan, F. (2010). Impacts of Task Conditions on Learners' Output in L2 Chinese Narrative Writing. *Journal of the Chinese Language Teachers Association*, *45*(1), 67–88.

Zamel, V. (1983). The Composing Processes of Advanced ESL Students: Six Case Studies. *TESOL Quarterly*, *17*(2), 165–78.

Zhang, D. (2017). Word Reading in L1 and L2 Learners of Chinese: Similarities and Differences in the Functioning of Component Processes. *Modern Language Journal*, *101*(2), 391–411.

Appendices

Appendix A: Background questionnaire

Background questionnaire for second language users of Chinese

Age:______________ Gender:______________

Native language(s):____

If you are a STUDENT:

University/Institute:____

Level of Study: ☐ Undergraduate ☐ Postgraduate ☐ Other ____________

Major:_____________________________

If you are NOT a student:

Education: ☐ Bachelor’s Degree ☐ Master’s Degree ☐ Other ____________

Field of Work:_____________________________

1. How long have you been learning Chinese? __________ month(s)

2. Have you ever studied Chinese before entering the university? Yes/No

If yes, please specify where you studied Chinese: _________________

And how long? __________ month(s)

3. Have you ever studied Chinese in a Chinese speaking country? Yes/No

If yes, please specify the country in which you have studied:_________________

And how long? __________ month(s)

4. Have you ever stayed in a Chinese speaking country? Yes/No

If yes, how long? __________ month(s)

5. Do you watch TV programmes, films or videos in Chinese? Yes/No

If yes, how long per week? Approximately ____________ hour (s)

6. Do you listen to Chinese songs? Yes/No

If yes, how long per week? Approximately ____________ hour (s)

7. Do you read (magazines, newspapers, and books etc.) in Chinese? Yes/No

If yes, how long per week? Approximately ______________ hour (s)

8. Do you write (course assignments, diaries, and emails etc.) in Chinese? Yes/No

If yes, how long per week? Approximately ______________ hour (s)

9. How often do you interact with (either in speaking or in writing) native Chinese speakers? _______________

Background questionnaire for first language users of Chinese

Age:______________ Gender:______________

If you are a STUDENT:

University/Institute:____

Level of Study: ☐ Undergraduate ☐ Postgraduate ☐ Other ____________

Major:_____________________________

If you are NOT a student:

Education: ☐ Bachelor’s Degree ☐ Master’s Degree ☐ Other ____________

Field of Work:_____________________________

1. How often do you watch TV programmes, films or videos in Chinese?
Approximately ____________ hour (s) per week.

2. How often do you you listen to Chinese songs, podcast or radio programmes?
Approximately ____________ hour (s) per week.

3. How often do you read (magazines, newspapers or books etc.) in Chinese?
Approximately ______________ hour (s) per week.

4. How often do you write (diaries, blogs or emails etc.) in Chinese?
Approximately ______________ hour (s) per week.

5. How often do you interact with other Chinese speakers? ______________________

6. What foreign language(s) do you speak? ____________________________________

7. How often do you watch TV programmes, films or vidoes in the language(s)?

Approximately ____________ hour (s) per week.

8. How often do you you listen to songs, podcast or radio programmes in the language(s)?

Approximately ____________ hour (s) per week.

9. How often do you read (magazines, newspapers or books etc.) in the language(s)?

Approximately ______________ hour (s) per week.

10. How often do you write (diaries, blogs or emails etc.) in the language(s)?

Approximately ______________ hour (s) per week.

11. How often do you interact with others in the language(s)?______________________

12. How long have you stayed in the UK?

______________years, and______________months

Appendix B: Writing prompts

Argumentative Writing Task One

Write about the following topic:

如今，几乎人人都有手机。一些人认为，手机给生活带来了便利。另一些人认为，我们每天花太多时间看手机，也因此带来了很多问题。讨论这两种观点，并对使用手机给出你的看法。

Almost everyone has a mobile phone nowadays. Some people think that a mobile phone brings convenience to our life. Others say we spend too much time on them, and there are many problems associated with them. Discuss both views and give your own opinion on the massive use of mobile phones.

Give reasons for your answers and include any relevant examples from your own knowledge or experience. You have 30 minutes to finish the task. There is no word limit for the essay, but you should write a complete essay.

Argumentative Writing Task Two

Write about the following topic:

一些人认为，学外语的就是为了去国外旅游或工作。另一些人认为，去国外旅游或工作并不是学外语的唯一理由。讨论这两种观点，并对为什么学习外语给出你的看法。

Some people think that the only reason for learning a foreign language is in order to travel to or work in a foreign country. Other say these are not the only reasons why someone should learn a foreign language. Discuss both views and give your own opinion on the motivation of learning a foreign language.

Give reasons for your answers and include any relevant examples from your own knowledge or experience. You have 30 minutes to finish the task. There is no word limit for the essay, but you should write a complete essay.

Narrative Writing Task One

Write about the following topic:

回忆童年的时候，我们经常会想起一些对我们产生重要影响的事。叙述一件发生在你童年时的重要事。请写出这件事的内容，这件事发生的时间、地点，你看见了什么、做了什么，还有哪些人参与，并谈谈这件事对你的影响。

When we reflect upon our childhood, we often come back to a few key events that had a major impact on us. Write a story about one of those defining events from your childhood. You may include what the event was, when and where it happened, what you saw and did, who else was involved, and explained how it impacted on you.

Be sure to develop your story by including relevant examples and specific details. You have 30 minutes to finish the task. There is no word limit for the essay, but you should write a complete essay.

Narrative Writing Task Two

Write about the following topic:

第一次尝试新事物会令我们感到既兴奋又紧张。叙述一个让你难忘的“第一次”。请写出你做了什么，事情发生在什么时间、什么地点，还有哪些人参与，

并谈谈这次经历给你的感受。

The first time we try something new can be both exciting and nervous. Write a story about a memorable “first” in your life. You may include what you did, when and where you did it, who else was involved, and explain how you felt about the experience.

Be sure to develop your story by including relevant examples and specific details. You have 30 minutes to finish the task. There is no word limit for the essay, but you should write a complete essay.

Appendix C: Chinese Proficiency Test

选词填空。共有 9 篇文章。请在 30 分钟内完成。

Choose the best word/phrases for each blank. There are 9 passages. You have 30 minutes to finish the task. Please do not use dictionaries or discuss with others.

短文 1：

九月五日是小红的 1.____。大家都 2.____她庆祝。她 3.____到很多礼物。所以，她今天非常 4.____。她希望明年能 5.____德国去玩。

1. A. 生活　B. 生日　C. 星期日　　Answer:__________

2. A. 帮　B. 让　C. 对　　Answer:__________

3. A. 收　B. 寄　C. 借　　Answer:__________

4. A. 热闹　B. 舒服　C. 高兴　　Answer:__________

5. A. 到　B. 去　C. 来　　Answer:__________

短文 2：

王小姐 1.____旅游很感兴趣。她放假的时候 2.____出国玩。她跟我说她第一次旅行时，在旅馆 3.____丢了行李。她紧张 4.____好几天，三天后才找到行李。从那一次开始，她旅行的时候就更 5.____自己的东西了。

1. A. 跟　B. 和　C. 对　　Answer:__________

2. A. 多少　B. 可能　C. 常常　　Answer:__________

3. A. 做　B. 弄　C. 走　　Answer:__________

4. A. 着　B. 了　C. 得　　Answer:__________

5. A. 小心　B. 知道　C. 认识　　Answer:__________

短文 3:

我喜欢到附近很大的一家 1.____买菜。那里的东西不怎么便宜，2.____比较新鲜，吃起来很放心。听说今天八点那里有特别活动，所以我早上六点 3.____起床了。到了那里以后，我买了好多 4.____我爱吃的饼干和一些青菜。不过 5.____付钱的人太多了，等的时间比买东西的时间还久。

1. A. 药店　B. 书店　C. 超市　　Answer:__________

2. A. 但是　B. 或是　C. 还是　　Answer:__________

3. A. 再　B. 才　C. 就　　Answer:__________

4. A. 支　B. 包　C. 口　　Answer:__________

5. A. 挂号　B. 点菜　C. 排队　　Answer:__________

短文 4:

昨晚张太太的肚子很痛，感觉小孩快要生出来了，所以她 1.____送进了医院。张先生也 2.____张太太到医院去，他又高兴又紧张，因为他就要 3.____爸爸了。五个小时以后，小孩生出来了，是一个可爱的小女孩。张先生和张太太 4.____这个孩子想了一个名字，叫“张美乐”，5.____她将来“美丽”又“快乐”。

1. A. 把　B. 能　C. 被　D. 会　　Answer:__________

2. A. 帮　B. 害　C. 陪　D. 请　　Answer:__________

3. A. 当　B. 成　C. 装　D. 弄　　Answer:__________

4. A. 教　B. 为　C. 让　D. 养　　Answer:__________

5. A. 觉得　B. 要求　C. 决定　D. 希望　　Answer:__________

短文 5：

小陈和小李是大学 1.____的好朋友，大学毕业以后，他们已经 2.____ 年没见面了。有一次小陈 3.____想起和小李一起读书的事，他很想知道小李现在过得怎么样，所以他找出小李以前寄给他的信，4.____着信上的地址去找小李。可是小李的邻居说小李一家人早就搬走了。小陈很难过，因为他 5.____以后都见不到小李了。

1. A. 时代 B. 时间 C. 年级 D. 年纪 Answer:__________
2. A. 多久 B. 好几 C. 很长 D. 很久 Answer:__________
3. A. 忽然 B. 总是 C. 随时 D. 从来 Answer:__________
4. A. 跟 B. 接 C. 写 D. 照 Answer:__________
5. A. 能够 B. 只有 C. 可能 D. 要是 Answer:__________

短文 6：

王小姐喜欢跳舞，她认为让身体随着音乐摆动不是一件困难的事，所以 1.____是老人还是小孩，1.____可以将跳舞当做日常的运动。王小姐平时的工作压力很大，她利用跳舞，2.____自己的压力，忘掉工作上的不愉快。3.____三年的舞蹈训练，她的身体更加健康，她还参加舞蹈比赛，4.____冠军。她热爱跳舞，5.____工作再怎么忙，她还是会到舞蹈教室报到。

1. A. 不论…都… B. 不但…也… C. 虽然…但是… D. 由于…因此… Answer:__________
2. A. 防止 B. 缩短 C. 减轻 D. 缺少 Answer:__________
3. A. 根据 B. 经过 C. 按照 D. 至于 Answer:__________
4. A. 得了 B. 受了 C. 接了 D. 收了 Answer:__________
5. A. 否则 B. 另外 C. 免得 D. 就算 Answer:__________

短文 7：

从 15 世纪开始，欧洲便出现了手抄报纸，报导政治、战争、市场等消息。1.____它突破了传统私人信件的形式，但传播的范围还很小，2.____要等到印刷技术进步，报纸才 3.____大量发行。报纸之所以普及，也跟近代商业的发展 4.____关系。经济的快速发展使得不同地区、不同国家间的关系更为密切，人们需要互相了解，信息需要快速传播，人们对报纸的依赖也就 5.____加深，报业因此蓬勃发展。

1. A. 由于　B. 尽管　C. 既然　D. 然而　　Answer:__________
2. A. 一向　B. 一时　C. 一旦　D. 一直　　Answer:__________
3. A. 得以　B. 加以　C. 予以　D. 足以　　Answer:__________
4. A. 受不了　B. 脱不了　C. 去不了　D. 起不了　　Answer:__________
5. A. 及其　B. 相当　C. 日益　D. 随后　　Answer:__________

短文 8：

目前许多中小学都鼓励学生利用电脑进行学习，但是成人的电脑让孩子使用，1.____会有各种顾虑，不是担心重要文件被孩子删掉，就是怕孩子乱逛网站造成电脑中毒。当然，对父母来说，电脑中毒事小，2.____孩子逛到色情或赌博网站，那就更糟糕了。正因为儿童使用电脑的比例逐年增加，有厂商看好这 3.____趋势，推出专为儿童设计的电脑，4.____内部安装了各种学习软件，4.____可让父母监控孩子上网的情形，一上市便广受好评，创造出亮眼的 5.____，厂商也因此大捞一笔。

1. A. 居然　B. 未必　C. 难免　D. 哪怕　　Answer:__________
2. A. 恐怕　B. 万一　C. 甚至　D. 而且　　Answer:__________
3. A. 股　B. 款　C. 片　D. 串　　Answer:__________
4. A. 虽然…但…　B. 除非…才…　C. 不论…也…　D. 不仅…还…
　　Answer:__________
5. A. 客户　B. 销售　C. 业绩　D. 景气　　Answer:__________

短文 9：

每每回顾生命历程，总能发现某些当年所 1.____的意外和经历的曲折，后来好像都转为一种能量和养分。若 2.____这些意外与曲折，我似乎就不会在人生的路上与难得的人、事相遇；而这些人、那些事在时间渐渐 3.____一切后，只留下了由欢笑与泪水交织而成的一股暖意。曾经的悔恨和不满仿佛都已 4.____。我经常拿这些人与事和朋友分享，他们鼓励我将之化为文字，5.____自己带着回忆走进棺材，5.____集结成册，当我有一天什么都不记得时，至少还有人帮我记得这些人、那些事。

1. A. 盲从　B. 遵循　C. 叫好　D. 遭遇　　Answer:__________

2. A. 未　B. 甭　C. 勿　D. 非　　Answer:__________

3. A. 冲淡　B. 裁剪　C. 消耗　D. 剥削　　Answer:__________

4. A. 香消玉殒　B. 云消雾散　C. 卷土重来　D. 魂飞魄散

Answer:__________

5. A. 为了…因此…　B. 就算…也能…　C. 总得…进而…　D. 与其…不如…

Answer:__________

Appendix D: Information sheet

Participant Information Sheet for Second Language (L2) Users of Chinese

Title of Study: Writing in a non-alphabetic language using a keyboard: Behaviours, cognitive activities and text quality

Department: Department of Communication, Culture and Media, Institute of Education, University College London

Name and contact details of the researcher: Xiaojun Lu (x.lu.16@ucl.ac.uk)

Supervisor: Dr Andrea Révész (a.revesz@ucl.ac.uk)

You are being invited to take part in a research project. Before you decided it is important for you to understand why the research I being done and what participation will involve. Please take time to read the following information carefully and discuss it with others if you wish. Ask if there is anything that is not clear or if you would like more information. Take time to decide whether or not you wish to take part.

The aim of this project is to examine the cognitive processes and behaviours of L2 users of Chinese while carrying out writing tasks, and the ways in which the cognitive processes and behaviours relate to the quality of the written texts, with data from L1 users of Chinese as the baseline. An additional aim is to assess whether the writing processes and text quality are influenced by types of writing tasks. The results of the study will inform previous research on second language writing and help teachers to facilitate L2 Chinese writing effectively.

Around 40 L2 users of Chinese will be invited to participate. All of them shall **be able to tell stories and express ideas on daily topics in written forms using the Pinyin input method on a computer**. Data will be collected from the same number of L1 users of Chinese who are comfortable with writing essays with the Pinyin input method so as to form the baseline.

It is up to you to decide whether or not to take part. If you do decide to take part you will be given this information sheet to keep and be asked to sign a consent form. You are free to withdraw at any time without giving a reason and without it affecting any benefits that you are entitled to. If you decide to withdraw, any data collected from you will be destroyed. I hope that if you do choose to be involved then you will find it a valuable experience.

If you decide to participate, I will ask you to take part in two sessions. In both sessions, you will be asked to complete two writing tasks using a computer. Your keystrokes and mouse movements will be recorded by the computer while you are writing. You will also be asked to participate in a follow-up interview, in which I will ask you to watch the recording of your last writing task in the second session and describe what you were thinking while writing. Your response will be video recorded. You will also be asked to complete a brief questionnaire about yourself (in the first session) and a cloze test (in the second session). All activities will be separated by short breaks.

It will take **approximately 4 hours in total**, with 1.5 hours in the first session and 2.5 hours in the second session. In return for your participation, I am able to offer you **a £25 Amazon gift card and up to 3 hours of free private tutoring of Chinese**. I will also share the overall results of the study with you. It is also a good opportunity to practice your essay writing skills in Chinese outside the classroom, and I can **provide feedback** on your essays if you want.

All the information that I collect about you during the course of the research will be kept strictly confidential. You will not be able to be identified in any ensuing reports or publications. However, there are **limits to confidentiality**, which include:

Please note that assurances on confidentiality will be strictly adhered to unless evidence of wrongdoing or potential harm is uncovered. In such cases the University may be obliged to contact relevant statutory bodies/agencies.

Please note that confidentiality will be maintained as far as it is possible, unless during our conversation I hear anything which makes me worried that someone might be in danger of harm, I might have to inform relevant agencies of this.

Please note that confidentiality may not be guaranteed; due to the limited size of the participant sample.

Confidentiality will be respected subject to legal constraints and professional guidelines.

Confidentiality will be respected unless there are compelling and legitimate reasons for this to be breached. If this was the case we would inform you of any decisions that might limit your confidentiality.

Confidentiality may be limited and conditional and the researcher has a duty of care to report to the relevant authorities possible harm/danger to the participant or others.

The overall results of the study will also be shared with the participants of the study. The results may also be presented at professional conferences and in research publications.

Thank you very much for taking the time to read this information sheet. If you have any further questions before you decide whether to take part, please reach me at **x.lu.16@ucl.ac.uk**. If you have any complaints about the project, please contact my Supervisor, Dr Andrea Révész at **a.revesz@ucl.ac.uk**.

Participant Information Sheet for First Language (L1) Users of Chinese

Title of Study: Writing in a non-alphabetic language using a keyboard: Behaviours, cognitive activities and text quality

Department: Department of Communication, Culture and Media, Institute of Education, University College London

Name and contact details of the researcher: Xiaojun Lu (x.lu.16@ucl.ac.uk)

Supervisor: Dr Andrea Révész (a.revesz@ucl.ac.uk)

You are being invited to take part in a research project. Before you decided it is important for you to understand why the research I being done and what participation will involve. Please take time to read the following information carefully and discuss it with others if you wish. Ask if there is anything that is not clear or if you would like more information. Take time to decide whether or not you wish to take part.

The aim of this project is to examine the cognitive processes and behaviours of L2 users of Chinese while carrying out writing tasks, and the ways in which the cognitive processes and behaviours relate to the quality of the written texts, with data from L1 users of Chinese as the baseline. An additional aim is to assess whether the writing processes and text quality are influenced by types of writing tasks. The results of the study will inform previous research on second language writing and help teachers to facilitate L2 Chinese writing effectively.

Around 40 L2 users of Chinese will be invited to participate. All of them shall be able to tell stories and express ideas on daily topics in written forms using the Pinyin input method on a computer. Data will be collected from **the same number of L1 users of Chinese** who **are comfortable with writing essays with the Pinyin input method** so as to form the baseline.

It is up to you to decide whether or not to take part. If you do decide to take part you will be given this information sheet to keep and be asked to sign a consent form. You are free to withdraw at any time without giving a reason and without it affecting any benefits that you are entitled to. If you decide to withdraw, any data collected from you will be destroyed. I hope that if you do choose to be involved then you will find it a valuable experience.

If you decide to participate, I will ask you to take part in two sessions. In both sessions, you will be asked to complete two writing tasks using a computer. Your keystrokes and mouse movements will be recorded by the computer while you are writing. You will also be asked to participate in a follow-up interview, in which I will ask you to watch the recording of your last writing task in the second session and describe what you were thinking while writing. Your response will be video recorded. You will also be asked to complete a brief questionnaire about yourself (in the first session) and a cloze test (in the second session). All activities will be separated by short breaks.

It will take **<u>approximately 3 hours in total</u>**, with 1 hour in the first session and 2 hours in the second session. In return for your participation, I am able to offer you **<u>a £20 Amazon gift card</u>**. I will also share the overall results of the study with you.

All the information that I collect about you during the course of the research will be kept strictly confidential. You will not be able to be identified in any ensuing reports or publications. However, there are **<u>limits to confidentiality</u>**, which include:

Please note that assurances on confidentiality will be strictly adhered to unless evidence of wrongdoing or potential harm is uncovered. In such cases the University may be obliged to contact relevant statutory bodies/agencies.

Please note that confidentiality will be maintained as far as it is possible, unless during our conversation I hear anything which makes me worried that someone might be in danger of harm, I might have to inform relevant agencies of this.

Please note that confidentiality may not be guaranteed; due to the limited size of the participant sample.

Confidentiality will be respected subject to legal constraints and professional guidelines.

Confidentiality will be respected unless there are compelling and legitimate reasons for this to be breached. If this was the case we would inform you of any decisions that might limit your confidentiality.

Confidentiality may be limited and conditional and the researcher has a duty of care to report to the relevant authorities possible harm/danger to the participant or others.

The overall results of the study will also be shared with the participants of the study. The results may also be presented at professional conferences and in research publications.

Thank you very much for taking the time to read this information sheet. If you have any further questions before you decide whether to take part, please reach me at **x.lu.16@ucl.ac.uk**. If you have any complaints about the project, please contact my Supervisor, Dr Andrea Révész at **a.revesz@ucl.ac.uk**.

Appendix E: Consent form

CONSENT FORM RESEARCH STUDIES

Please complete this form after you have read the Information Sheet and/or listened to an explanation about the research.

Title of Study: Writing in a non-alphabetic language using a keyboard: Behaviours, cognitive activities and text quality

Department: Department of Culture, Communication and Media, Institute of Education, University College London

Name and Contact Details of the Researcher: Xiaojun Lu (x.lu.16@ucl.ac.uk)

Supervisor: Dr Andrea Révész (a.revesz@ucl.ac.uk)

Name and Contact Details of the UCL Data Protection Officer: Spenser Crouch (s.crouch@ucl.ac.uk)

This study has been approved by the UCL Research Ethics Committee since March 6th, 2017.

Thank you for considering taking part in this research. The person organising the research must explain the project to you before you agree to take part. If you have any questions arising from the Information Sheet or explanation already given to you, please ask the researcher before you decide whether to join in. You will be given a copy of this Consent Form to keep and refer to at any time.

I confirm that I understand that by ticking/initialling each box below I am consenting to this element of the study. I understand that it will be assumed that unticked/initialled boxes means that I DO NOT consent to that part of the study.

I understand that by not giving consent for any one element that I may be deemed ineligible for the study.

		Tick Box
1.	I confirm that I have read and understood the Information Sheet for the above study. I have had an opportunity to consider the information and what will be expected of me. I have also had the opportunity to ask questions which have been answered to my satisfaction and would like to take part in - a questionnaire survey - four computer-delivered writing activities - an individual interview	
2.	I understand that I will be able to withdraw my data up to 2 weeks after the second session is over.	
3.	I consent to the processing of my personal information (i.e., gender, age, language background, major/area of working, length of using the second language and frequencies of using the second language) for the purposes explained to me. I understand that such information will be handled in accordance with all applicable data protection legislation.	
4.	I understand that all personal information will remain confidential and that all efforts will be made to ensure I cannot be identified. I understand that my data gathered in this study will be stored anonymously and securely. It will not be possible to identify me in any publications.	
5.	I understand that my participation is voluntary and that I am free to withdraw at any time without giving a reason, [*without the care I receive or my legal rights being affected*]. I understand that if I decide to withdraw, any personal data I have provided up to that point will be deleted unless I agree otherwise.	
6.	I understand the potential risks of participating and the support that will be available to me should I become distressed during the course of the research.	
7.	I understand the direct/indirect benefits of participating.	
8.	I understand that the data will not be made available to any commercial organisations but is solely the responsibility of the researcher(s) undertaking this study.	
9.	I understand that I will not benefit financially from this study or from any possible outcome it may result in in the future.	
10.	I agree that my psuedonymised research data may be used by others for future research. [No one will be able to identify you when this data is shared.]	

11.	I understand that the information I have submitted will be published as a report and I wish to receive a copy of it. Yes/No	
12.	I consent to my interview being audio/video recorded and understand that the recordings will be stored anonymously and using password-protected software for research purposes.	
13.	I hereby confirm that I understand the inclusion criteria as detailed in the Information Sheet and explained to me by the researcher.	
14.	I am aware of who I should contact if I wish to lodge a complaint.	
15.	I voluntarily agree to take part in this study.	

If you would like your contact details to be retained so that you can be contacted in the future by UCL researchers who would like to invite you to participate in follow up studies to this project, or in future studies of a similar nature, please tick the appropriate box below.

	Yes, I would be happy to be contacted in this way	
	No, I would not like to be contacted	

________________________ ________________ __________________

Name of participant Date Signature

________________________ ________________ __________________

Researcher Date Signature

Appendix F: Assessment Criteria for the Writing Tasks

Assessment Criteria for Argumentative Essays

Band	Task response	Coherence and cohesion	Lexical resource	Grammatical range and accuracy
9	· fully addresses all parts of the task · presents a fully developed position in answer to the question with relevant, fully extended and well supported ideas	· uses cohesion in such a way that it attracts no attention · skilfully manages paragraphing	· uses a wide range of vocabulary with very natural and sophisticated control of lexical features; rare minor errors occur only as 'slips'	· uses a wide range of structures with full flexibility and accuracy; rare minor errors occur only as 'slips'
8	· sufficiently addresses all parts of the task · presents a well-developed response to the question with relevant, extended and supported ideas	· sequences information and ideas logically · manages all aspects of cohesion well · uses paragraphing sufficiently and appropriately	· uses a wide range of vocabulary fluently and flexibly to convey precise meanings · skilfully uses uncommon lexical items but there may be occasional inaccuracies in word choice and collocation · produces rare errors in Chinese characters and/or word formation	· uses a wide range of structures · the majority of sentences are error-free · makes only very occasional errors or inappropriacies
7	· addresses all parts of the task · presents a clear position throughout the response · presents, extends and supports main ideas, but there may be a tendency to over-generalise and/or supporting ideas may lack focus	· logically organises information and ideas; there is clear progression throughout · uses a range of cohesive devices appropriately although there may be some under-/over-use · presents a clear central topic within each paragraph	· uses a sufficient range of vocabulary to allow some flexibility and precision · uses less common lexical items with some awareness of style and collocation · may produce occasional errors in word choice, Chinese characters and/or word formation	· uses a variety of complex structures · produces frequent error-free sentences · has good control of grammar and punctuation but may make a few errors
6	· addresses all parts of the task although some parts may be more fully covered than others · presents a relevant position although the	· arranges information and ideas coherently and there is a clear overall progression · uses cohesive devices effectively, but cohesion within and/or between sentences	· uses an adequate range of vocabulary for the task · attempts to use less common vocabulary but with some inaccuracy	· uses a mix of simple and complex sentence forms · makes some errors in grammar and punctuation but they rarely reduce

	conclusions may become unclear or repetitive • presents relevant main ideas but some may be inadequately developed/unclear	may be faulty or mechanical • may not always use referencing clearly or appropriately • uses paragraphing, but not always logically	• makes some errors in Chinese characters and/or word form, but they do not impede communication	communication
5	• addresses the task only partially; the format may be inappropriate in places • expresses a position but the development is not always clear and there may be no conclusions drawn • presents some main ideas but these are limited and not sufficiently developed; there may be irrelevant details	• presents information with some organisation but there may be a lack of overall progression • makes inadequate, inaccurate or over-use of cohesive devices • may be repetitive because of lack of referencing and substitution • the use of paragraphing may be confusing sometimes	• uses a limited range of vocabulary, but this is minimally adequate for the task • may make noticeable errors in Chinese characters and/or word formation that may cause some difficulty for the reader	• uses only a limited range of structures • attempts complex sentences but these tend to be less accurate than simple sentences • may make frequent grammatical errors and punctuation may be faulty; errors can cause some difficulty for the reader
4	• responds to the task only in a minimal way or the answer is tangential; the format may be inappropriate • presents a position but this is unclear • presents some main ideas but these are difficult to identify and may be repetitive, irrelevant or not well supported	• presents information and ideas but these are not arranged coherently and there is no clear progression in the response • uses some basic cohesive devices but these may be inaccurate or repetitive • the use of paragraphing is largely confusing	• uses only basic vocabulary which may be used repetitively or which may be inappropriate for the task • has limited control of word formation and/or Chinese characters; errors may cause strain for the reader	• uses only a very limited range of structures with only rare use of subordinate clauses • some structures are accurate but errors predominate, and punctuation is often faulty
3	• does not adequately address any part of the task • does not express a clear position • presents few ideas, which are largely undeveloped or irrelevant	• does not organise ideas logically • may use a very limited range of cohesive devices, and those used may not indicate a logical relationship between ideas	• uses only a very limited range of words and expressions with very limited control of word formation and/or characters • errors may severely distort the message	• attempts sentence forms but errors in grammar and punctuation predominate and distort the meaning

2	• barely responds to the task • does not express a position • may attempt to present one or two ideas but there is no development	• has very little control of organisational features	• uses an extremely limited range of vocabulary; essentially no control of word formation and/or Chinese characters	• cannot use sentence forms except in memorised phrases
1	• answer is completely unrelated to the task	• fails to communicate any message	• can only use a few isolated words	• cannot use sentence forms at all

议论文写作评分标准

分数	扣题程度	连贯与衔接	词汇丰富度	句法多样度及准确度
9	· 能够全面回应题目中的各项要求。 · 能够就题目中的问题展开充分的讨论，并提出相关论据予以支持。	· 衔接手段运用自如，行文连贯。 · 分段合理。	· 用词丰富，且能够正确、恰当地使用高级词汇；极少出现错误，偶有出现仅属打字错误。	· 能够准确、灵活地使用各类句法结构；极少出现错误，偶有出现仅属笔误。
8	· 能够充分回应题目中的各项要求。 · 能够就题目中的问题进行较为充分的讨论，并提出相关论据予以支持。	· 论点组织清晰，符合逻辑。 · 衔接手段运用得当。 · 分段合理。	· 用词丰富，达意准确。 · 能够熟练地使用高级词汇，但在用词及搭配方面有时偶尔出现错误。 · 极少出现汉字以及/或者构词方面的错误。	· 运用丰富多样的句法结构。 · 大多数句子准确无误。 · 偶尔出现句法结构错误或使用不当的情况。
7	· 能够回应题目中的各项要求。 · 有一个明确的主论点贯穿全文。 · 能够提出相关论据支持该论点，但有时会出现以偏概全以及/或者论据缺乏重点的情况。	· 论点组织条理清晰，行文流畅。 · 衔接手段使用基本恰当，但有时会出现衔接手段使用失当的情况。 · 各段落中心明确。	· 用词丰富，且基本准确恰当。 · 能够使用一些高级词汇，对语体及搭配有一定了解。 · 在词汇选用、汉字以及/或者构词方面偶有错误。	· 能够运用复杂的句法结构。 · 多数句子准确无误。 · 对句法及标点符号掌握较好，但有时会出现少许错误。
6	· 能够回应题目中的各项要求，但对某些部分的论证不够充分。 · 观点总体明确，但有些分论点不甚清晰，甚至出现重复。 · 能够提出多个分论点，但某些分论点论证不够充分。	· 论点组织连贯，行文较为流畅 · 能够有效地使用衔接手段，但句内以及/或者句间的衔接有时会过于机械，甚至出现错误。 · 有时会出现指代不明的情况。 · 可能存在段落之间逻辑不明的情况。	· 用词比较丰富。 · 尝试使用高级词汇，但存在使用不当的情况。 · 在汉字和构词方面有错误，但不影响对文章意思的理解。	· 能够综合使用简单与复杂的句式。 · 在句法及标点符号方面有一些错误，但不影响对文章意思的理解。
5	· 仅回应了题目中的部分要求；写作格式有时不甚恰当。 · 全文有一个主论点，但论证的过程不够清晰，无法得出明确的结论。 · 能够提出一些分论点，但讨论不够充分；	· 信息呈现有一定条理，但总体来说行文不够流畅。 · 衔接手段使用不足或者失当。 · 指代不够明确；缺乏以不同方式指代同一事物的能力，存在较多重复的部分。	· 用词有限，但能够基本完成题目要求。 · 在汉字和构词方面存在明显错误，且会造成一定的阅读困难。	· 句法结构使用有限。 · 尝试使用复杂句，但准确度常不及简单句。 · 经常出现句法及标点符号使用的错误，并对会造成一定的阅读困难。

	有时会出现无关细节。	· 有时分段混乱。		
4	· 仅在最低限度上回应了题目要求，所写内容与题目相关性不大；写作格式有时不甚恰当。 · 能够提出一个主论点，但并不清晰。 · 试图提出一些分论点，但未能清晰的表达出来；分论点之间重复，有些论点与题目不相关或缺乏论据支持。	· 论点组织不连贯，行文不流畅。 · 使用了一些基本的衔接手段，但有时出现使用不当或重复使用的情况。 · 分段混乱。	· 用词有限，词汇使用重复，且有些词汇与题目不相关。 · 对构词以及/或者汉字掌握有限；词汇错误会造成阅读困难。	· 仅能够使用有限的句法结构，偶尔尝试使用偏正复句。 · 部分句法结构使用正确，但错误占大多数，标点符号使用经常出错。
3	· 基本未能回应题目要求。 · 没有明确的主论点。 · 分论点数量很少且不切题，基本未能展开论证。	· 论点组织没有逻辑。 · 衔接手段使用非常有限，且基本使用不当。	· 用词非常有限，对构词以及/或者汉字的掌握也非常有限。 · 存在很多错误，严重影响读者对文章的理解。	· 尝试写出完整的句子，但存在很多句法和标点符号使用错误；基本无法传达意思。
2	· 基本未能回应题目要求。 · 没有主论点。 · 尝试提出少量分论点，但未能展开论证。	· 内容组织能力非常有限。	· 词汇使用及其有限，基本未能掌握构词以及/或者汉字。	· 除了使用预先背诵的短语外，无法写出完整的句子。
1	· 内容与题目无关。	· 无法传达任何信息。	· 仅能孤立地使用少数词汇。	· 无法写出句子。

Assessment Criteria for Narrative Essays

Band	Task response	Coherence and cohesion	Lexical resource	Grammatical range and accuracy
9	· fully addresses all parts of the task · tells a thorough and complete story in answer to the question with relevant and fully extended information	· uses cohesion in such a way that it attracts no attention	· uses a wide range of vocabulary with very natural and sophisticated control of lexical features; rare minor errors occur only as 'slips'	· uses a wide range of structures with full flexibility and accuracy; rare minor errors occur only as 'slips'
8	· sufficiently addresses all parts of the task · presents a well-developed story with relevant and extended information	· sequences information logically · manages all aspects of cohesion well	· uses a wide range of vocabulary fluently and flexibly to convey precise meanings · skilfully uses uncommon lexical items but there may be occasional inaccuracies in word choice and collocation · produces rare errors in Chinese characters and/or word formation	· uses a wide range of structures · the majority of sentences are error-free · makes only very occasional errors or inappropriacies
7	· addresses all parts of the task · tells a complete story but may lack detail /elaboration or have minor inconsistencies in its development	· logically organises information; there is clear progression throughout · uses a range of cohesive devices appropriately although there may be some under-/over-use	· uses a sufficient range of vocabulary to allow some flexibility and precision · uses less common lexical items with some awareness of style and collocation · may produce occasional errors in word choice, Chinese characters and/or word formation	· uses a variety of complex structures · produces frequent error-free sentences · has good control of grammar and punctuation but may make a few errors
6	· addresses all parts of the task although some parts may be more fully covered than others · tells a complete story with relevant information, some of which may be inadequately developed/unclear	· arranges information coherently and there is a clear overall progression · uses cohesive devices effectively, but cohesion within and/or between sentences may be faulty or mechanical · may not always use referencing clearly or	· uses an adequate range of vocabulary for the task · attempts to use less common vocabulary but with some inaccuracy · makes some errors in Chinese characters and/or word form, but they do not impede	· uses a mix of simple and complex sentence forms · makes some errors in grammar and punctuation but they rarely reduce communication

		appropriately	communication	
5	• addresses the task only partially; the format may be inappropriate in places • tells a relatively complete story but the development is not always clear • presents some information which is limited and not sufficiently developed; there may be irrelevant details	• presents information with some organisation but there may be a lack of overall progression • makes inadequate, inaccurate or over-use of cohesive devices • may be repetitive because of lack of referencing and substitution	• uses a limited range of vocabulary, but this is minimally adequate for the task • may make noticeable errors in Chinese characters and/or word formation that may cause some difficulty for the reader	• uses only a limited range of structures • attempts complex sentences but these tend to be less accurate than simple sentences • may make frequent grammatical errors and punctuation may be faulty; errors can cause some difficulty for the reader
4	• responds to the task only in a minimal way or the answer is tangential • tells a basic story but it is inconsistent in its development • presents some details but they are difficult to identify and may be repetitive or irrelevant	• presents information but these are not arranged coherently and there is no clear progression in the response • uses some basic cohesive devices but these may be inaccurate or repetitive	• uses only basic vocabulary which may be used repetitively or which may be inappropriate for the task • has limited control of word formation and/or Chinese characters; errors may cause strain for the reader	• uses only a very limited range of structures with only rare use of subordinate clauses • some structures are accurate but errors predominate, and punctuation is often faulty
3	• does not adequately address any part of the task • attempts to tell a story, but the story is largely list-like and with little narration • presents few ideas, which are largely scattered or irrelevant	• does not organise ideas logically • may use a very limited range of cohesive devices, and those used may not indicate a logical relationship between events	• uses only a very limited range of words and expressions with very limited control of word formation and/or characters • errors may severely distort the message	• attempts sentence forms but errors in grammar and punctuation predominate and distort the meaning
2	• barely responds to the task • may attempt to tell a story but it lacks elements of narrative • may attempt to present one or two ideas but there is no development	• has very little control of organisational features	• uses an extremely limited range of vocabulary; essentially no control of word formation and/or Chinese characters	• cannot use sentence forms except in memorised phrases

1	• answer is completely unrelated to the task	• fails to communicate any message	• can only use a few isolated words	• cannot use sentence forms at all

记叙文写作评分标准

分数	扣题程度	连贯与衔接	词汇丰富度	句法多样度及准确度
9	· 能够全面回应题目中的各项要求。 · 故事叙述完整，细节描述相关且充分。	· 衔接手段运用自如，行文连贯。	· 用词丰富，且能够正确、恰当地使用高级词汇；极少出现错误，偶有出现仅属打字错误。	· 能够准确、灵活地使用各类句法结构；极少出现错误，偶有出现仅属笔误。
8	· 能够充分回应题目中的各项要求。 · 故事叙述线索清晰，细节描述相关且较为充分	· 信息组织清晰，符合逻辑。 · 衔接手段运用得当。	· 用词丰富，达意准确。 · 能够熟练地使用高级词汇，但在用词及搭配方面有时偶尔出现错误。 · 极少出现汉字以及/或者构词方面的错误。	· 运用丰富多样的句法结构。 · 大多数句子准确无误。 · 偶尔出现句法结构错误或使用不当的情况。
7	· 能够回应题目中的各项要求。 · 故事叙述完整，但有时缺乏细节描述，或逻辑不明。	· 信息组织条理清晰，行文流畅。	· 用词丰富，且基本准确恰当。 · 能够使用一些高级词汇，对语体及搭配有一定了解。 · 在词汇选用、汉字以及/或者构词方面偶有错误。	· 能够运用复杂的句法结构。 · 多数句子准确无误。 · 对句法及标点符号掌握较好，但有时会出现少许错误。
6	· 能够回应题目中的各项要求，但对某些部分的叙述不过充分。 · 观点总体清晰，但有些分论点不甚清晰，甚至出现重复。 · 故事叙述完整，但某些细节展开不够。	· 信息组织连贯，行文较为流畅 · 能够有效地使用衔接手段，但句内以及/或者句间的衔接有时会过于机械，甚至出现错误。 · 有时会出现指代不明的情况。	· 用词比较丰富。 · 尝试使用高级词汇，但存在使用不当的情况。 · 在汉字和构词方面有错误，但不影响对文章意思的理解。	· 能够综合使用简单与复杂的句式。 · 在句法及标点符号方面有一些错误，但不影响对文章意思的理解。
5	· 仅回应了题目中的部分要求；写作格式有时不甚恰当。 · 故事叙述较为完整，但对于故事发展的描述不甚清晰。 · 能够写出一些细节，但描述不够充分；	· 信息呈现有一定条理，但总体来说行文不够流畅。 · 衔接手段使用不足或者失当。 · 指代不够明确；缺乏以不同方式指代同一事物的能力，存在较多重复的部分。	· 用词有限，但能够基本完成题目要求。 · 在汉字和构词方面存在明显错误，且会造成一定的阅读困难。	· 句法结构使用有限。 · 尝试使用复杂句，但准确度常不及简单句。 · 经常出现句法及标点符号使用的错误，并对会造成一定的阅读困难。

	有时还会出现无关细节。			
4	· 仅在最低限度上回应了题目要求，叙述内容与题目关联不大。 · 故事叙述简单，逻辑不明。 · 试图写出一些细节，但未能清晰表达出来；存在细节重复的情况，某些细节与题目无关。	· 信息组织不连贯，行文不流畅。 · 使用了一些基本的衔接手段，但有时出现使用不当或重复使用的情况。	· 用词有限，词汇使用重复，且有些词汇与题目不相关。 · 对构词以及/或者汉字掌握有限；词汇错误会造成阅读困难。	· 仅能够使用有限的句法结构，偶尔尝试使用偏正复句。 · 部分句法结构使用正确，但错误占大多数，标点符号使用经常出错。
3	· 基本未能回应题目要求。 · 试图叙述故事，但多为事实罗列，缺乏叙述文的特征。 · 仅有一些零散的细节，且基本与题目无关。	· 信息组织没有逻辑。 · 衔接手段使用非常有限，且基本使用不当。	· 用词非常有限，对构词以及/或者汉字的掌握也非常有限。 · 存在很多错误，严重影响读者对文章的理解。	· 尝试写出完整的句子，但存在很多句法和标点符号使用错误；基本无法传达意思。
2	· 基本未能回应题目要求。 · 文章未包含记叙文应有的要素。 · 未能对细节进行展开。	· 内容组织能力非常有限。	· 词汇使用及其有限，基本未能掌握构词以及/或者汉字。	· 除了使用预先背诵的短语外，无法写出完整的句子。
1	· 内容与题目无关。	· 无法传达任何信息。	· 仅能孤立地使用少数词汇。	· 无法写出句子。

Appendix G: Full descriptive statistics for writing behaviours in Chinese writing

Table 1 Speed fluency by stage in Chinese writing

	L2 Chinese (N = 32)					L1 Chinese (N = 32)				
	Stage	M	SD	95% CI low	95% CI up	Stage	M	SD	95% CI low	95% CI up
Production rate	1	46.80	13.33	44.47	49.13	1	69.66	21.43	73.40	65.92
	2	43.91	13.06	41.55	46.17	2	63.41	16.13	60.62	66.21
	3	42.25	11.59	40.24	44.26	3	64.13	18.92	60.85	67.41
	4	46.00	19.42	42.64	49.36	4	65.86	17.53	62.82	68.89
	5	43.09	35.99	36.86	49.32	5	57.30	13.92	54.89	59.71
	Total	43.17	8.74	41.70	44.70	Total	63.28	11.12	61.35	65.21
P-burst length	1	4.51	1.86	4.18	4.83	1	5.73	2.72	6.23	5.23
	2	4.29	1.91	3.96	4.62	2	5.90	2.94	5.36	6.44
	3	4.53	2.20	4.14	4.92	3	6.04	3.07	5.49	6.59
	4	4.46	1.66	4.17	4.75	4	5.98	3.25	5.39	6.58
	5	3.96	1.73	3.66	4.26	5	5.50	2.89	4.99	6.02
	Total	4.36	1.39	4.12	4.60	Total	5.71	1.64	5.42	6.00

Note: Production rate = Number of characters produced divided by active writing time, P-burst length = Number of characters between two pauses

Table 2 Pause frequency and duration by location and stage in Chinese writing

	L2 Chinese (N = 32)					L1 Chinese (N = 32)				
	Stage	*M*	*SD*	95% *CI* low	95% *CI* up	Stage	*M*	*SD*	95% *CI* low	95% *CI* up
Frequency (number of pauses per minute)										
Total	1	3.85	1.26	3.63	4.07	1	3.15	1.11	3.34	2.96
	2	4.54	1.17	4.34	4.74	2	3.69	1.05	3.51	3.87
	3	4.44	1.15	4.24	4.64	3	3.66	.95	3.49	3.82
	4	4.30	1.21	4.09	4.51	4	3.54	1.10	3.34	3.73
	5	3.74	1.05	3.65	3.92	5	3.45	.98	3.28	3.62
	Total	4.17	.85	4.02	4.32	Total	3.50	.70	3.38	3.62
Ww	1	.16	.17	0.13	0.19	1	.06	.12	.08	.04
	2	.21	.23	0.17	0.25	2	.07	.14	.04	.09
	3	.20	.22	0.16	0.24	3	.07	.14	.05	.10
	4	.14	.17	0.11	0.17	4	.05	.12	.03	.07
	5	.14	.19	0.11	0.17	5	.06	.14	.03	.08
	Total	.17	.10	.15	.19	Total	.06	.07	.05	.07
Bpc	1	.48	.57	0.38	0.58	1	.12	.24	.16	.08
	2	.60	.70	0.48	0.72	2	.20	.27	.15	.25
	3	.53	.56	0.43	0.63	3	.15	.23	.11	.19
	4	.49	.59	0.39	0.59	4	.17	.24	.13	.21
	5	.35	.44	0.27	0.43	5	.11	.21	.07	.14
	Total	.49	.49	.41	.58	Total	.15	.17	.13	.19
Bw	1	1.91	.89	1.76	2.06	1	1.37	.67	1.49	1.25
	2	2.27	.76	2.14	2.40	2	1.75	.77	1.62	1.89
	3	2.10	.77	1.97	2.23	3	1.75	.75	1.62	1.88

	4	2.06	.98	1.89	2.23	4	1.64	.81	1.50	1.78
	5	1.55	.86	1.40	1.70	5	1.36	.75	1.23	1.49
	Total	1.97	.55	1.87	2.07	Total	1.57	.47	1.49	1.65
Bc	1	.39	.3	.34	.44	1	.58	.39	.65	.51
	2	.52	.33	.46	.58	2	.74	.46	.66	.82
	3	.57	.37	.51	.63	3	.79	.47	.71	.87
	4	.52	.36	.46	.58	4	.80	.52	.71	.89
	5	.37	.32	.32	.43	5	.67	.52	.58	.76
	Total	.47	.18	.44	.50	Total	.71	.30	.66	.76
Bs	1	.80	.33	.74	.86	1	.84	.47	.92	.76
	2	.76	.30	.71	.81	2	.74	.42	.66	.81
	3	.76	.36	.70	.82	3	.68	.40	.61	.75
	4	.67	.36	.61	.73	4	.72	.41	.65	.79
	5	.65	.40	.58	.72	5	.77	.40	.70	.84
	Total	.73	.21	.69	.77	Total	.74	.23	.70	.78
Bcr	1	.11	.20	.08	.15	1	.18	.31	.23	.13
	2	.20	.30	.15	.25	2	.21	.34	.16	.27
	3	.29	.46	.21	.37	3	.22	.32	.16	.27
	4	.41	.60	.31	.51	4	.18	.26	.13	.22
	5	.70	.72	.58	.83	5	.48	.62	.38	.59
	Total	.34	.31	.29	.39	Total	.25	.21	.21	.29
Duration (median length of pause)										
Total	1	7.89	27.1	3.2	12.58	1	6.00	17.73	9.09	2.92
	2	4.25	1.42	4.00	4.50	2	4.17	3.22	4.73	3.61
	3	4.27	1.27	4.05	4.49	3	3.78	.89	3.93	3.62
	4	4.32	1.50	4.06	4.58	4	3.88	1.16	4.08	3.67

	5	5.81	4.72	4.99	6.63	5	4.55	2.24	4.94	4.16
	Total	4.19	.89	4.04	4.34	Total	3.81	.63	3.70	3.92
Ww	1	3.87	3.33	3.12	4.62	1	3.16	1.56	3.72	2.60
	2	3.79	3.11	3.12	4.46	2	2.98	1.10	2.59	3.38
	3	3.60	2.51	3.04	4.16	3	4.29	3.99	2.93	5.65
	4	5.89	16.52	1.87	9.91	4	3.20	1.49	2.56	3.84
	5	5.84	15.93	1.74	9.94	5	2.89	.91	2.54	3.25
	Total	3.20	1.31	2.97	3.43	Total	3.19	1.64	2.83	3.55
Bpc	1	3.78	3.01	3.21	4.34	1	2.88	1.12	3.19	2.56
	2	3.50	1.66	3.19	3.81	2	3.24	3.10	2.48	4.00
	3	3.57	3.25	2.96	4.18	3	3.13	1.27	2.78	3.48
	4	3.25	1.31	3.00	3.50	4	2.96	1.12	2.67	3.26
	5	3.37	2.60	2.83	3.91	5	3.38	3.44	2.32	4.43
	Total	2.96	.59	2.86	3.06	Total	2.94	1.45	2.67	3.21
Bw	1	3.94	1.20	3.73	4.15	1	3.70	1.20	3.91	3.49
	2	4.11	1.77	3.80	4.42	2	3.65	1.17	3.44	3.85
	3	4.04	1.35	3.81	4.27	3	3.64	1.67	3.35	3.93
	4	4.04	1.39	3.80	4.28	4	3.72	1.96	3.38	4.06
	5	4.32	3.87	3.64	5.00	5	3.53	1.44	3.27	3.78
	Total	3.75	.75	3.62	3.88	Total	3.36	.51	3.27	3.45
Bc	1	6.81	5.5	5.77	7.85	1	5.40	4.54	6.24	4.56
	2	6.23	5.17	5.31	7.16	2	5.02	2.77	4.52	5.52
	3	5.43	3.54	4.78	6.08	3	4.81	2.39	4.39	5.24
	4	5.38	4.52	4.55	6.21	4	4.75	2.50	4.30	5.20
	5	5.91	4.77	4.98	6.84	5	5.19	3.72	4.50	5.88
	Total	4.64	1.78	4.32	4.94	Total	4.18	1.41	3.93	4.42

Bs	1	24.47	43.21	16.98	31.96	1	28.23	44.59	35.96	2.51
	2	10.30	7.86	8.93	11.67	2	8.12	6.89	6.89	9.35
	3	11.20	14.98	8.58	13.82	3	6.72	5.21	5.79	7.65
	4	12.14	12.32	9.96	14.32	4	6.63	4.90	5.76	7.50
	5	13.75	13.76	11.32	16.18	5	1.18	8.33	8.73	11.62
	Total	8.97	5.23	8.06	9.88	Total	6.74	2.76	6.26	7.22
Bcr	1	5.64	6.09	3.86	7.42	1	5.01	3.89	6.08	3.94
	2	6.62	4.29	5.55	7.69	2	4.68	4.70	3.49	5.86
	3	6.46	7.10	4.84	8.08	3	5.32	3.17	4.51	6.13
	4	9.36	13.72	6.35	12.37	4	6.57	7.12	4.69	8.45
	5	14.21	1.98	12.07	16.35	5	1.32	1.37	8.15	12.48
	Total	9.17	9.08	7.56	10.80	Total	6.07	5.53	5.05	7.09

Note: Ww = Within a word, Bpc = Between Pinyin and character(s), Bw = Between words, Bc = Between clauses, Bs = Between sentences, Bcr = Between contextual revisions

Table 3 Revision frequency (number of revisions per minute) by dimension and stage in Chinese writing

	L2 Chinese (*N* = 32)					L1 Chinese (*N* = 32)				
	Stage	*M*	*SD*	95% *CI* low	95% *CI* up	Stage	*M*	*SD*	95% *CI* low	95% *CI* up
Total revision	1	3.22	2.45	2.80	3.64	1	5.43	2.61	4.98	5.88
	2	4.15	3.23	3.59	4.71	2	6.43	2.31	6.03	6.83
	3	4.20	2.80	3.72	4.69	3	6.60	2.43	6.18	7.02
	4	4.18	3.03	3.65	4.71	4	6.61	2.46	6.18	7.04
	5	3.57	2.25	3.18	3.96	5	5.90	2.20	5.52	6.28
	Total	3.86	2.55	3.24	4.12	Total	6.18	2.00	5.83	6.53
Linguistic domain										
Below word	1	1.63	1.55	1.36	1.90	1	2.19	1.29	1.97	2.41
	2	2.00	2.02	1.65	2.35	2	2.57	1.24	2.35	2.78
	3	2.01	1.75	1.71	2.31	3	2.70	1.31	2.47	2.93
	4	2.06	1.77	1.75	2.37	4	2.73	1.31	2.50	2.96
	5	1.56	1.35	1.33	1.79	5	2.35	1.29	2.13	2.57
	Total	1.85	1.51	1.59	2.11	Total	2.51	.95	2.34	2.67
Word	1	.86	.60	.76	.96	1	1.40	.86	1.25	1.55
	2	1.10	.81	.96	1.24	2	1.66	.86	1.51	1.81
	3	1.13	.86	.98	1.28	3	1.75	.93	1.41	1.73
	4	1.15	1.02	.97	1.33	4	1.73	.90	1.57	1.89
	5	1.03	.72	.91	1.61	5	1.55	.71	1.43	1.63
	Total	1.06	.65	.95	1.17	Total	1.62	.63	1.51	1.73
Below clause	1	.51	.47	.43	.59	1	1.36	.47	1.28	1.44
	2	.70	.59	.60	.80	2	1.68	.59	1.58	1.78
	3	.72	.62	.61	.83	3	1.65	.62	1.54	1.76

	4	.67	.50	.58	.76	4	1.66	.50	1.57	1.75
	5	.63	.44	.55	.71	5	1.45	.44	1.37	1.53
	Total	.65	.40	.58	.72	Total	1.56	.66	1.45	1.67
Clause and above	1	.10	.15	.01	.13	1	.22	.31	.17	.27
	2	.15	.18	.12	.18	2	.26	.31	.21	.31
	3	.16	.24	.12	.20	3	.28	.32	.23	.34
	4	.16	.21	.12	.20	4	.27	.35	.17	.29
	5	.17	.22	.13	.21	5	.30	.37	.24	.36
	Total	.15	.13	.13	.17	Total	.27	.21	.23	.31
Context										
Pre-contextual	1	2.91	2.41	2.49	3.33	1	4.86	2.49	4.26	5.12
	2	3.66	2.23	3.27	4.05	2	5.77	2.15	5.40	6.14
	3	3.68	2.81	3.19	4.17	3	5.94	2.38	5.53	6.35
	4	3.53	3.04	3.00	4.06	4	6.04	2.43	5.62	6.46
	5	2.71	2.23	2.31	3.11	5	5.11	2.33	4.71	5.51
	Total	3.30	2.51	2.86	3.73	Total	5.55	1.93	5.22	5.88
Contextual	1	.31	.33	.25	.37	1	.48	.58	.38	.58
	2	.48	.47	.40	.56	2	.66	.72	.54	.79
	3	.51	.59	.41	.61	3	.65	.71	.53	.77
	4	.66	.74	.53	.79	4	.57	.54	.48	.66
	5	.86	.78	.73	1.00	5	.79	.77	.66	.92
	Total	.57	.42	.50	.64	Total	.63	.48	.55	.71
Level of transcription										
Pinyin	1	2.05	1.71	1.75	2.35	1	2.38	1.38	2.14	2.62
	2	2.50	2.08	2.14	2.86	2	2.80	1.29	2.58	3.02
	3	2.59	2.01	2.24	2.94	3	2.88	1.47	2.63	3.13

	4	2.43	1.39	2.19	2.67	4	2.94	1.46	2.69	3.19
	5	1.83	1.49	1.57	2.09	5	2.50	1.41	2.26	2.74
	Total	2.28	1.62	2.00	2.56	Total	2.70	1.08	2.51	2.89
Character	1	1.07	.91	.91	1.23	1	2.79	1.66	2.50	3.08
	2	1.45	1.31	1.22	1.68	2	3.37	1.66	3.08	3.66
	3	1.43	1.15	1.23	1.63	3	3.49	1.61	3.21	3.77
	4	1.61	1.89	1.28	1.94	4	3.44	1.67	3.15	3.73
	5	1.55	1.02	1.37	1.73	5	3.16	1.40	2.92	3.40
	Total	1.42	1.04	1.24	1.60	Total	3.25	1.36	3.01	3.49

Table 4 Speed fluency by genre and stage in Chinese writing

Stage	Argumentative								Narrative							
	Production rate				P-burst length				Production rate				P-burst length			
	M	*SD*	95% *CI* low	95% *CI* up	*M*	*SD*	95% *CI* low	95% *CI* up	*M*	*SD*	95% *CI* low	95% *CI* up	*M*	*SD*	95% *CI* low	95% *CI* up
L2 Chinese (N =32)																
1	47.17	13.25	43.93	50.42	4.73	1.93	4.26	5.21	46.41	13.51	4.27	1.78	4.27	1.78	3.82	4.72
2	44.68	15.58	40.86	48.50	4.23	1.84	3.77	4.68	43.14	9.98	4.36	1.99	4.36	1.99	3.87	4.85
3	42.74	11.49	39.93	45.56	4.67	1.97	4.18	5.16	41.75	11.75	4.39	2.41	4.39	2.41	3.79	4.98
4	48.52	24.19	42.59	54.45	4.47	1.66	4.06	4.87	43.48	12.75	4.45	1.67	4.45	1.67	4.04	4.86
5	41.16	10.43	38.60	43.71	4.07	1.84	3.62	4.52	45.02	49.95	3.84	1.61	3.84	1.61	3.44	4.24
Total	43.95	8.57	41.85	46.05	4.47	1.37	4.13	4.81	42.38	8.73	40.24	44.52	4.26	1.42	3.91	4.61
L1 Chinese (N =32)																
1	69.24	16.34	65.24	73.25	5.52	2.29	4.94	6.10	70.09	25.79	63.67	76.51	5.96	3.15	5.12	6.81
2	64.17	14.00	60.74	67.61	5.86	2.52	5.21	6.52	62.65	18.09	58.22	67.08	5.94	3.34	5.07	6.80
3	64.96	18.80	60.36	69.57	6.18	3.17	5.39	6.97	63.29	19.14	58.61	67.98	5.89	2.98	5.13	6.66
4	69.02	19.75	64.18	73.85	5.60	3.07	4.81	6.39	62.69	14.47	59.15	66.24	6.37	3.41	5.49	7.25
5	58.11	13.49	54.80	61.41	5.81	3.18	5.01	6.61	56.49	14.40	52.96	60.01	5.20	2.54	4.55	5.84
Total	64.63	11.28	61.87	67.40	5.77	1.59	5.38	6.16	61.92	10.87	59.26	64.58	5.65	1.69	5.23	6.07

Note: Production rate = Number of characters produced divided by active writing time, P-burst length = Number of characters between two pauses

Table 5 Pause frequency and duration by genre, location and stage in Chinese writing

		Argumentative								Narrative							
		Frequency (Number of pauses per minute)				Duration (median length of pause)				Frequency (Number of pauses per minute)				Duration (median length of pause)			
Pause location	Stage	*M*	*SD*	95% *CI* low	95% *CI* up	*M*	*SD*	95% *CI* low	95% *CI* up	*M*	*SD*	95% *CI* low	95% *CI* up	*M*	*SD*	95% *CI* low	95% *CI* up
L2 Chinese (N =32)																	
Total	1	4.05	1.10	3.78	4.32	4.57	1.46	4.21	4.92	3.64	1.38	3.30	3.98	11.22	38.16	1.87	20.57
	2	4.51	1.23	4.21	4.81	4.49	1.70	4.07	4.91	4.58	1.11	4.30	4.85	4.01	1.04	3.76	4.27
	3	4.29	1.03	4.04	4.55	4.43	1.43	4.08	4.78	4.58	1.25	4.27	4.89	4.10	1.07	3.84	4.36
	4	4.39	1.23	4.08	4.69	4.33	1.07	4.07	4.59	4.21	1.19	3.92	4.50	4.30	1.85	3.85	4.76
	5	3.77	1.01	3.52	4.02	5.53	3.84	4.59	6.47	3.72	1.09	3.45	3.98	6.10	5.48	4.76	7.44
	Total	4.21	.84	4.00	4.42	4.27	.85	4.06	4.48	4.14	.86	3.93	4.35	4.11	.93	3.88	4.34
Ww	1	.18	.17	.14	.22	4.11	4.22	2.81	5.40	.14	.17	.10	.18	3.59	1.84	2.98	4.20
	2	.23	.26	.16	.29	3.93	3.86	2.79	5.07	.19	.21	.14	.24	3.62	1.96	3.00	4.24
	3	.22	.22	.16	.27	3.68	2.76	2.86	4.51	.18	.22	.12	.23	3.51	2.20	2.78	4.23
	4	.16	.17	.12	.20	6.85	2.93	.10	13.59	.12	.17	.08	.16	4.62	7.74	1.75	7.48
	5	.14	.20	.09	.19	4.21	4.21	2.65	5.77	.13	.18	.09	.18	7.36	21.84	-.46	15.17
	Total	.18	.11	.15	.21	3.22	1.31	2.89	3.55	.15	.09	.13	.17	3.17	1.33	2.84	3.50
Bpc	1	.47	.44	.36	.57	3.50	2.80	2.79	4.22	.45	.46	.33	.56	4.10	3.24	3.20	5.00
	2	.53	.56	.40	.67	3.76	1.80	3.27	4.24	.61	.52	.48	.73	3.25	1.49	2.86	3.64
	3	.44	.46	.33	.56	4.19	4.54	2.94	5.43	.59	.53	.46	.72	3.02	1.07	2.74	3.30
	4	.47	.51	.34	.59	3.22	1.50	2.80	3.63	.48	.47	.36	.59	3.27	1.13	2.98	3.57
	5	.33	.42	.23	.44	3.58	3.25	2.62	4.55	.35	.39	.25	.44	3.17	1.78	2.65	3.68

	Total	.48	.53	.35	.61	3.04	.63	2.88	3.20	.51	.46	.40	.62	2.88	.54	2.75	3.01
Bw	1	2.01	.95	1.78	2.24	3.98	1.16	3.70	4.27	1.81	.82	1.61	2.01	3.90	1.26	3.58	4.21
	2	2.22	.68	2.05	2.39	4.26	2.00	3.77	4.74	2.32	.82	2.12	2.52	3.97	1.52	3.60	4.34
	3	2.08	.76	1.89	2.26	4.24	1.54	3.86	4.62	2.12	.80	1.92	2.31	3.84	1.11	3.57	4.11
	4	2.15	1.01	1.91	2.40	4.15	1.40	3.81	4.50	1.97	.94	1.74	2.20	3.92	1.37	3.58	4.26
	5	1.64	.91	1.42	1.86	4.73	5.15	3.46	6.00	1.45	.80	1.25	1.65	3.90	1.71	3.47	4.33
	Total	2.02	.58	1.88	2.16	3.84	.81	3.64	4.04	1.92	.53	1.79	2.05	3.67	.69	3.50	3.84
Bc	1	.43	.30	.35	.50	7.68	5.90	6.18	9.19	.35	.31	.28	.43	5.74	4.80	4.38	7.10
	2	.51	.33	.43	.59	5.71	4.01	4.66	6.76	.54	.32	.46	.62	6.69	5.99	5.23	8.16
	3	.54	.37	.45	.63	5.72	4.32	4.60	6.84	.59	.37	.50	.68	5.15	2.55	4.49	5.81
	4	.47	.33	.39	.55	6.11	5.89	4.56	7.67	.58	.38	.48	.67	4.68	2.53	4.03	5.33
	5	.33	.27	.27	.40	6.47	5.63	4.91	8.03	.41	.36	.32	.50	5.36	3.72	4.34	6.38
	Total	.46	.16	.42	.50	4.78	1.88	4.32	5.24	.49	.20	.44	.54	4.50	1.67	4.09	4.91
Bs	1	.82	.33	.73	.90	16.63	18.14	12.18	21.07	.78	.33	.70	.86	32.31	57.53	18.21	46.40
	2	.76	.30	.68	.83	1.98	9.10	8.75	13.21	.76	.30	.69	.83	9.61	6.36	8.04	11.18
	3	.72	.31	.64	.80	1.19	9.00	7.98	12.39	.80	.40	.70	.89	12.26	19.34	7.44	17.07
	4	.68	.35	.60	.77	12.67	13.12	9.41	15.94	.67	.38	.57	.76	11.60	11.54	8.70	14.49
	5	.67	.40	.57	.77	11.91	8.06	9.91	13.92	.63	.41	.53	.73	15.61	17.66	11.18	2.04
	Total	.73	.21	.68	.78	9.39	5.46	8.05	1.73	.73	.22	.68	.78	8.56	5.01	7.33	9.79
Bcr	1	.12	.20	.07	.17	5.07	2.62	3.97	6.16	.10	.19	.06	.15	6.19	8.19	2.84	9.54
	2	.23	.33	.15	.31	6.93	4.43	5.42	8.45	.17	.27	.10	.23	6.25	4.16	4.74	7.77
	3	.29	.45	.18	.40	7.47	9.10	4.49	1.44	.29	.46	.18	.40	5.51	4.39	4.11	6.91
	4	.44	.62	.28	.59	11.76	17.92	6.41	17.12	.37	.58	.23	.52	6.57	4.88	5.00	8.14
	5	.64	.65	.48	.80	15.88	13.58	12.08	19.68	.77	.79	.57	.96	12.64	7.60	1.57	14.70
	Total	.34	.30	.27	.41	9.70	11.06	6.90	12.50	.34	.32	.26	.42	8.66	6.69	6.99	1.33

L1 Chinese (N =32)

Total	1	3.44	1.01	3.19	3.69	4.21	1.55	3.83	4.59	2.86	1.13	2.58	3.14	7.82	25.09	1.63	14.02
	2	3.75	.89	3.53	3.97	3.97	1.07	3.71	4.24	3.63	1.19	3.34	3.92	4.37	4.44	3.28	5.46
	3	3.71	.99	3.46	3.95	3.89	.81	3.69	4.09	3.60	.90	3.38	3.83	3.67	.96	3.43	3.90
	4	3.67	1.10	3.40	3.94	3.91	.97	3.67	4.15	3.40	1.11	3.13	3.67	3.84	1.33	3.52	4.17
	5	3.51	1.07	3.25	3.77	4.39	1.86	3.94	4.85	3.39	.89	3.17	3.60	4.71	2.57	4.08	5.34
	Total	3.61	.72	3.44	3.79	3.83	.62	3.68	3.99	3.39	.67	3.22	3.55	3.79	.64	3.63	3.94
Ww	1	.07	.13	.04	.10	3.12	1.60	2.36	3.88	.06	.12	.03	.08	3.21	1.57	2.35	4.06
	2	.08	.14	.04	.11	3.19	1.28	2.60	3.79	.06	.14	.02	.09	2.67	.69	2.28	3.06
	3	.07	.12	.04	.10	4.09	2.91	2.75	5.44	.07	.15	.04	.11	4.52	5.10	1.94	7.10
	4	.05	.13	.02	.08	3.56	1.79	2.44	4.67	.05	.11	.02	.07	2.88	1.15	2.20	3.56
	5	.06	.13	.02	.09	2.88	1.00	2.32	3.45	.06	.14	.02	.09	2.90	.86	2.43	3.37
	Total	.07	.07	.05	.08	3.17	1.19	2.82	3.53	.06	.08	.04	.08	3.20	2.06	2.55	3.86
Bpc	1	.13	.30	.06	.20	2.67	.81	2.35	3.00	.11	.17	.07	.16	3.08	1.36	2.54	3.62
	2	.17	.25	.11	.23	2.86	.89	2.54	3.19	.23	.28	.16	.30	3.56	4.11	2.20	4.92
	3	.13	.20	.08	.18	3.54	1.59	2.90	4.18	.17	.26	.10	.23	2.75	.73	2.47	3.03
	4	.12	.20	.07	.17	2.61	.61	2.35	2.88	.22	.26	.16	.28	3.19	1.30	2.74	3.63
	5	.10	.18	.05	.14	3.05	1.76	2.31	3.78	.12	.23	.06	.18	3.76	4.73	1.63	5.89
	Total	.13	.17	.09	.17	2.85	.94	2.59	3.11	.17	.16	.13	.21	3.03	1.79	2.57	3.50
Bw	1	1.52	.69	1.35	1.69	3.73	1.17	3.45	4.02	1.22	.62	1.07	1.37	3.67	1.24	3.36	3.99
	2	1.79	.71	1.62	1.97	3.63	1.13	3.36	3.91	1.71	.82	1.51	1.91	3.66	1.22	3.36	3.96
	3	1.85	.85	1.64	2.06	3.89	2.18	3.35	4.43	1.65	.63	1.49	1.80	3.39	.88	3.18	3.61
	4	1.82	.84	1.61	2.02	3.92	2.61	3.28	4.56	1.45	.73	1.27	1.63	3.53	.95	3.30	3.76
	5	1.44	.78	1.25	1.63	3.48	1.27	3.17	3.80	1.29	.70	1.11	1.46	3.57	1.61	3.17	3.97
	Total	1.68	.53	1.55	1.81	3.39	.55	3.25	3.52	1.47	.38	1.37	1.56	3.32	.47	3.21	3.44
Bc	1	.66	.39	.57	.76	5.10	3.48	4.21	5.99	.49	.38	.40	.59	5.74	5.49	4.26	7.22
	2	.78	.46	.67	.89	5.78	3.43	4.92	6.65	.70	.46	.59	.81	4.22	1.48	3.83	4.60

	3	.73	.45	.62	.84	4.93	2.50	4.30	5.57	.85	.49	.73	.97	4.70	2.28	4.12	5.27
	4	.76	.51	.64	.89	4.84	2.56	4.18	5.49	.84	.53	.71	.97	4.66	2.45	4.03	5.28
	5	.71	.55	.58	.85	5.21	2.74	4.48	5.94	.62	.48	.50	.74	5.17	4.47	4.02	6.32
	Total	.73	.30	.65	.80	4.32	1.61	3.92	4.71	.70	.30	.63	.77	4.04	1.18	3.75	4.33
Bs	1	.88	.47	.77	1.00	15.89	13.45	12.60	19.19	.79	.46	.68	.90	4.58	59.31	26.05	55.11
	2	.75	.40	.66	.85	8.34	7.22	6.53	1.15	.72	.44	.61	.83	7.89	6.58	6.23	9.56
	3	.68	.37	.59	.77	8.02	6.57	6.38	9.67	.68	.44	.57	.79	5.39	2.78	4.69	6.10
	4	.75	.40	.66	.85	6.46	4.10	5.43	7.49	.68	.41	.58	.79	6.80	5.62	5.39	8.21
	5	.75	.38	.66	.85	9.75	8.07	7.77	11.72	.78	.42	.68	.88	1.61	8.63	8.50	12.72
	Total	.76	.21	.71	.81	7.18	3.06	6.43	7.93	.73	.26	.67	.79	6.31	2.38	5.73	6.90
Bcr	1	.17	.26	.11	.24	5.48	4.90	3.66	7.29	.19	.35	.10	.27	4.44	2.09	3.59	5.30
	2	.20	.34	.12	.28	4.68	6.64	2.17	7.18	.23	.34	.15	.31	4.67	2.20	3.92	5.42
	3	.25	.34	.17	.33	5.32	3.41	4.13	6.50	.19	.30	.11	.26	5.33	2.93	4.22	6.43
	4	.19	.25	.13	.25	8.62	8.95	5.36	11.88	.17	.28	.10	.24	4.29	3.07	3.10	5.47
	5	.44	.55	.31	.58	8.73	8.80	6.00	11.45	.52	.68	.35	.69	11.64	11.44	8.41	14.88
	Total	.25	.20	.20	.30	5.21	2.55	4.55	5.88	.26	.22	.20	.31	6.93	7.33	5.01	8.85

Note: Ww = Within a word, Bpc = Between Pinyin and character(s), Bw = Between words, Bc = Between clauses, Bs = Between sentences, Bcr = Between contextual revisions

Table 6 Revision frequency (number of revisions per minute) by genre, dimension and stage in Chinese writing

		Argumentative				Narrative			
	Stage	*M*	*SD*	95% *CI* low	95% *CI* up	*M*	*SD*	95% *CI* low	95% *CI* up
L2 Chinese (N = 32)									
Total revision	1	3.49	2.70	2.83	4.15	2.96	2.16	2.43	3.48
	2	3.98	3.20	3.20	4.77	4.31	3.28	3.51	5.11
	3	4.11	3.13	3.34	4.87	4.29	2.44	3.69	4.88
	4	4.04	2.71	3.38	4.70	4.33	3.34	3.51	5.14
	5	3.64	2.30	3.08	4.20	3.50	2.21	2.96	4.05
	Total	3.85	2.63	3.21	4.49	3.86	2.47	3.25	4.46
Linguistic domain									
Below word	1	1.80	1.68	1.39	2.21	1.47	1.40	1.12	1.81
	2	1.94	1.96	1.46	2.42	2.07	2.10	1.56	2.58
	3	1.94	1.91	1.47	2.41	2.07	1.57	1.69	2.46
	4	1.94	1.57	1.56	2.33	2.17	1.96	1.69	2.65
	5	1.60	1.30	1.28	1.92	1.52	1.40	1.17	1.86
	Total	1.84	1.54	1.46	2.22	1.86	1.48	1.50	2.22
Word	1	.90	.60	.76	1.05	.82	.60	.68	.97
	2	1.05	.83	.85	1.25	1.14	.80	.95	1.34
	3	1.16	.77	.97	1.35	1.12	.77	.93	1.31
	4	1.14	1.01	.89	1.39	1.16	1.04	.90	1.41
	5	1.04	.79	.85	1.24	1.01	.64	.85	1.17
	Total	1.06	.68	.89	1.23	1.04	.62	.89	1.19
Below clause	1	.56	.55	.42	.69	.47	.38	.38	.56
	2	.67	.63	.52	.82	.73	.55	.60	.86

	3	.66	.64	.50	.82	.77	.60	.63	.92
	4	.66	.46	.55	.77	.68	.54	.55	.81
	5	.63	.44	.52	.74	.64	.44	.53	.75
	Total	.64	.43	.54	.75	.65	.38	.56	.74
Clause and above	1	.10	.13	.06	.13	.11	.16	.07	.14
	2	.16	.19	.12	.21	.13	.18	.09	.18
	3	.19	.27	.12	.25	.14	.19	.09	.19
	4	.18	.24	.12	.24	.14	.18	.10	.19
	5	.17	.24	.11	.23	.17	.20	.12	.22
	Total	.16	.14	.13	.19	.14	.11	.11	.17
Context									
Pre-contextual	1	3.16	2.71	2.50	3.83	2.65	2.05	2.15	3.16
	2	3.48	3.21	2.70	4.27	3.84	3.27	3.04	4.65
	3	3.61	3.13	2.85	4.38	3.74	2.44	3.14	4.33
	4	3.37	2.69	2.71	4.03	3.68	3.37	2.85	4.51
	5	2.82	2.22	2.28	3.37	2.60	2.26	2.05	3.16
	Total	3.29	2.60	2.65	3.93	3.29	2.41	2.70	3.88
Contextual	1	.33	.34	.24	.41	.30	.33	.22	.38
	2	.50	.46	.39	.61	.47	.48	.35	.58
	3	.49	.60	.35	.64	.55	.61	.40	.70
	4	.67	.75	.48	.85	.65	.73	.47	.83
	5	.82	.74	.63	1.00	.90	.82	.70	1.10
	Total	.56	.40	.46	.66	.57	.44	.46	.68
Level of transcription									
Pinyin	1	2.25	1.82	1.80	2.69	1.84	1.57	1.46	2.23
	2	2.37	1.92	1.90	2.84	2.62	2.24	2.08	3.17

	3	2.51	2.19	1.97	3.04	2.67	1.80	2.23	3.11
	4	2.33	1.66	1.92	2.73	2.52	2.11	2.01	3.04
	5	1.88	1.40	1.53	2.22	1.79	1.57	1.41	2.18
	Total	2.27	1.63	1.87	2.67	2.28	1.61	1.89	2.67
Character	1	1.11	1.03	.85	1.36	1.03	.78	.84	1.22
	2	1.45	1.46	1.09	1.80	1.45	1.14	1.17	1.73
	3	1.44	1.18	1.16	1.73	1.44	1.14	1.16	1.72
	4	1.60	1.33	1.27	1.93	1.62	1.45	1.27	1.98
	5	1.57	1.12	1.29	1.84	1.54	.93	1.31	1.77
	Total	1.43	1.12	1.16	1.70	1.41	.95	1.18	1.64
L1 Chinese (N = 32)									
Total revision	1	5.78	2.45	5.18	6.38	4.91	2.71	4.24	5.57
	2	6.56	2.19	6.03	7.10	6.29	2.44	5.70	6.89
	3	6.40	2.52	5.78	7.02	6.79	2.33	6.22	7.36
	4	6.68	2.65	6.03	7.32	6.55	2.27	5.99	7.10
	5	6.10	2.27	5.54	6.66	5.70	2.12	5.18	6.22
	Total	6.30	2.06	5.80	6.81	6.05	1.96	5.57	6.53
Linguistic domain									
Below word	1	2.45	1.33	2.12	2.77	1.92	1.20	1.63	2.22
	2	2.57	1.19	2.28	2.86	2.57	1.29	2.25	2.88
	3	2.57	1.45	2.22	2.93	2.83	1.16	2.54	3.11
	4	2.72	1.40	2.37	3.06	2.74	1.22	2.44	3.04
	5	2.55	1.32	2.22	2.87	2.16	1.25	1.86	2.47
	Total	2.57	1.02	2.32	2.82	2.44	.89	2.23	2.66
Word	1	1.51	.88	1.30	1.72	1.29	.83	1.09	1.50
	2	1.74	.90	1.52	1.96	1.57	.82	1.37	1.78

	3	1.77	.99	1.53	2.02	1.72	.86	1.50	1.93
	4	1.88	.96	1.64	2.12	1.58	.81	1.38	1.77
	5	1.61	.75	1.42	1.79	1.50	.68	1.33	1.66
	Total	1.70	.68	1.53	1.87	1.53	.56	1.39	1.67
Below clause	1	1.38	.76	1.19	1.56	1.35	1.09	1.08	1.62
	2	1.67	.89	1.45	1.88	1.69	.90	1.47	1.92
	3	1.59	.79	1.39	1.78	1.71	.92	1.48	1.93
	4	1.62	.97	1.39	1.86	1.69	.81	1.49	1.89
	5	1.38	.76	1.20	1.57	1.51	.87	1.30	1.72
	Total	1.53	.59	1.38	1.67	1.59	.72	1.41	1.77
Clause and above	1	.25	.32	.18	.33	.19	.29	.12	.26
	2	.28	.31	.20	.35	.23	.31	.16	.31
	3	.29	.34	.20	.37	.28	.30	.20	.35
	4	.24	.27	.17	.30	.30	.41	.20	.40
	5	.32	.45	.21	.43	.28	.26	.22	.34
	Total	.28	.22	.22	.33	.26	.20	.21	.31
Context									
Pre-contextual	1	5.32	2.37	4.74	5.90	4.41	2.54	3.79	5.03
	2	5.87	2.04	5.37	6.37	5.66	2.27	5.11	6.22
	3	5.71	2.45	5.11	6.31	6.18	2.30	5.62	6.75
	4	6.05	2.64	5.41	6.70	6.02	2.22	5.48	6.57
	5	5.32	2.37	4.75	5.90	4.90	2.30	4.33	5.46
	Total	5.66	2.00	5.17	6.14	5.44	1.88	4.98	5.90
Contextual	1	.46	.46	.35	.58	.49	.69	.33	.66
	2	.69	.71	.51	.86	.63	.73	.45	.81
	3	.69	.77	.51	.88	.61	.64	.45	.77

	4	.62	.55	.49	.76	.52	.52	.39	.65
	5	.77	.75	.59	.96	.80	.80	.61	1.00
	Total	.65	.50	.53	.77	.61	.47	.50	.73
Level of transcription									
Pinyin	1	2.61	1.37	2.27	2.94	2.16	1.35	1.83	2.49
	2	2.76	1.18	2.47	3.05	2.83	1.41	2.49	3.18
	3	2.77	1.59	2.38	3.16	3.00	1.35	2.66	3.33
	4	2.90	1.49	2.54	3.27	2.97	1.45	2.62	3.33
	5	2.62	1.43	2.27	2.97	2.38	1.39	2.04	2.72
	Total	2.73	1.10	2.46	3.00	2.67	1.06	2.41	2.93
Character	1	2.98	1.57	2.60	3.37	2.59	1.74	2.17	3.02
	2	3.50	1.65	3.09	3.90	3.24	1.67	2.83	3.65
	3	3.45	1.68	3.04	3.86	3.53	1.54	3.15	3.91
	4	3.55	1.82	3.11	4.00	3.34	1.51	2.97	3.71
	5	3.24	1.45	2.89	3.59	3.07	1.35	2.74	3.40
	Total	3.34	1.41	3.00	3.69	3.15	1.30	2.83	3.47

Appendix H: Non-significant results from linear mixed-effects regressions and post hoc Bonferroni tests for Research Questions 1-5

Table 1 Non-significant results from post-hoc Bonferroni tests examining the effects of stage on writing behaviours in L2 Chinese (N = 32, $p \geq .01$)

	Stage	*SE*	*t*	*p*
Speed fluency				
Production rate	S1 - S2	.03	2.40	.17
	S1 - S4	.03	1.56	1.00
	S2 - S3	.03	1.21	1.00
	S2 - S4	.03	-.85	1.00
	S2 - S5	.03	2.64	.08
	S3 - S4	.03	-2.05	.41
	S3 - S5	.03	1.44	1.00
P-burst length	S1 - S2	.04	1.23	1.00
	S1 - S3	.04	.32	1.00
	S1 - S4	.04	-.10	1.00
	S2 - S3	.04	-.91	1.00
	S2 - S4	.04	-1.34	1.00
	S2 - S5	.04	2.21	.27
	S3 - S4	.04	-.43	1.00
	S3 - S5	.04	3.11	.02
Pausing				
Total pause frequency	S1 - S5	.11	.91	1.00
	S2 - S3	.11	.81	1.00
	S2 - S4	.11	2.03	.43
	S3 - S4	.11	1.22	1.00
Bpc pause frequency	S1 - S2	.03	-2.43	.15
	S1 - S3	.03	-1.22	1.00
	S1 - S4	.03	-.09	1.00
	S2 - S3	.03	1.21	1.00
	S2 - S4	.03	2.34	.20
	S3 - S4	.03	1.13	1.00
Bw pause frequency	S1 - S3	.09	-2.08	.38
	S1 - S4	.09	-1.66	.97
	S2 - S3	.09	1.88	.61
	S2 - S4	.09	2.29	.22
	S3 - S4	.09	0.42	1.00
Bc pause frequency	S1 - S5	.04	.59	1.00
	S2 - S3	.04	-1.10	1.00
	S2 - S4	.04	.04	1.00

	S3 - S4	.04	1.14	1.00
Bs pause frequency	S1 - S2	.04	.96	1.00
	S1 - S3	.04	.98	1.00
	S1 - S4	.04	3.20	.01
	S2 - S3	.04	.02	1.00
	S2 - S4	.04	2.24	.25
	S2 - S5	.04	2.93	.04
	S3 - S4	.04	2.22	.27
	S3 - S5	.04	2.91	.04
	S4 - S5	.04	.68	1.00
Bcr pause frequency	S1 - S2	.03	-1.98	.48
	S2 - S3	.03	-1.89	.60
	S3 - S4	.03	-2.26	.24
Total pause duration	S1 - S2	.01	-2.50	.13
	S1 - S3	.01	-1.90	.58
	S1 - S4	.01	-1.71	.87
	S1 - S5	.01	2.67	.08
	S2 - S3	.01	.60	1.00
	S2 - S4	.01	.79	1.00
	S3 - S4	.01	.19	1.00
Bc pause duration	S1 - S2	.07	.87	1.00
	S1 - S3	.07	1.87	.62
	S1 - S4	.07	2.61	.09
	S1 - S5	.07	1.97	.49
	S2 - S3	.07	1.04	1.00
	S2 - S4	.07	1.81	.71
	S2 - S5	.07	1.18	1.00
	S3 - S4	.07	.76	1.00
	S3 - S5	.07	.17	1.00
	S4 - S5	.07	-.57	1.00
Bs pause duration	S1 - S5	.08	2.75	.06
	S2 - S3	.08	.20	1.00
	S2 - S4	.08	-.81	1.00
	S2 - S5	.08	-2.83	.05
	S3 - S4	.09	-1.01	1.00
	S3 - S5	.09	-3.02	.03
	S4 - S5	.09	-2.00	.45
Bcr pause duration	S1 - S2	.13	-1.57	1.00
	S1 - S3	.13	-.82	1.00
	S1 - S4	.13	-2.89	.04
	S2 - S3	.12	.89	1.00
	S2 - S4	.12	-1.36	1.00
	S3 - S4	.11	-2.38	.18
Revision				

Total	S2 - S3	.03	-.61	1.00
	S2 - S4	.03	-.50	1.00
	S2 - S5	.03	3.02	.03
	S3 - S4	.03	.11	1.00
Below word	S1 - S5	.03	.19	1.00
	S2 - S3	.03	-.51	1.00
	S2 - S4	.03	-.98	1.00
	S3 - S4	.03	-.47	1.00
Word	S1 - S5	.03	-2.48	.13
	S2 - S3	.03	-.84	1.00
	S2 - S4	.03	-.20	1.00
	S2 - S5	.03	.88	1.00
	S3 - S4	.03	.64	1.00
	S3 - S5	.03	1.72	.86
	S4 - S5	.03	1.08	1.00
Below clause	S1 - S5	.03	-3.11	.02
	S2 - S3	.03	-.24	1.00
	S2 - S4	.03	.29	1.00
	S2 - S5	.03	1.00	1.00
	S3 - S4	.03	.54	1.00
	S3 - S5	.03	1.24	1.00
	S4 - S5	.03	.71	1.00
Clause and above	S1 - S2	.02	-2.26	.24
	S1 - S3	.02	-2.68	.08
	S1 - S4	.02	-2.86	.04
	S1 - S5	.02	-3.20	.01
	S2 - S3	.02	-.43	1.00
	S2 - S4	.02	-.60	1.00
	S2 - S5	.02	-.94	1.00
	S3 - S4	.02	-.18	1.00
	S3 - S5	.02	-.52	1.00
	S4 - S5	.02	-.34	1.00
Pre-contextual	S1 - S5	.04	1.33	1.00
	S2 - S3	.04	-.38	1.00
	S2 - S4	.04	.80	1.00
	S3 - S4	.04	1.18	1.00
Contextual	S2 - S3	.03	-.19	1.00
	S2 - S4	.03	-2.46	.14
	S3 - S4	.03	-2.27	.23
Pinyin	S1 - S5	.04	1.60	1.00
	S2 - S3	.04	-.69	1.00
	S2 - S4	.04	.39	1.00
	S3 - S4	.04	1.09	1.00
Character	S2 - S3	.03	-.15	1.00

	S2 - S4	.03	-2.11	.35
	S2 - S5	.03	-2.28	.22
	S3 - S4	.03	-1.96	.50
	S3 - S5	.03	-2.14	.33
	S4 - S5	.03	-.17	1.00

Note: Ww = Within a word, Bpc = Between Pinyin and character(s), Bw = Between words, Bc = Between clauses, Bs = Between sentences, Bcr = Between contextual revisions

Table 2 Non-significant results from post-hoc Bonferroni tests examining the effects of stage on writing behaviours in L1 Chinese (N = 32, $p \geq .01$)

	Stage	*SE*	*t*	*p*
Speed fluency				
Production rate	S1 - S4	.02	2.17	.30
	S2 - S3	.02	-.07	1.00
	S2 - S4	.02	-1.52	1.00
	S3 - S4	.02	-1.46	.41
Pause				
Total pause frequency	S1 - S5	.11	-2.72	.07
	S2 - S3	.11	.30	1.00
	S2 - S4	.11	1.43	1.00
	S2 - S5	.11	2.23	.26
	S3 - S4	.11	1.13	1.00
	S3 - S5	.11	1.94	.53
	S4 - S5	.11	.80	1.00
Bpc pause frequency	S1 - S3	.02	-1.46	1.00
	S1 - S4	.02	-2.38	.18
	S1 - S5	.02	.41	1.00
	S2 - S3	.02	2.19	.29
	S2 - S4	.02	1.27	1.00
	S3 - S4	.02	-.92	1.00
	S3 - S5	.02	1.87	.62
	S4 - S5	.02	2.79	.05
Bw pause frequency	S1 - S4	.08	-3.18	.02
	S1 - S5	.08	.07	1.00
	S2 - S3	.08	.01	1.00
	S2 - S4	.08	1.40	1.00
	S3 - S4	.08	1.39	1.00
	S4 - S5	.08	3.25	.01
Bc pause frequency	S1 - S2	.05	-3.06	02
	S1 - S5	.05	-1.71	.87
	S2 - S3	.05	-1.05	1.00
	S2 - S4	.05	-1.26	1.00
	S2 - S5	.05	1.35	1.00
	S3 - S4	.05	-.21	1.00
	S3 - S5	.05	2.40	.17
	S4 - S5	.05	2.61	.09
Bs pause frequency	S1 - S2	.05	2.05	.41
	S1 - S4	.05	2.41	.16
	S1 - S5	.05	1.51	1.00
	S2 - S3	.05	1.33	1.00

	S2 - S4	.05	.36	1.00
	S2 - S5	.05	-.54	1.00
	S3 - S4	.05	-.97	1.00
	S3 - S5	.05	-1.87	1.00
	S4 - S5	.05	-.90	1.00
Bcr pause frequency	S1 - S2	.03	-.91	1.00
	S1 - S3	.03	-1.10	1.00
	S1 - S4	.03	-.20	1.00
	S2 - S3	.03	-.19	1.00
	S2 - S4	.03	.71	1.00
	S3 - S4	.03	-.90	1.00
Total pause duration	S1 - S2	.02	-2.01	.44
	S1 - S3	.02	-2.89	.04
	S1 - S4	.02	-2.49	.13
	S1 - S5	.02	.72	1.00
	S2 - S3	.02	-.88	1.00
	S2 - S4	.02	-.47	1.00
	S2 - S5	.02	2.74	.06
	S3 - S4	.02	.41	1.00
	S4 - S5	.02	3.21	.01
Bs pause duration	S2 - S3	.08	1.90	.58
	S2 - S4	.08	2.07	.38
	S2 - S5	.08	-2.68	.07
	S3 - S4	.08	.17	1.00
Bcr pause duration	S1 - S2	.12	.53	1.00
	S1 - S3	.12	-.92	1.00
	S1 - S4	.12	-1.32	1.00
	S2 - S3	.11	-1.53	1.00
	S2 - S4	.11	-1.95	.53
	S3 - S4	.11	-.44	1.00
Revision				
Total	S1 - S5	.21	-2.65	.08
	S2 - S3	.21	-.81	1.00
	S2 - S4	.21	-.88	1.00
	S2 - S5	.21	2.57	.11
	S3 - S4	.21	-.08	1.00
Below word	S1 - S5	.04	-1.79	.74
	S2 - S3	.04	-1.02	1.00
	S2 - S4	.04	-1.19	1.00
	S2 - S5	.04	1.96	.50
	S3 - S4	.04	-.16	1.00
	S3 - S5	.04	2.99	.03
	S4 - S5	.04	3.15	.01
Word	S1 - S2	.03	-3.28	.01

	S1 - S5	.03	-2.49	.13
	S2 - S3	.03	-.94	1.00
	S2 - S4	.03	-.85	1.00
	S2 - S5	.03	.80	1.00
	S3 - S4	.03	.10	1.00
	S3 - S5	.03	1.74	.83
	S4 - S5	.03	1.64	1.00
Below clause	S1 - S5	.03	-1.58	1.00
	S2 - S3	.03	.32	1.00
	S2 - S4	.03	.33	1.00
	S2 - S5	.03	2.68	.08
	S3 - S4	.03	.02	1.00
	S3 - S5	.03	2.37	.18
	S4 - S5	.03	2.35	.19
Clause and above	S1 - S2	.02	-1.08	1.00
	S1 - S3	.02	-1.75	.81
	S1 - S4	.02	-1.22	1.00
	S1 - S5	.02	-2.17	.30
	S2 - S3	.02	-.67	1.00
	S2 - S4	.02	-.15	1.00
	S2 - S5	.02	-1.09	1.00
	S3 - S4	.02	.52	1.00
	S3 - S5	.02	-.42	1.00
	S4 - S5	.02	-.95	1.00
Pre-contextual	S1 - S5	.21	-1.16	1.00
	S2 - S3	.21	-.83	1.00
	S2 - S4	.21	-1.28	1.00
	S2 - S5	.21	3.16	.02
	S3 - S4	.21	-.45	1.00
Contextual	S1 - S2	.03	-2.78	.06
	S1 - S3	.03	-2.76	.06
	S1 - S4	.03	-2.02	.44
	S2 - S3	.03	.01	1.00
	S2 - S4	.03	.75	1.00
	S2 - S5	.03	-2.24	.26
	S3 - S4	.03	.75	1.00
	S3 - S5	.03	-2.24	.25
	S4 - S5	.03	-2.99	.03
Pinyin	S1 - S5	.04	-1.09	1.00
	S2 - S3	.04	-.34	1.00
	S2 - S4	.04	-.79	1.00
	S2 - S5	.04	2.82	.05
	S3 - S4	.04	-.45	1.00
	S3 - S5	.04	3.16	.02

Character	S1 - S5	.14	-2.72	.07
	S2 - S3	.14	-.93	1.00
	S2 - S4	.14	-.60	1.00
	S2 - S5	.14	1.54	1.00
	S3 - S4	.14	.34	1.00
	S3 - S5	.14	2.47	.14
	S4 - S5	.14	2.13	.33

Note: Ww = Within a word, Bpc = Between Pinyin and character(s), Bw = Between words, Bc = Between clauses, Bs = Between sentences, Bcr = Between contextual revisions

Table 3 Non-significant results from linear mixed-effects regressions examining the effects of genre on writing behaviours in L2 Chinese

Dependent variable	Fixed effect: genre	*Est*	*SE*	*t*	*p*
Pause frequency					
Total pause	Narrative	-.06	.13	-.42	.70
Model = lmer(total_pause_frequency) ~ genre + (1|participant) + (1|prompt), data, REML = F)					
Bpc pause	Narrative	.04	.03	1.22	.30
Model = lmer(log(bpc_pause_frequency) ~ genre + (1+genre|participant) + (1|prompt), data, REML = F)					
Bw pause	Narrative	-.05	.03	-1.65	.10
Model = lmer(log(bw_pause_frequency) ~ genre + (1|participant), data, REML = F)					
Bc pause	Narrative	.03	.03	1.26	.29
Model = lmer(bc_pause_frequency ~ genre + (1|participant) + (1|prompt), data, REML = F)					
Bs pause	Narrative	< .01	.06	.01	.99
Model = lmer(bs_pause_frequency ~ genre + (1|participant) + (1|prompt), data, REML = F)					
Bcr pause	Narrative	< .01	.02	.06	.96
Model = lmer(sqrt(bcr_pause_frequency) ~ genre + (1|participant), data, REML = F)					
Pause duration					
Total pause	Narrative	.04	.02	-1.82	.16
Model = lmer(log(total_pause_duration) ~ genre + (1|participant) + (1|prompt), data, REML = F)					
Ww pause	Narrative	.01	.03	.40	.69
Model = lmer(1/log(ww_pause_duration) ~ genre + (1|participant), data, REML = F)					
Bpc pause	Narrative	.05	.03	1.78	.08
Model = lmer(1/log(bpc_pause_duration) ~ genre + (1|participant) + (1|prompt), data, REML = F)					
Bw pause	Narrative	-.04	.03	-1.55	.13
Model = lmer(log(bw_pause_duration) ~ genre + (1|participant), data, REML = F)					
Bc pause	Narrative	-.05	.05	-1.01	.31
Model = lmer(bc_pause_duration ~ genre + (1|participant), data, REML = F)					
Bs pause	Narrative	-.10	.06	-1.75	.08
Model = lmer(log(bs_pause_duration) ~ genre + (1|participant), data, REML = F)					
Bcr pause	Narrative	-.07	.11	-.61	.55
Model = lmer(log(bcr_pause_duration) ~ genre + (1|participant), data, REML = F)					
Revision					
Total	Narrative	.01	.02	.28	.78
Model = lmer(log(total_revision) ~ genre + (1+genre|participant), data, REML = F)					
Below word	Narrative	< .01	.02	.20	.84
Model = lmer(log(below_word_revision) ~ genre + (1|participant), data, REML = F)					
Word	Narrative	< -.01	.02	-.04	.97
Model = lmer(log(word_revision) ~ genre + (1|participant), data, REML=F)					
Below clause	Narrative	.02	.02	1.06	.36
Model = lmer(log(below_clause_revision) ~ genre + (1|participant) + (1|prompt), data, REML = F)					

Clause and above	Narrative	-.02	.01	-1.36	.18
Model = lmer(log(clause_and_above_revision) ~ genre + (1|participant), data, REML = F)					
Pre-contextual	Narrative	.01	.02	-.41	.68
Model = lmer(log(precontextual_revision) ~ genre + (1|participant), data, REML = F)					
Contextual	Narrative	< .01	.02	-.12	.91
Model = lmer(log(contextual_revision) ~ genre + (1|participant), data, REML = F)					
Pinyin	Narrative	< .01	.02	.01	.92
Model = lmer(log(pinyin_revision) ~ genre + (1|participant), data, REML = F)					
Character	Narrative	< .01	.02	.18	.86
Model = lmer(log(character_revision) ~ genre + (1|participant), data, REML = F)					

Note: Ww = Within a word, Bpc = Between Pinyin and character(s), Bw = Between words, Bc = Between clauses, Bs = Between sentences, Bcr = Between contextual revisions

Table 4 Non-significant results from linear mixed-effects regressions examining the effects of genre on writing behaviours in L1 Chinese

Dependent variable	Fixed effect: genre	*Est*	*SE*	*t*	*p*
Speed fluency					
P-burst length	Narrative	-.03	.04	-.66	.51
Model = lmer(log(pburst_length) ~ genre + (1|participant), data, REML = F)					
Pause frequency					
Total pause	Narrative	-.23	.17	-1.39	.24
Model = lmer(total_pause_frequency) ~ genre + (1+genre|participant) + (1|prompt), data, REML = F)					
Ww pause	Narrative	-.01	.01	-.96	.34
Model = lmer(log(ww_pause_frequency) ~ genre + (1|participant), data, REML = F)					
Bpc pause	Narrative	.03	.02	1.58	.20
Model = lmer(log(bpc_pause_frequency) ~ genre + (1|participant) + (1|prompt), data, REML = F)					
Bc pause	Narrative	-.03	.04	-.72	.47
Model = lmer(bc_pause_frequency ~ genre + (1|participant), data, REML = F)					
Bs pause	Narrative	-.03	.03	-1.11	.27
Model = lmer(bs_pause_frequency ~ genre + (1|participant), data, REML = F)					
Bcr pause	Narrative	.01	.03	.27	.79
Model = lmer(bcr_pause_frequency ~ genre + (1+genre|participant), data, REML = F)					
Pause duration					
Total pause	Narrative	-.01	.02	-.65	.52
Model = lmer(log(total_pause_duration) ~ genre + (1+genre|participant), data, REML = F)					
Ww pause	Narrative	.07	.05	1.34	.18
Model = lmer(1/log(ww_pause_duration) ~ genre + (1|participant), data, REML = F)					
Bpc pause	Narrative	< .01	.05	-.09	.93
Model = lmer(1/log(bpc_pause_duration) ~ genre + (1+genre|participant), data, REML = F)					
Bw pause	Narrative	-.01	.02	-.71	.48
Model = lmer(log(bw_pause_duration) ~ genre + (1|participant) + (1|prompt), data, REML = F)					
Bc pause	Narrative	-.05	.04	-1.26	.22
Model = lmer(log(bc_pause_duration) ~ genre + (1+genre|participant), data, REML = F)					
Bs pause	Narrative	-.11	.08	-1.46	.21
Model = lmer(log(bs_pause_duration) ~ genre + (1+genre|participant) + (1|prompt), data, REML = F)					
Bcr pause	Narrative	.13	.08	1.54	.20
Model = lmer(log(bcr_pause_duration) ~ genre + (1|participant) + (1|prompt), data, REML = F)					
Revision					
Total	Genre (narrative)	-.03	.04	-.88	.44
Model = lmer(log(total_revision) ~ genre + (1|participant) + (1|prompt), data, REML = F)					
Below word	Narrative	-.03	.04	-.73	.50
Model = lmer(log(below_word_revision) ~ genre + (1+genre|participant) + (1|prompt), data, REML = F)					

Word	Narrative	-.06	.03	-1.88	.15
Model = lmer(log(word_revision) ~ genre + (1|participant) + (1|prompt), data, REML = F)					
Below clause	Narrative	.01	.03	.46	.68
Model = lmer(log(below_clause_revision) ~ genre + (1|participant) + (1|prompt), data, REML = F)					
Clause and above	Narrative	-.01	.02	-.68	.50
Model = lmer(log(clause_and_above_revision) ~ genre + (1|participant) + (1|prompt), data, REML = F)					
Pre-contextual	Narrative	-.03	.04	-.82	.47
Model = lmer(log(precontextual_revision) ~ genre + (1|participant), data, REML = F)					
Contextual	Narrative	-.02	.03	-.67	.51
Model = lmer(log(contextual_revision) ~ genre + (1+genre|participant), data, REML = F)					
Pinyin	Narrative	-.02	.04	-.45	.67
Model = lmer(log(pinyin_revision) ~ genre + (1|participant) + (1|prompt), data, REML = F)					
Character	Narrative	-.04	.04	-1.03	.37
Model = lmer(log(character_revision) ~ genre + (1|participant) + (1|prompt), data, REML = F)					

Note: Ww = Within a word, Bpc = Between Pinyin and character(s), Bw = Between words, Bc = Between clauses, Bs = Between sentences, Bcr = Between contextual revisions

Table 5 Non-significant results from linear mixed-effects regressions examining the effects of interaction between genre and stage on writing behaviours in L2 Chinese (N = 32)

Dependent variable	Predictor: genre*stage	*Est*	*SE*	*t*	*p*
Speed fluency					
Production rate	Narrative * 2	.02	.06	.30	.76
	Narrative * 3	< -.01	.06	-.08	.94
	Narrative * 4	-.05	.06	-.89	.37
	Narrative * 5	-.03	.06	-.46	.64
Model = lmer(log(production_rate) ~ genre*stage + (1|participant), data, REML = F)					
P-burst length	Narrative * 2	.13	.08	1.58	.12
	Narrative * 3	.01	.08	.16	.87
	Narrative * 4	.10	.08	1.20	.23
	Narrative * 5	.06	.08	.73	.47
Model = lmer(sqrt(pburst_length) ~ genre*stage + (1+genre|participant), data, REML = F)					
Pause frequency					
Ww pause	Narrative * 2	.01	.05	.23	.82
	Narrative * 3	< -.01	.05	-.05	.96
	Narrative * 4	< -.01	.05	-.49	.97
	Narrative * 5	.04	.05	.90	.37
Model = lmer(log(ww_pause_frequency) ~ genre*stage + (1|participant), data, REML = F)					
Bw pause	Narrative * 2	.31	.18	1.69	.09
	Narrative * 3	.24	.18	1.33	.19
	Narrative * 4	.02	.18	.09	.93
	Narrative * 5	.02	.18	.09	.93
Model = lmer(bw_pause_frequency ~ genre*stage + (1+genre|participant), data, REML = F)					
Bs pause	Narrative * 2	.04	.08	.54	.59
	Narrative * 3	.11	.08	1.51	.13
	Narrative * 4	.03	.08	.33	.74
	Narrative * 5	.01	.08	.08	.94
Model = lmer(bs_pause_frequency ~ genre*stage + (1|participant) + (1|prompt), data, REML = F)					
Bcr pause	Narrative * 2	-.03	.06	-.46	.65
	Narrative * 3	.02	.06	.29	.77
	Narrative * 4	-.02	.06	-.38	.71
	Narrative * 5	.07	.06	1.30	.20
Model = lmer(log(bcr_pause_frequency) ~ genre*stage + (1|participant) + (1|prompt), data, REML = F)					
Pause duration					
Total pause	Narrative * 2	.04	.03	1.20	.23
	Narrative * 3	.02	.03	.66	.51
	Narrative * 4	.04	.03	1.45	.15

	Narrative * 5	.01	.03	.30	.76
Model = lmer(1/log(total_pause_duration) ~ genre*stage + (1|participant) + (1|prompt), data, REML = F)					
Ww pause	Narrative * 2	.03	.16	.20	.84
	Narrative * 3	< -.01	.16	-.03	.98
	Narrative * 4	-.01	.17	-.09	.93
	Narrative * 5	< .01	.17	.02	.99
Model = lmer(log(ww_pause_duration) ~ genre*stage + (1|participant), data, REML = F)					
Bw pause	Narrative * 2	-.01	.04	-.18	.86
	Narrative * 3	.01	.04	.19	.85
	Narrative * 4	.02	.04	.63	.53
	Narrative * 5	.05	.04	1.23	.22
Model = lmer(1/log(bw_pause_duration) ~ genre*stage + (1+genre|participant), data, REML = F)					
Bc pause	Narrative * 2	-.07	.06	-1.15	.25
	Narrative * 3	-.07	.06	-1.14	.26
	Narrative * 4	-.05	.06	-.80	.42
	Narrative * 5	-.03	.06	-.51	.61
Model = lmer(1/log(bc_pause_duration) ~ genre*stage + (1|participant), data, REML = F)					
Bs pause	Narrative * 2	-.29	.17	-1.75	.08
	Narrative * 3	-.27	.17	-1.61	.11
	Narrative * 4	-.31	.17	-1.87	.06
	Narrative * 5	-.09	.17	-.54	.59
Model = lmer(log(bs_pause_duration) ~ genre*stage + (1|participant), data, REML = F)					
Bcr pause	Narrative * 2	-.05	.26	-.21	.83
	Narrative * 3	-.13	.25	-.51	.61
	Narrative * 4	-.27	.25	-1.09	.27
	Narrative * 5	-.01	.24	-.03	.98
Model = lmer(log(bcr_pause_duration) ~ genre*stage + (1|participant) + (1|prompt), data, REML = F)					
Revision					
Word	Narrative * 2	.09	.06	1.57	.12
	Narrative * 3	.03	.06	.54	.59
	Narrative * 4	.06	.06	.94	.35
	Narrative * 5	.05	.06	.86	.39
Model = lmer(log(word_revision) ~ genre*stage + (1+genre|participant), data, REML=F)					
Clause and above	Narrative * 2	-.03	.03	-.92	.36
	Narrative * 3	-.03	.03	-.98	.33
	Narrative * 4	-.03	.03	-.99	.32
	Narrative * 5	< -.01	.03	-.09	.93
Model = lmer(log(clause_and_above_revision) ~ genre*stage + (1|participant), data, REML=F)					
Contextual	Narrative * 2	-.01	.06	-.10	.92
	Narrative * 3	.06	.06	.89	.38
	Narrative * 4	< .01	.06	.05	.96
	Narrative * 5	.05	.06	.78	.44

Model = lmer(log(contextual_revision) ~ genre*stage + (1|participant), data, REML=F)					
Character	Narrative * 2	.05	.06	.86	.39
	Narrative * 3	.02	.06	.42	.67
	Narrative * 4	.02	.06	.42	.67
	Narrative * 5	.02	.06	.40	.69
Model = lmer(log(character_revision) ~ genre*stage + (1+genre|participant), data, REML=F)					

Note: Ww = Within a word, Bw = Between words, Bc = Between clauses, Bs = Between sentences, Bcr = Between contextual revisions

Table 6 Non-significant results from linear mixed-effects regressions examining the effects of interaction between genre and stage on writing behaviours in L1 Chinese (N = 32)

Dependent variable	Predictor: genre*stage	*Est*	*SE*	*t*	*p*
Speed fluency					
Production rate	Narrative * 2	-2.02	2.98	-.68	.50
	Narrative * 3	-2.22	2.98	-.74	.46
	Narrative * 4	-5.53	2.98	-1.86	.06
	Narrative * 5	-1.43	2.98	-.48	.63
Model = lmer(production_rate ~ genre*stage + (1|participant), data, REML = F)					
P-burst length	Narrative * 2	-.04	.12	-.35	.72
	Narrative * 3	-.07	.12	-.59	.55
	Narrative * 4	.11	.12	.90	.37
	Narrative * 5	-.14	.12	-1.17	.24
Model = lmer(log(pburst_length) ~ genre*stage + (1|participant) + (1|prompt), data, REML = F)					
Pause frequency					
Ww pause	Narrative * 2	-.01	.03	-.24	.81
	Narrative * 3	.01	.03	.45	.65
	Narrative * 4	.01	.03	.45	.65
	Narrative * 5	.01	.03	.37	.71
Model = lmer(ww_pause_frequency ~ genre*stage + (1+genre|participant), data, REML = F)					
Bw pause	Narrative * 2	.22	.17	1.32	.18
	Narrative * 3	.10	.17	.60	.55
	Narrative * 4	-.06	.17	-.39	.69
	Narrative * 5	.15	.17	.91	.36
Model = lmer(bw_pause_frequency ~ genre*stage + (1|participant) + (1|prompt), data, REML = F)					
Bs pause	Narrative * 2	.05	.09	.58	.56
	Narrative * 3	.08	.09	.88	.38
	Narrative * 4	.02	.09	.22	.82
	Narrative * 5	.12	.09	1.25	.21
Model = lmer(bs_pause_frequency ~ genre*stage + (1|participant) + (1|prompt), data, REML = F)					
Bcr pause	Narrative * 2	.03	.08	.34	.74
	Narrative * 3	-.07	.08	-.85	.39
	Narrative * 4	-.03	.08	-.41	.68
	Narrative * 5	.05	.08	.60	.55
Model = lmer(bcr_pause_frequency ~ genre*stage + (1+genre|participant), data, REML = F)					
Pause duration					
Ww pause	Narrative * 2	-.49	.95	-.52	.60
	Narrative * 3	-.05	.93	-.05	.95
	Narrative * 4	-.74	1.04	-.72	.48
	Narrative * 5	.19	1.00	.19	.85

Model = lmer(ww_pause_duration) ~ genre*stage + (1|participant), data, REML = F)					
Bw pause	Narrative * 2	.02	.06	.38	.97
	Narrative * 3	-.06	.06	-.98	.33
	Narrative * 4	-.01	.06	-.20	.85
	Narrative * 5	.03	.06	.97	.62
Model = lmer(log(bw_pause_duration) ~ genre*stage + (1|participant), data, REML = F)					
Bcr pause	Narrative * 2	.25	.22	1.11	.26
	Narrative * 3	.06	.22	.25	.81
	Narrative * 4	-.41	.22	-1.79	.07
	Narrative * 5	.37	.21	1.79	.07
Model = lmer(log(bcr_pause_duration) ~ genre*stage + (1|participant) + (1|prompt), data, REML = F)					
Revision					
Word	Narrative * 2	.04	.07	.61	.54
	Narrative * 3	.10	.07	1.46	.15
	Narrative * 4	.10	.07	.01	.99
	Narrative * 5	.07	.07	.98	.33
Model = lmer(log(word_revision) ~ genre*stage + (1|participant) + (1|prompt), data, REML = F)					
Below clause	Narrative * 2	.07	.07	1.00	.32
	Narrative * 3	.09	.07	1.31	.19
	Narrative * 4	.10	.07	1.49	.14
	Narrative * 5	.09	.07	1.38	.17
Model = lmer(log(word_revision) ~ genre*stage + (1+genre|participant) + (1|prompt), data, REML = F)					
Clause and above	Narrative * 2	.01	.05	.17	.86
	Narrative * 3	.04	.05	.90	.37
	Narrative * 4	.08	.05	1.64	.10
	Narrative * 5	.04	.05	.89	.37
Model = lmer(log(clause_and_above_revision) ~ genre*stage + (1|participant), data, REML = F)					
Contextual	Narrative * 2	-.02	.07	-.33	.74
	Narrative * 3	-.03	.07	-.39	.70
	Narrative * 4	-.05	.07	-.73	.46
	Narrative * 5	.02	.07	.36	.72
Model = lmer(log(contextual_revision) ~ genre*stage + (1|participant), data, REML = F)					

Note: Ww = Within a word, Bw = Between words, Bs = Between sentences, Bcr = Between contextual revisions

Table 7 Non-significant results from post-hoc Bonferroni tests examining the effects of stage on writing behaviours by genre in L2 Chinese (N = 32, p ≥ .01)

	Argumentative				*Narrative*			
	Stage	*SE*	*t*	*p*	Stage	*SE*	*t*	*p*
Pausing								
Total pause	S1 - S2	.01	-2.83	.05	S1 - S5	.17	-.45	1.00
frequency	S1 - S3	.01	-1.57	1.00	S2 - S3	.17	-.03	1.00
	S1 - S4	.01	-2.19	.30	S2 - S4	.17	2.17	.31
	S1 - S5	.01	1.86	.63	S3 - S4	.17	2.20	.29
	S2 - S3	.01	1.25	1.00	S4 - S5	.17	2.94	.04
	S2 - S4	.01	.64	1.00				
	S3 - S4	.01	.61	1.00				
Bpc pause	S1 - S2	.03	-.78	1.00	S1 - S2	.04	-2.64	.09
frequency	S1 - S3	.03	.77	1.00	S1 - S3	.04	-2.42	.17
	S1 - S4	.03	.58	1.00	S1 - S4	.04	-.67	1.00
	S1 - S5	.03	3.00	.03	S1 - S5	.04	1.79	.74
	S2 - S3	.03	1.55	1.00	S2 - S3	.04	.22	1.00
	S2 - S4	.03	1.36	1.00	S2 - S4	.04	1.97	.49
	S3 - S4	.03	-.20	1.00	S3 - S4	.04	1.75	.81
	S3 - S5	.03	2.22	.27	S4 - S5	.04	2.46	.14
	S4 - S5	.03	2.42	.16				
Bc pause	S1 - S2	.05	-1.51	1.00	S1 - S5	.04	-.86	1.00
frequency	S1 - S3	.05	-2.17	.31	S2 - S3	.04	-.65	1.00
	S1 - S4	.05	-.81	1.00	S2 - S4	.04	-.39	1.00
	S1 - S5	.05	1.70	.90	S2 - S5	.04	2.96	.03
	S2 - S3	.05	-.66	1.00	S3 - S4	.04	.26	1.00
	S2 - S4	.05	.70	1.00				
	S2 - S5	.05	3.21	.01				
	S3 - S4	.05	1.36	1.00				
	S4 - S5	.05	2.51	.13				
Bpc pause	S1 - S2	.04	2.35	.19	S1 - S2	.04	-1.88	.61
duration	S1 - S3	.04	2.34	.20	S1 - S3	.04	-2.03	.44
	S1 - S4	.04	.66	1.00	S1 - S4	.04	-.70	1.00
	S1 - S5	.04	.03	1.00	S1 - S5	.04	-1.87	.63
	S2 - S3	.04	.02	1.00	S2 - S3	.04	-.14	1.00
	S2 - S4	.04	-1.61	1.00	S2 - S4	.04	1.22	1.00
	S2 - S5	.04	-2.16	.32	S2 - S5	.04	-.08	1.00
	S3 - S4	.04	-1.62	1.00	S3 - S4	.04	1.36	1.00
	S3 - S5	.04	-2.15	.33	S3 - S5	.04	.06	1.00
	S4 - S5	.04	-.59	1.00	S4 - S5	.04	-1.23	1.00
Revision								
Total	S1 - S2	.04	-2.47	.14	S2 - S3	.04	-.56	1.00

	S1 - S3	.04	-2.79	.06	S2 - S4	.04	-.13	1.00
	S1 - S4	.04	-3.08	.02	S2 - S5	.04	3.18	.02
	S1 - S5	.04	-1.27	1.00	S3 - S4	.04	.44	1.00
	S2 - S3	.04	-.32	1.00	S4 - S5	.04	3.30	.01
	S2 - S4	.04	-.61	1.00				
	S2 - S5	.04	1.20	1.00				
	S3 - S4	.04	-.29	1.00				
	S3 - S5	.04	1.52	1.00				
	S4 - S5	.04	1.81	.71				
Below word		No stage effect			S1 - S5	.05	-.82	1.00
					S2 - S3	.05	-.75	1.00
					S2 - S4	.05	-.88	1.00
					S2 - S5	.05	3.26	.02
					S3 - S4	.05	-.13	1.00
Below clause	S1 - S2	.04	-1.73	.85	S1 - S4	.04	-3.29	.01
	S1 - S3	.04	-1.54	1.00	S1 - S5	.04	-2.84	.05
	S1 - S4	.04	-2.16	.31	S2 - S3	.04	-.54	1.00
	S1 - S5	.04	-1.61	1.00	S2 - S4	.04	.87	1.00
	S2 - S3	.04	.19	1.00	S2 - S5	.04	1.32	1.00
	S2 - S4	.04	-.44	1.00	S3 - S4	.04	1.41	1.00
	S2 - S5	.04	.12	1.00	S3 - S5	.04	1.86	.64
	S3 - S4	.04	-.63	1.00	S4 - S5	.04	.45	1.00
	S3 - S5	.04	-.07	1.00				
	S4 - S5	.04	.56	1.00				
Pre-contextual	S1 - S2	.05	-1.47	1.00	S1 - S5	.05	.51	1.00
	S1 - S3	.05	-1.92	.56	S2 - S3	.05	-.12	1.00
	S1 - S4	.05	-1.29	1.00	S2 - S4	.05	.92	1.00
	S1 - S5	.05	1.45	1.00	S3 - S4	.05	1.04	1.00
	S2 - S3	.05	-.45	1.00				
	S2 - S4	.05	.18	1.00				
	S2 - S5	.05	2.92	.04				
	S3 - S4	.05	.63	1.00				
	S4 - S5	.05	2.74	.07				
Pinyin	S1 - S2	.05	-.85	1.00	S1 - S5	.05	.17	1.00
	S1 - S3	.05	-1.22	1.00	S2 - S3	.05	-.61	1.00
	S1 - S4	.05	-.83	1.00	S2 - S4	.05	.51	1.00
	S1 - S5	.05	2.29	.23	S3 - S4	.05	1.12	1.00
	S2 - S3	.05	.37	1.00				
	S2 - S4	.05	.02	1.00				

S2 - S5	.05	3.14	.02
S3 - S4	.05	.39	1.00
S4 - S5	.05	3.12	.02

Note: Bpc = Between Pinyin and character(s), Bc = Between clauses

Table 8 Non-significant results from post-hoc Bonferroni tests examining the effects of stage on writing behaviours by genre in L1 Chinese (N = 32, p ≥ .01)

	Argumentative				*Narrative*			
	Stage	*SE*	*t*	*p*	Stage	*SE*	*t*	*p*
Pausing								
Total pause	S1 - S2	.14	-2.16	.32	S1 - S5	.16	-3.23	.01
frequency	S1 - S3	.14	-1.90	.59	S2 - S3	.16	.17	1.00
	S1 - S4	.14	-1.59	1.00	S2 - S4	.16	1.44	1.00
	S1 - S5	.14	-.50	1.00	S2 - S5	.16	1.55	1.00
	S2 - S3	.14	.26	1.00	S3 - S4	.16	1.26	1.00
	S2 - S4	.14	.57	1.00	S3 - S5	.16	1.38	1.00
	S2 - S5	.14	1.66	.98	S4 - S5	.16	.12	1.00
	S3 - S4	.14	.31	1.00				
	S3 - S5	.14	1.40	1.00				
	S4 - S5	.14	1.10	1.00				
Bpc pause		No stage effect			S1 - S2	.03	-3.29	.01
frequency					S1 - S3	.03	-1.43	1.00
					S1 - S4	.03	-3.05	.03
					S1 - S5	.03	-.08	1.00
					S2 - S3	.03	1.87	.63
					S2 - S4	.03	.24	1.00
					S2 - S5	.03	3.22	.01
					S3 - S4	.03	-1.62	1.00
					S3 - S5	.03	1.35	1.00
					S4 - S5	.03	2.97	.03
Bc pause		No stage effect			S1 - S2	.07	-2.76	.06
frequency					S1 - S5	.07	-1.75	.81
					S2 - S3	.07	-2.24	.26
					S2 - S4	.07	-2.08	.39
					S2 - S5	.07	1.01	1.00
					S3 - S4	.07	.16	1.00
					S3 - S5	.07	3.24	.01
					S4 - S5	.07	3.08	.02
Total pause		No stage effect			S1 - S2	.02	-2.15	.32
duration					S1 - S4	.02	-2.68	.08
					S1 - S5	.02	.35	1.00
					S2 - S3	.02	-1.36	1.00
					S2 - S4	.02	-.53	1.00
					S2 - S5	.02	2.51	.13
					S3 - S4	.02	.83	1.00
					S4 - S5	.02	3.04	.03
Bpc pause	S1 - S2	.07	1.00	1.00		No stage effect		

duration	S1 - S3	.07	2.78	.06				
	S1 - S4	.07	< -.01	1.00				
	S1 - S5	.07	1.04	1.00				
	S2 - S3	.07	1.92	.57				
	S2 - S4	.07	-.96	1.00				
	S2 - S5	.07	.13	1.00				
	S3 - S4	.07	-2.73	.08				
	S3 - S5	.07	-1.69	.94				
	S4 - S5	.07	1.03	1.00				
Bc pause	No stage effect				S1 - S2	.08	2.08	.38
duration					S1 - S3	.08	1.27	1.00
					S1 - S4	.08	1.51	1.00
					S1 - S5	.08	1.34	1.00
					S2 - S3	.08	-.86	1.00
					S2 - S4	.08	-.60	1.00
					S2 - S5	.08	-.76	1.00
					S3 - S4	.07	.26	1.00
					S3 - S5	.07	.09	1.00
					S4 - S5	.08	-.16	1.00
Bs pause	S2 - S3	.11	.44	1.00	S2 - S3	.13	2.16	.31
duration	S2 - S4	.10	1.66	.97	S2 - S4	.13	1.39	1.00
	S2 - S5	.10	-1.68	.94	S2 - S5	.13	-2.12	.35
	S3 - S4	.10	1.22	1.00	S3 - S4	.13	-.78	1.00
	S3 - S5	.10	-2.13	.34				
Revision								
Total	S1 - S3	.04	-2.70	.07	S1 - S5	.29	-2.71	.07
	S1 - S5	.04	-1.70	.90	S2 - S3	.29	-1.71	.88
	S2 - S3	.04	.82	1.00	S2 - S4	.29	-.87	1.00
	S2 - S4	.04	-.07	1.00	S2 - S5	.29	2.05	.41
	S2 - S5	.04	1.82	.70	S3 - S4	.29	.85	1.00
	S3 - S4	.04	-.89	1.00	S4 - S5	.29	2.92	.04
	S3 - S5	.04	1.00	1.00				
	S4 - S5	.04	1.89	.60				
Below word	No stage effect				S1 - S5	.05	-1.76	.80
					S2 - S3	.05	-1.72	.86
					S2 - S4	.05	-1.09	1.00
					S2 - S5	.05	2.46	.14
					S3 - S4	.05	.63	1.00
Pre-contextual	S1 - S2	.04	-2.67	.08	S1 - S5	.06	-2.35	.19
	S1 - S3	.04	-1.77	.78	S2 - S3	.06	-1.47	1.00
	S1 - S4	.04	-2.81	.05	S2 - S4	.06	-1.05	1.00
	S1 - S5	.04	-.09	1.00	S2 - S5	.06	2.64	.09
	S2 - S3	.04	.91	1.00	S3 - S4	.06	.42	1.00
	S2 - S4	.04	-.14	1.00				

	S2 - S5	.04	2.58	.10				
	S3 - S4	.04	-1.05	1.00				
	S3 - S5	.04	1.68	.94				
	S4 - S5	.04	2.72	.07				
Pinyin	No stage effect				S1 - S5	.05	-1.51	1.00
					S2 - S3	.05	-.89	1.00
					S2 - S4	.05	-.69	1.00
					S2 - S5	.05	2.77	.06
					S3 - S4	.05	.20	1.00
Character	S1 - S2	.04	-2.93	.04	S2 - S3	.05	-1.82	.71
	S1 - S3	.04	-2.71	.07	S2 - S4	.05	-.83	1.00
	S1 - S4	.04	-3.07	.02	S2 - S5	.05	.32	1.00
	S1 - S5	.04	-1.92	.55	S3 - S4	.05	.98	1.00
	S2 - S3	.04	.22	1.00	S3 - S5	.05	2.14	.33
	S2 - S4	.04	-.14	1.00	S4 - S5	.05	1.16	1.00
	S2 - S5	.04	1.01	1.00				
	S3 - S4	.04	-.36	1.00				
	S3 - S5	.04	.79	1.00				
	S4 - S5	.04	1.15	1.00				

Note: Bpc = Between Pinyin and character(s), Bc = Between clauses, Bs = Between sentences

Table 9 Non-significant results from linear mixed-effects regressions examining the relationships between L2 proficiency and writing behaviours (N = 32)

Dependent variable	Fixed effect	*Est*	*SE*	*t*	*p*
Speed fluency					
P-burst length	Proficiency	< .01	.01	.72	.48
Model = lmer(log(pburst_length) ~ proficiency + (1|participant) + (1|prompt), data, REML=F)					
Pause frequency					
Total pause	Proficiency	-.01	< .01	-1.94	.06
Model = lmer(log(total_pause_frequency) ~ proficiency + (1|participant) + (1|prompt), data, REML=F)					
Bw pause	Proficiency	-.01	.01	-1.26	.22
Model = lmer(bw_pause_frequency ~ proficiency + (1|participant) + (1|prompt), data, REML=F)					
Bc pause frequency	Proficiency	< -.01	< .01	-.25	.81
Model = lmer(bc_pause_frequency ~ proficiency + (1|participant) + (1|prompt), data, REML=F)					
Bs pause	Proficiency	< -.01	< .01	-.46	.65
Model = lmer(bs_pause_frequency ~ proficiency + (1|participant) + (1|prompt), data, REML=F)					
Bcr pause	Proficiency	< -.01	< .01	-.83	.41
Model = lmer(1/log(bcr_pause_frequency) ~ proficiency + (1|participant) + (1|prompt), data, REML=F)					
Pause duration					
Total pause	Proficiency	.02	.02	.95	.35
Model = lmer(total_pause_duration ~ proficiency + (1|participant) + (1|prompt), data, REML=F)					
Ww pause	Proficiency	< -.01	< .01	-1.43	.16
Model = lmer(1/log(ww_pause_duration) ~ proficiency + (1|participant) + (1|prompt), data, REML=F)					
Bpc pause	Proficiency	< -.01	< .01	-.41	.69
Model = lmer(1/log(bpc_pause_duration) ~ proficiency + (1|participant) + (1|prompt), data, REML=F)					
Bw pause	Proficiency	.01	< .01	.56	.58
Model = lmer(log(bw_pause_duration) ~ proficiency + (1|participant) + (1|prompt), data, REML=F)					
Bc pause	Proficiency	< -.01	.01	.21	.83
Model = lmer(log(bc_pause_duration) ~ proficiency + (1|participant), data, REML=F)					
Bs pause	Proficiency	-.01	.01	-1.54	.13
Model = lmer(log(bs_pause_duration) ~ proficiency + (1|participant) + (1|prompt), data, REML=F)					
Bcr pause	Proficiency	.01	.01	.55	.58
Model = lmer(log(bcr_pause_duration) ~ proficiency + (1|participant), data, REML=F)					
Revision					
Below word	Proficiency	.01	.01	1.56	.13
Model = lmer(log(below_word_revision) ~ proficiency + (1|participant), data, REML=F)					
Word	Proficiency	.01	.01	1.55	.13

Model = lmer(log(word_revision) ~ proficiency + (1|participant), data, REML=F)					
Pre-contextual	Proficiency	.02	.01	1.96	.06
Model = lmer(log(precontextual_revision) ~ proficiency + (1|participant), data, REML=F)					
Contextual	Proficiency	.01	.01	1.28	.21
Model = lmer(log(contextual_revision) ~ proficiency + (1|participant), data, REML=F)					
Pinyin	Proficiency	.01	.01	1.27	.21
Model = lmer(log(pinyin_revision) ~ proficiency + (1|participant), data, REML=F)					

Note: Ww = Within a word, Bpc = Between Pinyin and character(s), Bw = Between words, Bc = Between clauses, Bs = Between sentences, Bcr = Between contextual revisions

Table 10 Non-significant results from linear mixed-effects regressions examining how stage mediated the relationships between L2 proficiency and writing behaviours (N = 32)

Dependent variable	Predictor: proficiency * stage	*Est*	*SE*	*t*	*p*
Speed fluency					
Production rate	Proficiency * 2	< -.01	< .01	-.44	.66
	Proficiency * 3	-.01	< .01	-1.34	.18
	Proficiency * 4	< .01	< .01	.05	.96
	Proficiency * 5	< -.01	< .01	-1.17	.24
Model = lmer(log(production_rate) ~ stage*cproficiency + (1|participant) + (1|prompt), data, REML=F)					
Pause frequency					
Total pause	Proficiency * 2	-.02	.01	-1.35	.18
	Proficiency * 3	-.01	.01	-.80	.42
	Proficiency * 4	< .01	.01	.24	.81
	Proficiency * 5	.02	.01	1.21	.23
Model = lmer(total_pause_frequency) ~ stage*cproficiency + (1|participant) + (1|prompt), data, REML=F)					
Ww pause	Proficiency * 2	< -.01	< .01	-.56	.57
	Proficiency * 3	< -.01	< .01	-.97	.33
	Proficiency * 4	< .01	< .01	1.20	.23
	Proficiency * 5	< .01	< .01	1.56	.12
Model = lmer(ww_pause_frequency) ~ stage*cproficiency + (1|participant) + (1|prompt), data, REML=F)					
Bpc pause	Proficiency * 2	-.01	< .01	-1.52	.13
	Proficiency * 3	< -.01	< .01	-.23	.82
	Proficiency * 4	< .01	< .01	1.03	.31
	Proficiency * 5	< .01	< .01	.16	.87
Model = lmer(log(bpc_pause_frequency) ~ stage*cproficiency + (1|participant) + (1|prompt), data, REML=F)					
Bw pause	Proficiency * 2	-.01	.01	-1.20	.23
	Proficiency * 3	-.01	.01	-.94	.35
	Proficiency * 4	< .01	.01	.21	.83
	Proficiency * 5	-.01	.01	-.58	.56
Model = lmer(bw_pause_frequency ~ stage*cproficiency + (1|participant) + (1|prompt), data, REML=F)					
Bc pause	Proficiency * 2	< -.01	.01	-.29	.77
	Proficiency * 3	-.01	.01	-1.20	.23
	Proficiency * 4	< -.01	.01	-.64	.52
	Proficiency * 5	< -.01	.01	-.06	.95

Model = lmer(bc_pause_frequency ~ stage*cproficiency + (1|participant) + (1|prompt), data, REML=F)					
Bs pause	Proficiency * 2	< .01	.01	.58	.56
	Proficiency * 3	< .01	.01	.13	.90
	Proficiency * 4	< .01	.01	.18	.85
	Proficiency * 5	01	.01	1.08	.28
Model = lmer(bs_pause_frequency ~ stage*cproficiency + (1|participant) + (1|prompt), data, REML=F)					
Bcr pause	Proficiency * 2	< -.01	.01	-.01	.99
	Proficiency * 3	< .01	.01	.92	.36
	Proficiency * 4	< -.01	.01	-.89	.37
	Proficiency * 5	.01	.01	1.46	.14
Model = lmer(log(bcr_pause_frequency) ~ stage*cproficiency + (1|participant) + (1|prompt), data, REML=F)					
Pause duration					
Total pause	Proficiency * 2	< .01	< .01	1.45	.15
	Proficiency * 3	< .01	< .01	.81	.42
	Proficiency * 4	< .01	< .01	.37	.71
	Proficiency * 5	< .01	< .01	1.03	.31
Model = lmer(1/log(total_pause_duration) ~ stage*cproficiency + (1|participant), data, REML=F)					
Ww pause	Proficiency * 2	.01	.01	1.66	.10
	Proficiency * 3	.01	.01	.99	.32
	Proficiency * 4	< -.01	.01	-.13	.90
	Proficiency * 5	.01	.01	1.58	.12
Model = lmer(1/log(ww_pause_duration) ~ stage*cproficiency + (1|participant), data, REML=F)					
Bpc pause	Proficiency * 2	< -.01	< .01	-.12	.90
	Proficiency * 3	< -.01	< .01	-.78	.44
	Proficiency * 4	< -.01	< .01	-.39	.70
	Proficiency * 5	< .01	< .01	.54	.59
Model = lmer(1/log(bpc_pause_duration) ~ stage*cproficiency + (1|participant), data, REML=F)					
Bw pause	Proficiency * 2	< .01	< .01	1.45	.15
	Proficiency * 3	< .01	< .01	.28	.78
	Proficiency * 4	< -.01	< .01	-.43	.67
	Proficiency * 5	< .01	< .01	1.04	.30
Model = lmer(1/log(bw_pause_duration) ~ stage*cproficiency + (1|participant) + (1|prompt), data, REML=F)					
Bs pause	Proficiency * 2	< .01	.01	.33	.74
	Proficiency * 3	-.01	.01	-.66	.51
	Proficiency * 4	-.02	.01	-1.56	.12
	Proficiency * 5	-.02	.01	-1.89	.06
Model = lmer(log(bs_pause_duration) ~ stage*cproficiency + (1|participant), data, REML=F)					
Bcr pause	Proficiency * 2	< .01	.02	.12	.90
	Proficiency * 3	< .01	.02	.24	.81
	Proficiency * 4	.02	.02	1.54	.13

	Proficiency * 5	.01	.02	.45	.65
Model = lmer(log(bcr_pause_duration) ~ stage*cproficiency + (1|participant), data, REML=F)					
Revision					
Total	Proficiency * 2	-.01	< .01	-1.64	.10
	Proficiency * 3	< -.01	< .01	.53	.60
	Proficiency * 4	< -.01	< .01	.06	.95
	Proficiency * 5	< -.01	< .01	1.09	.27
Model = lmer(log(total_revision) ~ stage*cproficiency + (1|participant), data, REML=F)					
Below word	Proficiency * 2	-.01	< .01	-1.15	.25
	Proficiency * 3	< -.01	< .01	-.28	.78
	Proficiency * 4	< .01	< .01	.26	.80
	Proficiency * 5	< -.01	< .01	-.33	.74
Model = lmer(log(below_word_revision) ~ stage*cproficiency + (1|participant), data, REML=F)					
Word	Proficiency * 2	< -.01	< .01	-1.12	.26
	Proficiency * 3	< -.01	< .01	-.30	.76
	Proficiency * 4	< -.01	< .01	-.13	.90
	Proficiency * 5	< .01	< .01	.76	.45
Model = lmer(log(word_revision) ~ stage*cproficiency + (1|participant), data, REML=F)					
Clause and above	Proficiency * 2	< .01	< .01	.63	.53
	Proficiency * 3	< .01	< .01	.76	.45
	Proficiency * 4	< .01	< .01	.33	.74
	Proficiency * 5	< .01	< .01	.36	.72
Model = lmer(log(clause_and_above_revision) ~ stage*cproficiency + (1|participant), data, REML=F)					
Pre-contextual	Proficiency * 2	-.01	< .01	-1.50	.13
	Proficiency * 3	< -.01	< .01	-.27	.79
	Proficiency * 4	< .01	< .01	.54	.59
	Proficiency * 5	< .01	< .01	.35	.72
Model = lmer(log(precontextual_revision) ~ stage*cproficiency + (1|participant), data, REML=F)					
Contextual	Proficiency * 2	< .01	< .01	.32	.75
	Proficiency * 3	.01	< .01	1.51	.13
	Proficiency * 4	< -.01	< .01	-.47	.64
	Proficiency * 5	.01	< .01	1.54	.12
Model = lmer(log(contextual_revision) ~ stage*cproficiency + (1|participant), data, REML=F)					

Note: Ww = Within a word, Bpc = Between Pinyin and character(s), Bw = Between words, Bc = Between clauses, Bs = Between sentences, Bcr = Between contextual revisions

Table 11 Non-significant results from linear mixed-effects regressions examining how genre mediated the relationships between L2 proficiency and writing behaviours (N = 32)

Dependent variable	Predictor: proficiency * genre	*Est*	*SE*	*t*	*p*
Speed fluency					
Production rate	Proficiency * narrative	.07	.09	.74	.47
Model = lmer(production_rate ~ genre*cproficiency + (1+genre|participant), data, REML = F)					
P-burst length	Proficiency * narrative	-.01	.01	-.61	.55
Model = lmer(pburst_length ~ genre*cproficiency + (1+genre|participant), data, REML = F)					
Pause frequency					
Total pause	Proficiency * narrative	-.01	.01	-.50	.62
Model = lmer(total_pause_frequency ~ genre*cproficiency + (1|participant) + (1|prompt), data, REML = F)					
Ww pause	Proficiency * narrative	< -.01	< .01	-.18	.86
Model = lmer(ww_pause_frequency ~ genre*cproficiency + (1|participant),data, REML = F)					
Bpc pause	Proficiency * narrative	< -.01	< .01	-.63	.53
Model = lmer(log(bpc_pause_frequency) ~ genre*cproficiency + (1|participant) + (1|prompt), data, REML = F)					
Bw pause	Proficiency * narrative	-.01	< .01	-1.16	.25
Model = lmer(log(bw_pause_frequency) ~ genre*cproficiency + (1|participant),data, REML = F)					
Bc pause	Proficiency * narrative	< .01	.01	.01	1.00
Model = lmer(bc_pause_frequency ~ genre*cproficiency + (1|participant) + (1|prompt), data, REML = F)					
Bs pause	Proficiency * narrative	< -.01	.01	-.59	.56
Model = lmer(bs_pause_frequency ~ genre*cproficiency + (1|participant) + (1|prompt), data, REML = F)					
Pause duration					
Total pause	Proficiency * narrative	< -.01	< .01	-.59	.56
Model = lmer(log(total_pause_duration) ~ genre*cproficiency + (1|participant) + (1|prompt), data, REML = F)					
Ww pause	Proficiency * narrative	< .01	.01	.33	.75
Model = lmer(log(ww_pause_duration) ~ genre*cproficiency + (1e|participant), data, REML = F)					
Bpc pause	Proficiency * narrative	< .01	< .01	.21	.83
Model = lmer(1/log(bpc_pause_duration) ~ genre*cproficiency + (1|participant) + (1|prompt), data, REML = F)					
Bw pause	Proficiency * narrative	< -.01	.01	-1.14	.26
Model = lmer(log(bw_pause_duration) ~ genre*cproficiency + (1|participant), data, REML = F)					
Bc pause	Proficiency * narrative	-.01	.01	-1.64	.10
Model = lmer(log(bc_pause_duration) ~ genre*cproficiency + (1|participant), data, REML = F)					
Bs pause	Proficiency * narrative	-.01	.01	-.82	.41
Model = lmer(log(bs_pause_duration) ~ genre*cproficiency + (1|participant), data, REML = F)					

<table>
<tr><td>Bcr pause</td><td>Proficiency * narrative</td><td>-.02</td><td>.01</td><td>-1.35</td><td>.18</td></tr>
<tr><td colspan="6">Model = lmer(log(bcr_pause_duration) ~ genre*cproficiency + (1|participant), data, REML = F)</td></tr>
<tr><td colspan="6">Revision</td></tr>
<tr><td>Total</td><td>Proficiency * narrative</td><td>< .01</td><td>.01</td><td>.42</td><td>.68</td></tr>
<tr><td colspan="6">Model = lmer(log(total_revision) ~ genre*cproficiency + (1+genre|participant), data, REML = F)</td></tr>
<tr><td>Below word</td><td>Proficiency * narrative</td><td>< .01</td><td>< .01</td><td>.26</td><td>.79</td></tr>
<tr><td colspan="6">Model = lmer(log(below_word_revision ~ genre*cproficiency + (1|participant), data, REML = F)</td></tr>
<tr><td>Word</td><td>Proficiency * narrative</td><td>< .01</td><td>< .01</td><td>.17</td><td>.87</td></tr>
<tr><td colspan="6">Model = lmer(log(word_revision) ~ genre*cproficiency + (1|participant), data, REML = F)</td></tr>
<tr><td>Below clause</td><td>Proficiency * narrative</td><td>< -.01</td><td>< .01</td><td>-.16</td><td>.87</td></tr>
<tr><td colspan="6">Model = lmer(log(below_clause_revision) ~ genre*cproficiency + (1|participant) + (1|prompt), data, REML = F)</td></tr>
<tr><td>Clause and above</td><td>Proficiency * narrative</td><td>< -.01</td><td>< .01</td><td>-.48</td><td>.64</td></tr>
<tr><td colspan="6">Model = lmer(log(clause_and_above_revision) ~ genre*cproficiency + (1|participant), data, REML = F)</td></tr>
<tr><td>Pre-contextual</td><td>Proficiency * narrative</td><td>< .01</td><td>< .01</td><td>.28</td><td>.78</td></tr>
<tr><td colspan="6">Model = lmer(log(precontextual_revision) ~ genre*cproficiency + (1+genre|participant), data, REML = F)</td></tr>
<tr><td>Contextual</td><td>Proficiency * narrative</td><td>.01</td><td>.01</td><td>.76</td><td>.45</td></tr>
<tr><td colspan="6">Model = lmer(log(contextual_revision) ~ genre*cproficiency + (1|participant), data, REML = F)</td></tr>
<tr><td>Pinyin</td><td>Proficiency * narrative</td><td>< -.01</td><td>< .01</td><td>-.09</td><td>.93</td></tr>
<tr><td colspan="6">Model = lmer(log(pinyin_revision) ~ genre*cproficiency + (1+genre|participant),data, REML = F)</td></tr>
<tr><td>Character</td><td>Proficiency * narrative</td><td>< .01</td><td>< .01</td><td>.47</td><td>.64</td></tr>
<tr><td colspan="6">Model = lmer(log(character_revision) ~ genre*cproficiency + (1+genre|participant),data, REML = F)</td></tr>
</table>

Note: Ww = Within a word, Bpc = Between Pinyin and character(s), Bw = Between words, Bc = Between clauses, Bs = Between sentences, Bcr = Between contextual revisions

Table 12 Non-significant results from linear mixed-effects regressions examining how writing behaviours predict L2 Chinese text quality (N = 32)

Dependent variable	Predictor	*Est*	*SE*	*t*	*p*
	Pause frequency				
Text quality	Total pause	-.13	.07	-1.81	.07
Model = lmer(text_score ~ total_pause_frequency + (1|participant) + (1|prompt), data, REML = F)					
Text quality	Ww pause	-.17	.46	-.36	.71
Model = lmer(text_score ~ ww_pause_frequency + (1|participant) + (1|prompt), data, REML = F)					
Text quality	Bpc pause	.15	.22	.70	.50
Model = lmer(text_score ~ bpc_pause_frequency + (1|participant) + (1|prompt), data, REML = F)					
Text quality	Bc pause	.24	.28	.88	.38
Model = lmer(text_score ~ bc_pause_frequency + (1|participant) + (1|prompt), data, REML = F)					
Text quality	Bs pause	.06	.26	.22	.82
Model = lmer(text_score ~ bs_pause_frequency + (1|participant) + (1|prompt), data, REML = F)					
Text quality	Bcr pause	.07	.20	.33	.74
Model = lmer(text_score ~ bcr_pause_frequency + (1|participant) + (1|prompt), data, REML = F)					
	Pause duration				
Text quality	Ww pause	.03	.04	.74	.46
Model = lmer(text_score ~ ww_pause_duration + (1|participant) + (1|prompt), data, REML = F)					
Text quality	Bpc pause	-.03	.06	-.39	.70
Model = lmer(text_score ~ bpc_pause_duration + (1|participant) + (1|prompt), data, REML = F)					
Text quality	Bs pause	-.01	.01	-.77	.44
Model = lmer(text_score ~ bs_pause_duration + (1|participant) + (1|prompt), data, REML = F)					
Text quality	Bcr pause	.01	.01	1.07	.29
Model = lmer(text_score ~ bcr_pause_duration + (1|participant) + (1|prompt), data, REML = F)					
Text quality	Bcr pause	.01	.01	1.07	.29
Model = lmer(text_score ~ bcr_pause_duration + (1|participant) + (1|prompt), data, REML = F)					
	Revision				
Text quality	Total	.13	.23	.54	.59
Model = lmer(text_score ~ log(total_revision) + (1|participant) + (1|prompt), data, REML = F)					
Text quality	Below word	.19	.23	.81	.42
Model = lmer(text_score ~ log(below_word_revision) + (1|participant), data, REML = F)					
Text quality	Word	-.03	.26	-.10	.92
Model = lmer(text_score ~ log(word_revision) + (1|participant) + (1|prompt), data, REML = F)					
Text quality	Below clause	.22	.31	.70	.49
Model = lmer(text_score ~ log(below_clause_revision) + (1|participant) + (1|prompt), data, REML = F)					
Text quality	Clause and above	-.30	.46	-.65	.52
Model = lmer(text_score ~ log(clause_and_above_revision) + (1|participant) + (1|prompt), data, REML = F)					
Text quality	Pre-contextual	.08	.24	.33	.74
Model = lmer(text_score ~ log(precontextual_revision) + (1|participant) + (1|prompt), data, REML					

= F)					
Text quality	Contextual	.14	.24	.58	.56
Model = lmer(text_score ~ log(contextual_revision) + (1|participant) + (1|prompt), data, REML = F)					
Text quality	Pinyin	.09	.23	.39	.70
Model = lmer(text_score ~ log(pinyin_revision) + (1|participant) + (1|prompt), data, REML = F)					
Text quality	Character	.08	.24	.30	.77
Model = lmer(text_score ~ log(character_revision) + (1|participant) + (1|prompt), data, REML = F)					

Note: Ww = Within a word, Bpc = Between Pinyin and character(s), Bc = Between clauses, Bs = Between sentences, Bcr = Between contextual revisions

Table 13 Non-significant results from linear mixed-effects regressions examining how writing behaviours predict L1 Chinese text quality (N = 32)

Dependent variable	Predictor	*Est*	*SE*	*t*	*p*
	Speed fluency				
Text quality	Production rate	.01	.01	.93	.36
Model = lmer(text_score ~ production_rate + (1|participant) + (1|prompt), data, REML = F)					
Text quality	P-burst length	.01	.03	.44	.66
Model = lmer(text_score ~ pburst_length + (1|participant) + (1|prompt), data, REML = F)					
	Pause frequency				
Text quality	Total pause	-.06	.08	-.81	.42
Model = lmer(text_score ~ total_pause_frequency + (1|participant) + (1|prompt), data, REML = F)					
Text quality	Ww pause	-.06	.66	-.09	.93
Model = lmer(text_score ~ ww_pause_frequency + (1|participant) + (1|prompt), data, REML = F)					
Text quality	Bpc pause	-.29	.38	-.78	.44
Model = lmer(text_score ~ bpc_pause_frequency + (1|participant) + (1|prompt), data, REML = F)					
Text quality	Bw pause	-.08	.10	-.73	.47
Model = lmer(text_score ~ bw_pause_frequency + (1|participant) + (1|prompt), data, REML = F)					
Text quality	Bc pause	-.10	.18	-.57	.57
Model = lmer(text_score ~ bc_pause_frequency + (1|participant) + (1|prompt), data, REML = F)					
Text quality	Bs pause	.18	.22	.84	.41
Model = lmer(text_score ~ bs_pause_frequency + (1|participant) + (1|prompt), data, REML = F)					
Text quality	Bcr pause	-.08	.24	-.33	.74
Model = lmer(text_score ~ bcr_pause_frequency + (1|participant) + (1|prompt), data, REML = F)					
	Pause duration				
Text quality	Total pause	-.10	.08	-1.21	.23
Model = lmer(text_score ~ total_pause_duration + (1|participant) + (1|prompt), data, REML = F)					
Text quality	Ww pause	.04	.04	.93	.36
Model = lmer(text_score ~ (ww_pause_duration)^2 + (1|participant) + (1|prompt), data, REML = F)					
Text quality	Bpc pause	.01	.04	.16	.88
Model = lmer(text_score ~ bpc_pause_duration + (1|participant) + (1|prompt), data, REML = F)					
Text quality	Bw pause	-.06	.09	-.67	.51
Model = lmer(text_score ~ bw_pause_duration + (1|participant) + (1|prompt), data, REML = F)					
Text quality	Bc pause	-.02	.04	-.52	.61
Model = lmer(text_score ~ bc_pause_duration + (1|participant) + (1|prompt), data, REML = F)					
Text quality	Bs pause	-.03	.02	-1.66	.10
Model = lmer(text_score ~ bs_pause_duration + (1|participant) + (1|prompt), data, REML = F)					
Text quality	Bcr pause	-.01	.01	-.40	.69
Model = lmer(text_score ~ bcr_pause_duration + (1|participant) + (1|prompt), data, REML = F)					
	Revision				
Text quality	Total	.14	.23	.60	.55
Model = lmer(text_score ~ log(total_revision) + (1|participant) + (1|prompt), data, REML = F)					

Text quality	Below word	.01	.22	.07	.95
Model = lmer(text_score ~ log(below_word_revision) + (1|participant), data, REML = F)					
Text quality	Word	.25	.24	-1.03	.31
Model = lmer(text_score ~ log(word_revision) + (1|participant) + (1|prompt), data, REML = F)					
Text quality	Below clause	.12	.23	.52	.61
Model = lmer(text_score ~ log(below_clause_revision) + (1|participant) + (1|prompt), data, REML = F)					
Text quality	Clause and above	.08	.32	.24	.81
Model = lmer(text_score ~ log(clause_and_above_revision) + (1|participant) + (1|prompt), data, REML = F)					
Text quality	Pre-contextual	.17	.21	.80	.43
Model = lmer(text_score ~ log(precontextual_revision) + (1|participant) + (1|prompt), data, REML = F)					
Text quality	Contextual	-.24	.22	-1.09	.28
Model = lmer(text_score ~ log(contextual_revision) + (1|participant) + (1|prompt), data, REML = F)					
Text quality	Pinyin	< .01	.21	.02	.98
Model = lmer(text_score ~ log(pinyin_revision) + (1|participant) + (1|prompt), data, REML = F)					
Text quality	Character	.18	.20	.91	.37
Model = lmer(text_score ~ log(character_revision) + (1|participant) + (1|prompt), data, REML = F)					

Note: Ww = Within a word, Bpc = Between Pinyin and character(s), Bc = Between clauses, Bs = Between sentences, Bcr = Between contextual revisions

Table 14 Results from linear mixed-effects regressions examining the effects of interaction between genre and writing behaviours on L2 Chinese text quality (N = 32)

Dependent variable	Predictor	*Est*	*SE*	*t*	*p*
	Speed fluency				
Text quality	Production rate * genre (narrative)	< -.01	.01	-.21	.83
Model = lmer(text_score ~ production_rate*genre + (1|participant), data, REML = F)					
Text quality	P-burst length * genre (narrative)	< -.01	.04	< -.01	1.00
Model = lmer(text_score ~ pburst_length*genre + (1|participant), data, REML = F)					
	Pause frequency				
Text quality	Total pause * genre (narrative)	.02	.07	.28	.78
Model = lmer(text_score ~ total_pause_frequency*genre + (1|participant), data, REML = F)					
Text quality	Ww pause * genre (narrative)	-.70	.63	-1.10	.27
Model = lmer(text_score ~ ww_pause_frequency*genre + (1|participant), data, REML = F)					
Text quality	Bpc pause * genre (narrative)	.01	.17	.08	.93
Model = lmer(text_score ~ bpc_pause_frequency*genre + (1|participant), data, REML = F)					
Text quality	Bw pause * genre (narrative)	-.03	.11	-.24	.81
Model = lmer(text_score ~ bw_pause_frequency*genre + (1|participant), data, REML = F)					
Text quality	Bc pause * genre (narrative)	.31	.35	-.89	.37
Model = lmer(text_score ~ bc_pause_frequency*genre + (1|participant), data, REML = F)					
Text quality	Bs pause * genre (narrative)	.28	.29	.96	.34
Model = lmer(text_score ~ bs_pause_frequency*genre + (1|participant), data, REML = F)					
Text quality	Bcr pause * genre (narrative)	.13	.20	.67	.50
Model = lmer(text_score ~ bcr_pause_frequency*genre + (1|participant), data, REML = F)					
	Pause duration				
Text quality	Total pause * genre (narrative)	-.02	.07	-.31	.76
Model = lmer(text_score ~ total_pause_duration*genre + (1|participant) + (1|prompt), data, REML = F)					
Text quality	Ww pause * genre (narrative)	-.01	.07	-.17	.86
Model = lmer(text_score ~ ww_pause_duration*genre + (1|participant), data, REML = F)					
Text quality	Bpc pause * genre (narrative)	-.05	.11	-.47	.64

Model = lmer(text_score ~ bpc_pause_duration*genre + (1|participant), data, REML = F)					
Text quality	Bw pause * genre (narrative)	-.07	.09	-.86	.40
Model = lmer(text_score ~ bc_pause_duration*genre + (1+genre|participant), data, REML = F)					
Text quality	Bc pause * genre (narrative)	-.01	.04	.22	.82
Model = lmer(text_score ~ bc_pause_duration*genre + (1|participant), data, REML = F)					
Text quality	Bs pause * genre (narrative)	-.02	.01	-1.42	.16
Model = lmer(text_score ~ bs_pause_duration*genre + (1|participant), data, REML = F)					
Text quality	Bcr pause * genre (narrative)	.01	.01	1.18	.24
Model = lmer(text_score ~ bcr_pause_duration*genre + (1|participant), data, REML = F)					
	Revision				
Text quality	Total * genre (narrative)	-.20	.17	-1.21	.23
Model = lmer(text_score ~ log(total_revision)*genre + (1|participant), data, REML = F)					
Text quality	Below word * genre (narrative)	-.24	.17	-1.47	.14
Model = lmer(text_score ~ log(below_word_revision)*genre + (1|participant), data, REML = F)					
Text quality	Word * genre (narrative)	-.10	.25	-.41	.68
Model = lmer(text_score ~ log(word_revision)*genre + (1|participant), data, REML = F)					
Text quality	Below clause * genre (narrative)	-.18	.30	-.60	.55
Model = lmer(text_score ~ log(below_clause_revision)*genre + (1|participant), data, REML = F)					
Text quality	Clause-and-above * genre (narrative)	.29	.60	.49	.63
Model = lmer(text_score ~ log(clause_and_above_revision)*genre + (1|participant) + (1|prompt), data, REML = F)					
Text quality	Pre-contextual * genre (narrative)	-.21	.16	-1.28	.20
Model = lmer(text_score ~ log(precontextual_revision)*genre + (1|participant), data, REML = F)					
Text quality	Contextual * genre (narrative)	.06	.25	.23	.82
Model = lmer(text_score ~ log(contextual_revision)*genre + (1|participant), data, REML = F)					
Text quality	Pinyin * genre (narrative)	-.21	.17	-1.28	.21
Model = lmer(text_score ~ log(pinyin_revision)*genre + (1|participant), data, REML = F)					
Text quality	Character * genre (narrative)	-.07	.18	-.37	.71
Model = lmer(text_score ~ log(character_revision)*genre + (1|participant), data, REML = F)					

Note: Ww = Within a word, Bpc = Between Pinyin and character(s), Bw = Between words, Bc = Between clauses, Bs = Between sentences, Bcr = Between contextual revisions

Table 15 Non-significant results from linear mixed-effects regressions examining the effects of interaction between genre and writing behaviours on L1 Chinese text quality (N = 32)

Dependent variable	Predictor	*Est*	*SE*	*t*	*p*
	Speed fluency				
Text quality	Production rate * genre (narrative)	< -.01	.01	-.54	.59
Model = lmer(text_score ~ production_rate*genre + (1|participant), data, REML = F)					
Text quality	P-burst length * genre (narrative)	-.03	.05	-.65	.52
Model = lmer(text_score ~ pburst_length*genre + (1|participant), data, REML = F)					
	Pause frequency				
Text quality	Total pause * genre (narrative)	.01	.10	.09	.93
Model = lmer(text_score ~ total_pause_frequency*genre + (1|participant), data, REML = F)					
Text quality	Ww pause * genre (narrative)	.90	1.06	.84	.40
Model = lmer(text_score ~ ww_pause_frequency*genre + (1|participant), data, REML = F)					
Text quality	Bpc pause * genre (narrative)	.42	.46	.92	.36
Model = lmer(text_score ~ bpc_pause_frequency*genre + (1|participant), data, REML = F)					
Text quality	Bw pause * genre (narrative)	-.16	.16	-.99	.32
Model = lmer(text_score ~ bw_pause_frequency*genre + (1|participant), data, REML = F)					
Text quality	Bc pause * genre (narrative)	-.20	.23	-.84	.40
Model = lmer(text_score ~ bc_pause_frequency*genre + (1|participant), data, REML = F)					
Text quality	Bs pause * genre (narrative)	-.05	.32	-.17	.87
Model = lmer(text_score ~ bs_pause_frequency*genre + (1+genre|participant) + (1|prompt), data, REML = F)					
	Pause duration				
Text quality	Total pause * genre (narrative)	.02	.12	.18	.85
Model = lmer(text_score ~ total_pause_duration*genre + (1|participant), data, REML = F)					
Text quality	Ww pause * genre (narrative)	-.01	.08	-.07	.94
Model = lmer(text_score ~ ww_pause_duration*genre + (1|participant), data, REML = F)					
Text quality	Bpc pause * genre (narrative)	.04	.10	.41	.68
Model = lmer(text_score ~ bpc_pause_duration*genre + (1+genre|participant) + (1|prompt), data,					

REML = F)					
Text quality	Bw pause * genre (narrative)	-.08	.15	-.51	.61
Model = lmer(text_score ~ bc_pause_duration*genre + (1|participant), data, REML = F)					
Text quality	Bc pause * genre (narrative)	-.04	.01	-.72	.47
Model = lmer(text_score ~ bc_pause_duration*genre + (1|participant), data, REML = F)					
Text quality	Bs pause * genre (narrative)	< -.01	.03	-.14	.89
Model = lmer(text_score ~ bs_pause_duration*genre + (1|participant), data, REML = F)					
Text quality	Bcr pause * genre (narrative)	-.01	.03	-.28	.78
Model = lmer(text_score ~ bcr_pause_duration*genre + (1|participant), data, REML = F)					
	Revision				
Text quality	Total * genre (narrative)	.19	.25	.76	.45
Model = lmer(text_score ~ log(total_revision)*genre + (1|participant), data, REML = F)					
Text quality	Below word * genre (narrative)	.23	.26	.89	.38
Model = lmer(text_score ~ log(below_word_revision)*genre + (1+genre|participant), data, REML = F)					
Text quality	Word * genre (narrative)	.17	.31	.56	.58
Model = lmer(text_score ~ log(word_revision)*genre + (1|participant), data, REML = F)					
Text quality	Below clause * genre (narrative)	-.05	.28	-.19	.85
Model = lmer(text_score ~ log(below_clause_revision)*genre + (1+genre|participant), data, REML = F)					
Text quality	Clause-and-above * genre (narrative)	.54	.46	1.17	.25
Model = lmer(text_score ~ log(clause_and_above_revision)*genre + (1+genre|participant), data, REML = F)					
Text quality	Pre-contextual * genre (narrative)	.17	.24	.71	.48
Model = lmer(text_score ~ log(precontextual_revision)*genre + (1|participant), data, REML = F)					
Text quality	Contextual * genre (narrative)	.16	.26	.61	.55
Model = lmer(text_score ~ log(contextual_revision)*genre + (1+genre|participant), data, REML = F)					
Text quality	Pinyin * genre (narrative)	.14	.23	-.58	.56
Model = lmer(text_score ~ log(pinyin_revision)*genre + (1+genre|participant), data, REML = F)					
Text quality	Character * genre (narrative)	.15	.22	.68	.50
Model = lmer(text_score ~ log(character_revision)*genre + (1|participant), data, REML = F)					

Note: Ww = Within a word, Bpc = Between Pinyin and character(s), Bw = Between words, Bc = Between clauses, Bs = Between sentences, Bcr = Between contextual revisions

Table 16 Non-Significant results from linear mixed-effects regressions examining the effects of interaction between proficiency and writing behaviours on L2 Chinese text quality (N = 32)

Dependent variable	Predictor	*Est*	*SE*	*t*	*p*
	Speed fluency				
Text quality	Production rate * proficiency	< .01	< .01	.02	.98
Model = lmer(text_score ~ production_rate*cproficiency + (1|participant) + (1|prompt), data, REML)					
Text quality	P-burst length * proficiency	-.01	.01	-1.63	.10
Model = lmer(text_score ~ pburst_length*cproficiency + (1|participant) + (1|prompt), data, REML)					
	Pause frequency				
Text quality	Total pause * proficiency	-.01	.01	-.87	.39
Model = lmer(text_score ~ total_pause_frequency*cproficiency + (1|participant) + (1|prompt), data, REML)					
Text quality	Ww pause * proficiency	-.02	.06	-.32	.75
Model = lmer(text_score ~ ww_pause_frequency*cproficiency + (1|participant) + (1|prompt), data, REML)					
Text quality	Bpc pause * proficiency	-.01	.03	-.42	.67
Model = lmer(text_score ~ bpc_pause_frequency*cproficiency + (1|participant) + (1|prompt), data, REML)					
Text quality	Bw pause * proficiency	< .01	.01	.04	.97
Model = lmer(text_score ~ bw_pause_frequency*cproficiency + (1|participant) + (1|prompt), data, REML)					
Text quality	Bc pause * proficiency	-.03	.03	-.99	.33
Model = lmer(text_score ~ bc_pause_frequency*cproficiency + (1|participant) + (1|prompt), data, REML)					
Text quality	Bcr pause * proficiency	< -.01	.02	< -.01	1.00
Model = lmer(text_score ~ bcr_pause_frequency*cproficiency + (1|participant) + (1|prompt), data, REML)					
	Pause duration				
Text quality	Total pause * proficiency	.01	.01	1.39	.17
Model = lmer(text_score ~ total_pause_duration*cproficiency + (1|participant) + (1|prompt), data, REML)					
Text quality	Ww pause * proficiency	< -.01	< .01	-.68	.50
Model = lmer(text_score ~ ww_pause_duration*cproficiency + (1|participant) + (1|prompt), data, REML)					
Text quality	Bpc pause * proficiency	< -.01	.01	.23	.82
Model = lmer(text_score ~ bpc_pause_duration*cproficiency + (1|participant) + (1|prompt), data,					

REML)					
Text quality	Bc pause * proficiency	< .01	< .01	< .01	1.00
Model = lmer(text_score ~ bc_pause_duration*cproficiency + (1|participant) + (1|prompt), data, REML)					
Text quality	Bs pause * proficiency	< .01	< .01	.15	.89
Model = lmer(text_score ~ bs_pause_duration*cproficiency + (1|participant) + (1|prompt), data, REML)					
Text quality	Bcr pause * proficiency	< .01	< .01	1.44	.15
Model = lmer(text_score ~ bcr_pause_duration*cproficiency + (1|participant) + (1|prompt), data, REML)					
	Revision				
Text quality	Total * proficiency	.04	.02	1.75	.08
Model = lmer(text_score ~ log(total_revision)*cproficiency + (1|participant) + (1|prompt), data, REML)					
Text quality	Below word * proficiency	.02	.02	.99	.33
Model = lmer(text_score ~ log(below_word_revision)*cproficiency + (1|participant) + (1|prompt), data, REML)					
Text quality	Below clause * proficiency	.04	.04	1.07	.29
Model = lmer(text_score ~ log(below_clause_revision)*cproficiency + (1|participant) + (1|prompt), data, REML)					
Text quality	Clause-and-above * proficiency	< .01	.06	.02	.98
Model = lmer(text_score ~ log(clause_and_above_revision)*cproficiency + (1|participant) + (1|prompt), data, REML)					
Text quality	Pre-contextual * proficiency	.04	.02	1.51	.13
Model = lmer(text_score ~ log(precontextual_revision)*cproficiency + (1|participant) + (1|prompt), data, REML)					
Text quality	Contextual * proficiency	.03	.03	1.10	.28
Model = lmer(text_score ~ log(contextual_revision)*cproficiency + (1|participant) + (1|prompt), data, REML)					
Text quality	Pinyin * proficiency	.03	.02	1.11	.27
Model = lmer(text_score ~ log(pinyin_revision)*cproficiency + (1|participant), data, REML)					
Text quality	Character * proficiency	.05	.03	1.63	.11
Model = lmer(text_score ~ log(character_revision)*cproficiency + (1|participant) + (1|prompt), data, REML)					

Note: Ww = Within a word, Bpc = Between Pinyin and character(s), Bw = Between words, Bc = Between clauses, Bs = Between sentences, Bcr = Between contextual revisions

Table 17 Non-significant results from linear mixed-effects regressions examining the effects of interaction between writing behaviours and stage on L2 Chinese text quality (N = 32)

Dependent variable	Predictor: behaviour index * stage	*E*	*SE*	*t*	*p*
	Speed fluency				
Text quality	Production rate * 2	< .01	< .01	.83	.41
	Production rate * 3	< .01	< .01	1.19	.23
	Production rate * 4	< -.01	< .01	.33	.74
	Production rate * 5	< .01	< .01	1.66	.10
Model = lmer(text_score ~ production_rate * stage + (1|participant) + (1|prompt), data, REML = F)					
Text quality	P-burst length * 2	-.01	.02	-.44	.66
	P-burst length * 3	-.01	.02	-.35	.72
	P-burst length * 4	-.01	.02	-.61	.54
	P-burst length * 5	-.02	.02	-1.07	.29
Model = lmer(text_score ~ pburst_length * stage + (1|participant) + (1|prompt), data, REML = F)					
	Pause frequency				
Text quality	Total pause * 2	-.02	.03	-.74	.46
	Total pause * 3	-.02	.03	-.68	.59
	Total pause * 4	-.04	.03	-1.30	.19
	Total pause * 5	-.02	.03	-.80	.42
Model = lmer(text_score ~ total_pause_frequency * stage + (1|participant) + (1|prompt), data, REML = F)					
Text quality	Ww pause * 2	-.02	.02	-.11	.91
	Ww pause * 3	-.11	.02	-.60	.55
	Ww pause * 4	-.26	.02	-1.14	.26
	Ww pause * 5	-.13	.02	-.60	.55
Model = lmer(text_score ~ ww_pause_frequency * stage + (1|participant) + (1|prompt), data, REML = F)					
Text quality	Bpc pause * 2	-.03	.06	-.47	.64
	Bpc pause * 3	.05	.07	.76	.45
	Bpc pause * 4	-.05	.07	-.76	.45
	Bpc pause * 5	-.05	.08	-.66	.51
Model = lmer(text_score ~ bpc_pause_frequency * stage + (1|participant) + (1|prompt), data, REML = F)					
Text quality	Bc pause * 2	.11	.12	.90	.37
	Bc pause * 3	.02	.12	.13	.90
	Bc pause * 4	-.05	.12	-.40	.70
	Bc pause * 5	.03	.13	.26	.79
Model = lmer(text_score ~ bc_pause_frequency * stage + (1|participant) + (1|prompt), data,					

REML = F)					
Text quality	Bs pause * 2	.07	.12	.60	.55
	Bs pause * 3	-.03	.11	-.30	.76
	Bs pause * 4	.08	.11	.71	.48
	Bs pause * 5	< .01	.11	.08	.94
Model = lmer(text_score ~ bs_pause_frequency * stage + (1|participant) + (1|prompt), data, REML = F)					
Text quality	Bcr pause * 2	.27	.17	1.56	.12
	Bcr pause * 3	.25	.16	1.57	.12
	Bcr pause * 4	.20	.15	1.37	.17
	Bcr pause * 5	.24	.15	1.62	.11
Model = lmer(text_score ~ bcr_pause_frequency * stage + (1|participant) + (1|prompt), data, REML = F)					
	Pause duration				
Text quality	Total pause * 2	-.02	.02	-.77	.44
	Total pause * 3	-.02	.02	-.68	.50
	Total pause * 4	-.03	.02	-1.32	.19
	Total pause * 5	< -.01	.01	-.50	.62
Model = lmer(text_score ~ total_pause_duration * stage + (1|participant) + (1|prompt), data, REML = F)					
Text quality	Bw pause * 2	-.02	.03	-.63	.53
	Bw pause * 3	-.01	.03	-.35	.73
	Bw pause * 4	-.02	.03	-.90	.37
	Bw pause * 5	-.01	.03	-.53	.59
Model = lmer(text_score ~ bw_pause_duration * stage + (1|participant) + (1|prompt), data, REML = F)					
Text quality	Bs pause * 2	< -.01	< .01	-.71	.48
	Bs pause * 3	< .01	< .01	.70	.48
	Bs pause * 4	< -.01	< .01	-1.30	.20
	Bs pause * 5	< -.01	< .01	-.67	.50
Model = lmer(text_score ~ bs_pause_duration * stage + (1|participant) + (1|prompt), data, REML = F)					
Text quality	Bcr pause * 2	-.01	.02	-.44	.66
	Bcr pause * 3	-.02	.01	-1.37	.17
	Bcr pause * 4	-.01	.01	-.82	.41
	Bcr pause * 5	< -.01	.01	-.21	.83
Model = lmer(text_score ~ bcr_pause_duration * stage + (1|participant), data, REML = F)					
	Revision				
Text quality	Total * 2	-.03	.09	-.30	.77
	Total * 3	.02	.08	.25	.81
	Total * 4	.03	.09	.32	.75
	Total * 5	.03	.09	.30	.76
Model = lmer(text_score ~ log(total_revision) * stage + (1|participant) + (1|prompt), data, REML = F)					

Text quality	Below word * 2	.04	.08	.45	.65
	Below word * 3	-.01	.08	-.11	.91
	Below word * 4	.06	.08	.68	.50
	Below word * 5	.09	.09	1.02	.31
Model = lmer(text_score ~ log(below_word_revision) * stage + (1|participant) + (1|prompt), data, REML = F)					
Text quality	Word * 2	-.22	.12	-1.76	.08
	Word * 3	.08	.12	.64	.52
	Word * 4	-.05	.12	-.42	.68
	Word * 5	-.11	.12	-.89	.39
Model = lmer(text_score ~ log(word_revision) * stage + (1|participant) + (1|prompt), data, REML = F)					
Text quality	Below clause * 2	-.03	.14	-.19	.85
	Below clause * 3	.12	.14	.85	.40
	Below clause * 4	.11	.15	.77	.44
	Below clause * 5	.03	.14	.21	.84
Model = lmer(text_score ~ log(below_clause_revision) * stage + (1|participant) + (1|prompt), data, REML = F)					
Text quality	Clause and above * 2	.08	.03	.28	.78
	Clause and above * 3	-.02	.03	-.07	.95
	Clause and above * 4	.02	.03	.07	.95
	Clause and above * 5	-.06	.03	-.22	.83
Model = lmer(text_score ~ log(clause_and_above_revision) * stage + (1|participant) + (1|prompt), data, REML = F)					
Text quality	Pre-contextual * 2	-.01	.08	-.11	.91
	Pre-contextual * 3	.02	.08	.19	.85
	Pre-contextual * 4	.04	.08	.50	.62
	Pre-contextual * 5	.04	.08	.47	.64
Model = lmer(text_score ~ log(precontextual_revision) * stage + (1|participant) + (1|prompt), data, REML = F)					
Text quality	Contextual * 2	-.06	.15	-.41	.68
	Contextual * 3	.04	.14	.26	.80
	Contextual * 4	-.08	.14	-.57	.57
	Contextual * 5	-.13	.14	-.97	.33
Model = lmer(text_score ~ log(contextual_revision) * stage + (1|participant) + (1|prompt), data, REML = F)					
Text quality	Pinyin * 2	-.06	.08	-.67	.51
	Pinyin * 3	< .01	.08	.06	.95
	Pinyin * 4	.03	.08	.41	.68
	Pinyin * 5	.03	.08	.37	.71
Model = lmer(text_score ~ log(pinyin_revision) * stage + (1|participant) + (1|prompt), data, REML = F)					
Text quality	Character * 2	.05	.10	.51	.61
	Character * 3	.07	.10	.75	.45

	Character * 4	.07	.10	.69	.49
	Character * 5	.03	.10	.27	.79

Model = lmer(text_score ~ log(character_revision) * stage + (1|participant) + (1|prompt), data, REML = F)

Note: Ww = Within a word, Bpc = Between Pinyin and character(s), Bw = Between words, Bc = Between clauses, Bs = Between sentences, Bcr = Between contextual revisions

Table 18 Non-significant results from linear mixed-effects regressions examining the effects of interaction between writing behaviours and stage on L1 Chinese text quality (N = 32)

Dependent variable	Predictor: behaviour index * stage	*E*	*SE*	*t*	*p*
	Speed fluency				
Text quality	Production rate * 2	< .01	< .01	.20	.84
	Production rate * 3	< .01	< .01	.66	.51
	Production rate * 4	< .01	< .01	.97	.34
	Production rate * 5	< .01	< .01	.43	.67
Model = lmer(text_score ~ production_rate * stage + (1|participant) + (1|prompt), data, REML = F)					
Text quality	P-burst length * 2	< .01	.02	.03	.98
	P-burst length * 3	.02	.02	.90	.37
	P-burst length * 4	< .01	.02	.08	.94
	P-burst length * 5	.02	.02	.91	.37
Model = lmer(text_score ~ pburst_length * stage + (1|participant) + (1|prompt), data, REML = F)					
	Pause frequency				
Text quality	Total pause * 2	-.01	.04	-.26	.80
	Total pause * 3	-.05	.04	-1.20	.23
	Total pause * 4	-.03	.04	-.77	.44
	Total pause * 5	-.05	.04	-1.13	.26
Model = lmer(text_score ~ total_pause_frequency * stage + (1|participant) + (1|prompt), data, REML = F)					
Text quality	Ww pause * 2	.23	.35	.68	.50
	Ww pause * 3	-.04	.35	-.13	.90
	Ww pause * 4	.21	.38	.54	.59
	Ww pause * 5	.31	.36	.84	.40
Model = lmer(text_score ~ ww_pause_frequency * stage + (1|participant) + (1|prompt), data, REML = F)					
Text quality	Bpc pause * 2	-.28	.20	-1.39	.17
	Bpc pause * 3	-.37	.21	-1.71	.09
	Bpc pause * 4	-.35	.21	-1.64	.10
	Bpc pause * 5	-.40	.23	-1.69	.09
Model = lmer(text_score ~ bpc_pause_frequency * stage + (1|participant) + (1|prompt), data, REML = F)					
Text quality	Bw pause * 2	-.08	.06	-1.37	.17
	Bw pause * 3	-.06	.06	-1.01	.31
	Bw pause * 4	-.04	.06	-.67	.51
	Bw pause * 5	-.05	.06	-.83	.41
Model = lmer(text_score ~ bw_pause_frequency * stage + (1|participant) + (1|prompt), data,					

REML = F)					
Text quality	Bc pause * 2	.04	.10	.41	.68
	Bc pause * 3	-.03	.10	-.26	.80
	Bc pause * 4	-.07	.10	-.75	.45
	Bc pause * 5	.01	.10	.09	.93
Model = lmer(text_score ~ bc_pause_frequency * stage + (1|participant) + (1|prompt), data, REML = F)					
Text quality	Bcr pause * 2	-.14	.15	-.93	.35
	Bcr pause * 3	-.10	.15	-.63	.53
	Bcr pause * 4	-.11	.17	-.65	.52
	Bcr pause * 5	-.20	.13	-1.53	.13
Model = lmer(text_score ~ bcr_pause_frequency * stage + (1|participant) + (1|prompt), data, REML = F)					
	Pause duration				
Text quality	Ww pause * 2	-.04	.09	-.46	.65
	Ww pause * 3	-.06	.06	-1.00	.32
	Ww pause * 4	-.04	.08	-.50	.62
	Ww pause * 5	-.07	.11	-.59	.56
Model = lmer(text_score ~ ww_pause_duration * stage + (1|participant) + (1|prompt), data, REML = F)					
Text quality	Bpc pause * 2	.05	.06	.89	.38
	Bpc pause * 3	-.01	.07	-.17	.86
	Bpc pause * 4	.06	.07	.80	.43
	Bpc pause * 5	.01	.06	.16	.87
Model = lmer(text_score ~ bpc_pause_duration * stage + (1|participant) + (1|prompt), data, REML = F)					
Text quality	Bw pause * 2	-.03	.04	-.69	.49
	Bw pause * 3	< -.01	.04	-.10	.92
	Bw pause * 4	-.02	.04	-.43	.67
	Bw pause * 5	-.03	.04	-.64	.53
Model = lmer(text_score ~ bw_pause_duration * stage + (1|participant) + (1|prompt), data, REML = F)					
Text quality	Bc pause * 2	.02	.02	1.09	.28
	Bc pause * 3	< .01	.02	.24	.82
	Bc pause * 4	-.02	.02	-1.43	.15
	Bc pause * 5	< .01	.01	.33	.75
Model = lmer(text_score ~ bc_pause_duration * stage + (1|participant) + (1|prompt), data, REML = F)					
Text quality	Bs pause * 2	< .01	.01	.20	.84
	Bs pause * 3	-.01	.01	-.89	.38
	Bs pause * 4	< -.01	.01	-.69	.49
	Bs pause * 5	-.01	< .01	-1.60	.11
Model = lmer(text_score ~ bs_pause_duration * stage + (1|participant) + (1|prompt), data, REML = F)					

Text quality	Bcr pause * 2	-.02	.02	-1.02	.31
	Bcr pause * 3	< -.01	.02	-.12	.90
	Bcr pause * 4	-.01	.02	-.66	.51
	Bcr pause * 5	-.02	.02	-1.17	.24
Model = lmer(text_score ~ bcr_pause_duration * stage + (1|participant) + (1|prompt), data, REML = F)					
	Revision				
Text quality	Total * 2	-.03	.12	-.23	.82
	Total * 3	-.03	.12	-.27	.79
	Total * 4	.02	.12	.18	.86
	Total * 5	< .01	.12	.02	.98
Model = lmer(text_score ~ log(total_revision) * stage + (1|participant) + (1|prompt), data, REML = F)					
Text quality	Below word * 2	-.03	.11	-.31	.76
	Below word * 3	< .01	.11	.04	.97
	Below word * 4	-.08	.11	-.69	.49
	Below word * 5	-.03	.11	-.28	.78
Model = lmer(text_score ~ log(below_word_revision) * stage + (1|participant) + (1|prompt), data, REML = F)					
Text quality	Word * 2	< -.01	.13	< -.01	1.00
	Word * 3	-.05	.13	-.43	.67
	Word * 4	.07	.13	.55	.59
	Word * 5	.07	.14	.54	.59
Model = lmer(text_score ~ log(word_revision) * stage + (1|participant) + (1|prompt), data, REML = F)					
Text quality	Below clause * 2	-.10	.12	-.84	.40
	Below clause * 3	-.17	.12	-1.39	.17
	Below clause * 4	.04	.12	.36	.72
	Below clause * 5	-.13	.12	-1.08	.28
Model = lmer(text_score ~ log(below_clause_revision) * stage + (1|participant) + (1|prompt), data, REML = F)					
Text quality	Clause and above * 2	.03	.20	.15	.88
	Clause and above * 3	.02	.20	.12	.90
	Clause and above * 4	-.18	.19	-.93	.35
	Clause and above * 5	.10	.20	.49	.63
Model = lmer(text_score ~ log(clause_and_above_revision) * stage + (1|participant) + (1|prompt), data, REML = F)					
Text quality	Pre-contextual * 2	.01	.11	.10	.92
	Pre-contextual * 3	-.02	.11	-.16	.88
	Pre-contextual * 4	.05	.11	.45	.66
	Pre-contextual * 5	.03	.10	.35	.72
Model = lmer(text_score ~ log(precontextual_revision) * stage + (1|participant) + (1|prompt), data, REML = F)					
Text quality	Pinyin * 2	-.03	.11	-.24	.81

	Pinyin * 3	< -.01	.11	-.03	.98
	Pinyin * 4	-.06	.11	-.59	.56
	Pinyin * 5	-.05	.10	-.52	.61
Model = lmer(text_score ~ log(pinyin_revision) * stage + (1|participant) + (1|prompt), data, REML = F)					
Text quality	Character * 2	-.01	.10	-.15	.88
	Character * 3	-.06	.11	-.62	.54
	Character * 4	.05	.10	.45	.65
	Character * 5	.01	.11	.13	.90
Model = lmer(text_score ~ log(character_revision) * stage + (1|participant) + (1|prompt), data, REML = F)					

Note: Ww = Within a word, Bpc = Between Pinyin and character(s), Bw = Between words, Bc = Between clauses, Bs = Between sentences, Bcr = Between contextual revisions

Table 19 Non-significant results from linear mixed-effects regressions examining the effects of interaction between writing behaviours and genre by stage on L2 Chinese text quality (N = 32)

Dependent variable	Predictor: behaviour index * genre	*E*	*SE*	*t*	*p*
	Speed fluency				
Text quality (S1)	Production rate * narrative	< .01	.01	.27	.79
Text quality (S2)	Production rate * narrative	< -.01	.01	-.62	.54
Text quality (S3)	Production rate * narrative	-.01	.01	-.92	.36
Text quality (S4)	Production rate * narrative	< .01	< .01	.53	.60
Text quality (S5)	Production rate * narrative	< -.01	< .01	-.73	.47

S1: Model = lmer(text_score ~ production_rate_s1 * genre + (1|participant), data, REML = F)
S2: Model = lmer(text_score ~ production_rate_s2 * genre + (1|participant), data, REML = F)
S3: Model = lmer(text_score ~ production_rate_s3 * genre + (1|participant), data, REML = F)
S4: Model = lmer(text_score ~ production_rate_s4 * genre + (1|participant), data, REML = F)
S5: Model = lmer(text_score ~ production_rate_s5 * genre + (1|participant), data, REML = F)

Dependent variable	Predictor: behaviour index * genre	*E*	*SE*	*t*	*p*
Text quality (S1)	P-burst length * narrative	-.01	.03	-.26	.79
Text quality (S2)	P-burst length * narrative	-.02	.04	-.45	.66
Text quality (S3)	P-burst length * narrative	.01	.03	.20	.84
Text quality (S4)	P-burst length * narrative	-.01	.04	-.27	.79
Text quality (S5)	P-burst length * narrative	< .01	.03	.05	.96

S1: Model = lmer(text_score ~pburst_length_s1 * genre + (1|participant) + (1|prompt), data, REML = F)
S2: Model = lmer(text_score ~pburst_length_s2 * genre + (1|participant), data, REML = F)
S3: Model = lmer(text_score ~pburst_length_s3 * genre + (1|participant), data, REML = F)
S4: Model = lmer(text_score ~pburst_length_s4 * genre + (1|participant), data, REML = F)
S5: Model = lmer(text_score ~pburst_length_s5 * genre + (1|participant), data, REML = F)

Dependent variable	Predictor: behaviour index * genre	*E*	*SE*	*t*	*p*
	Pause frequency				
Text quality (S1)	Total pause * narrative	.02	.05	.31	.76
Text quality (S2)	Total pause * narrative	-.08	.06	-1.38	.17
Text quality (S3)	Total pause * narrative	< .01	.06	.05	.96
Text quality (S4)	Total pause * narrative	.04	.05	.83	.41
Text quality (S5)	Total pause * narrative	.02	.06	.41	.68

S1: Model = lmer(text_score ~ total_pause_frequency_s1 * genre + (1|participant) + (1|prompt), data, REML = F)
S2: Model = lmer(text_score ~ total_pause_frequency_s2 * genre + (1|participant), data, REML = F)
S3: Model = lmer(text_score ~ total_pause_frequency_s3 * genre + (1|participant), data, REML = F)
S4: Model = lmer(text_score ~ total_pause_frequency_s4 * genre + (1|participant), data, REML = F)

S5: Model = lmer(text_score ~ total_pause_frequency_s5 * genre + (1|participant), data, REML = F)					
Text quality (S1)	Ww pause * narrative	-.44	.41	-1.08	.28
Text quality (S2)	Ww pause * narrative	-.17	.29	-.56	.57
Text quality (S3)	Ww pause * narrative	.15	.29	.51	.61
Text quality (S4)	Ww pause * narrative	-.30	.39	-.77	.44
Text quality (S5)	Ww pause * narrative	-.50	.36	-1.41	.16
S1: Model = lmer(text_score ~ ww_pause_frequency_s1 * genre + (1|participant), data, REML = F)					
S2: Model = lmer(text_score ~ ww_pause_frequency_s2 * genre + (1|participant), data, REML = F)					
S3: Model = lmer(text_score ~ ww_pause_frequency_s3 * genre + (1|participant), data, REML = F)					
S4: Model = lmer(text_score ~ ww_pause_frequency_s4 * genre + (1|participant), data, REML = F)					
S5: Model = lmer(text_score ~ ww_pause_frequency_s5 * genre + (1|participant), data, REML = F)					
Text quality (S1)	Bpc pause * narrative	-.02	.11	-.18	.86
Text quality (S2)	Bpc pause * narrative	-.09	.10	-.96	.34
Text quality (S3)	Bpc pause * narrative	-.05	.11	-.44	.66
Text quality (S4)	Bpc pause * narrative	.03	.10	.30	.76
Text quality (S5)	Bpc pause * narrative	.13	.15	.83	.41
S1: Model = lmer(text_score ~ bpc_pause_frequency_s1 * genre + (1|participant), data, REML = F)					
S2: Model = lmer(text_score ~ bpc_pause_frequency_s2 * genre + (1|participant), data, REML = F)					
S3: Model = lmer(text_score ~ bpc_pause_frequency_s3 * genre + (1|participant), data, REML = F)					
S4: Model = lmer(text_score ~ bpc_pause_frequency_s4 * genre + (1|participant), data, REML = F)					
S5: Model = lmer(text_score ~ bpc_pause_frequency_s5 * genre + (1|participant), data, REML = F)					
Text quality (S1)	Bw pause * narrative	.03	.07	.35	.73
Text quality (S2)	Bw pause * narrative	< .01	.09	.03	.97
Text quality (S3)	Bw pause * narrative	-.01	.08	-.11	.91
Text quality (S4)	Bw pause * narrative	.07	.06	1.11	.27
Text quality (S5)	Bw pause * narrative	-.06	.08	-.78	.44
S1: Model = lmer(text_score ~ bw_pause_frequency_s1 * genre + (1|participant), data, REML = F)					
S2: Model = lmer(text_score ~ bw_pause_frequency_s2 * genre + (1|participant), data, REML = F)					
S3: Model = lmer(text_score ~ bw_pause_frequency_s3 * genre + (1|participant), data, REML = F)					
S4: Model = lmer(text_score ~ bw_pause_frequency_s4 * genre + (1|participant), data, REML =					

F)
S5: Model = lmer(text_score ~ bw_pause_frequency_s5 * genre + (1|participant), data, REML = F)

Text quality (S1)	Bc pause * narrative	-.13	.22	-.60	.55
Text quality (S2)	Bc pause * narrative	-.22	.22	-1.01	.32
Text quality (S3)	Bc pause * narrative	.13	.17	.75	.46
Text quality (S4)	Bc pause * narrative	-.12	.18	-.67	.51
Text quality (S5)	Bc pause * narrative	-.15	.25	-.60	.55

S1: Model = lmer(text_score ~ bc_pause_frequency_s1 * genre + (1|participant), data, REML = F)
S2: Model = lmer(text_score ~ bc_pause_frequency_s2 * genre + (1|participant), data, REML = F)
S3: Model = lmer(text_score ~ bc_pause_frequency_s3 * genre + (1|participant), data, REML = F)
S4: Model = lmer(text_score ~ bc_pause_frequency_s4 * genre + (1|participant), data, REML = F)
S5: Model = lmer(text_score ~ bc_pause_frequency_s5 * genre + (1|participant), data, REML = F)

Text quality (S2)	Bs pause * narrative	-.06	.21	-.28	.78
Text quality (S3)	Bs pause * narrative	.09	.18	.50	.62
Text quality (S4)	Bs pause * narrative	.17	.19	.91	.37
Text quality (S5)	Bs pause * narrative	.05	.16	.28	.78

S2: Model = lmer(text_score ~ bs_pause_frequency_s2 * genre + (1|participant) + (1|prompt), data, REML = F)
S3: Model = lmer(text_score ~ bs_pause_frequency_s3 * genre + (1|participant) + (1|prompt), data, REML = F)
S4: Model = lmer(text_score ~ bs_pause_frequency_s4 * genre + (1|participant), data, REML = F)
S5: Model = lmer(text_score ~ bs_pause_frequency_s5 * genre + (1|participant), data, REML = F)

Text quality (S1)	Bcr pause * narrative	.24	.33	.73	.47
Text quality (S2)	Bcr pause * narrative	.08	.24	.33	.72
Text quality (S3)	Bcr pause * narrative	.01	.16	.06	.96
Text quality (S4)	Bcr pause * narrative	.05	.12	.43	.67
Text quality (S5)	Bcr pause * narrative	.10	.09	1.19	.24

S1: Model = lmer(text_score ~ bcr_pause_frequency_s1 * genre + (1|participant), data, REML = F)
S2: Model = lmer(text_score ~ bcr_pause_frequency_s2 * genre + (1|participant), data, REML = F)
S3: Model = lmer(text_score ~ bcr_pause_frequency_s3 * genre + (1|participant), data, REML = F)
S4: Model = lmer(text_score ~ bcr_pause_frequency_s4 * genre + (1|participant), data, REML = F)
S5: Model = lmer(text_score ~ bcr_pause_frequency_s5 * genre + (1|participant), data, REML = F)

Pause duration

Text quality (S1)	Total pause * narrative	< .01	.03	.14	.89
Text quality (S2)	Total pause * narrative	.05	.05	.88	.38
Text quality (S3)	Total pause * narrative	-.02	.06	-.27	.78
Text quality (S4)	Total pause * narrative	< .01	.05	.04	.97

S1: Model = lmer(text_score ~ total_pause_duration_s1 * genre + (1|participant), data, REML = F)

S2: Model = lmer(text_score ~ total_pause_duration_s2 * genre + (1|participant), data, REML = F)					
S3: Model = lmer(text_score ~ total_pause_duration_s3 * genre + (1|participant), data, REML = F)					
S4: Model = lmer(text_score ~ total_pause_duration_s4 * genre + (1|participant), data, REML = F)					
Text quality (S1)	Ww pause * narrative	< -.01	.05	-.02	.98
Text quality (S2)	Ww pause * narrative	.06	.05	1.22	.23
Text quality (S3)	Ww pause * narrative	-.01	.04	-.16	.87
Text quality (S4)	Ww pause * narrative	< -.01	.01	-.47	.64
Text quality (S5)	Ww pause * narrative	.02	.02	1.26	.22
S1: Model = lmer(text_score ~ ww_pause_duration_s1 * genre + (1|participant), data, REML = F)					
S2: Model = lmer(text_score ~ ww_pause_duration_s2 * genre + (1|participant), data, REML = F)					
S3: Model = lmer(text_score ~ ww_pause_duration_s3 * genre + (1|participant) + (1|prompt), data, REML = F)					
S4: Model = lmer(text_score ~ ww_pause_duration_s4 * genre + (1|participant) + (1|prompt), data, REML = F)					
S5: Model = lmer(text_score ~ ww_pause_duration_s5 * genre + (1|participant) + (1|prompt), data, REML = F)					
Text quality (S1)	Bpc pause * narrative	-.01	.03	-.27	.79
Text quality (S2)	Bpc pause * narrative	.02	.04	.56	.58
Text quality (S3)	Bpc pause * narrative	.03	.05	.62	.54
Text quality (S4)	Bpc pause * narrative	-.08	.07	-1.08	.29
S1: Model = lmer(text_score ~ bpc_pause_duration_s1 * genre + (1|participant), data, REML = F)					
S2: Model = lmer(text_score ~ bpc_pause_duration_s2 * genre + (1|participant), data, REML = F)					
S3: Model = lmer(text_score ~ bpc_pause_duration_s3 * genre + (1|participant), data, REML = F)					
S4: Model = lmer(text_score ~ bpc_pause_duration_s4 * genre + (1|participant), data, REML = F)					
Text quality (S1)	Bw pause * narrative	-.02	.06	-.31	.76
Text quality (S2)	Bw pause * narrative	-.01	.04	-.13	.90
Text quality (S3)	Bw pause * narrative	.05	.06	.85	.40
Text quality (S4)	Bw pause * narrative	-.02	.05	-.36	.72
Text quality (S5)	Bw pause * narrative	-.05	.04	-1.29	.20
S1: Model = lmer(text_score ~ bw_pause_duration_s1 * genre + (1|participant), data, REML = F)					
S2: Model = lmer(text_score ~ bw_pause_duration_s2 * genre + (1|participant), data, REML = F)					
S3: Model = lmer(text_score ~ bw_pause_duration_s3 * genre + (1|participant), data, REML = F)					
S4: Model = lmer(text_score ~ bw_pause_duration_s4 * genre + (1|participant), data, REML = F)					
S5: Model = lmer(text_score ~ bw_pause_duration_s5 * genre + (1|participant), data, REML = F)					
Text quality (S2)	Bc pause * narrative	.01	.02	.35	.73
Text quality (S3)	Bc pause * narrative	-.02	.03	-.68	.50
Text quality (S4)	Bc pause * narrative	.04	.02	1.74	.08
Text quality (S5)	Bc pause * narrative	-.02	.02	-1.05	.30
S2: Model = lmer(text_score ~ bc_pause_duration_s2 * genre + (1|participant), data, REML = F)					
S3: Model = lmer(text_score ~ bc_pause_duration_s3 * genre + (1|participant), data, REML = F)					
S4: Model = lmer(text_score ~ bc_pause_duration_s4 * genre + (1|participant), data, REML = F)					
S5: Model = lmer(text_score ~ bc_pause_duration_s5 * genre + (1|participant), data, REML = F)					
Text quality (S1)	Bs pause * narrative	< -.01	< .01	-1.45	.15
Text quality (S2)	Bs pause * narrative	-.01	.01	-.97	.34

Text quality (S3)	Bs pause * narrative	-.01	.01	-1.04	.30
Text quality (S4)	Bs pause * narrative	< -.01	.01	-.71	.48
Text quality (S5)	Bs pause * narrative	-.01	.01	-.69	.49
S1: Model = lmer(text_score ~ bs_pause_duration_s1 * genre + (1|participant), data, REML = F)					
S2: Model = lmer(text_score ~ bs_pause_duration_s2 * genre + (1|participant) + (1|prompt), data, REML = F)					
S3: Model = lmer(text_score ~ bs_pause_duration_s3 * genre + (1|participant) + (1|prompt), data, REML = F)					
S4: Model = lmer(text_score ~ bs_pause_duration_s4 * genre + (1|participant), data, REML = F)					
S5: Model = lmer(text_score ~ bs_pause_duration_s5 * genre + (1|participant), data, REML = F)					
Text quality (S2)	Bcr pause * narrative	.03	.02	1.15	.26
Text quality (S3)	Bcr pause * narrative	-.04	.03	-1.41	.17
Text quality (S4)	Bcr pause * narrative	.01	.02	.70	.49
Text quality (S5)	Bcr pause * narrative	.01	.01	1.75	.08
S2: Model = lmer(text_score ~ bcr_pause_duration_s2 * genre + (1|participant), data, REML = F)					
S3: Model = lmer(text_score ~ bcr_pause_duration_s3 * genre + (1|participant), data, REML = F)					
S4: Model = lmer(text_score ~ bcr_pause_duration_s4 * genre + (1|participant), data, REML = F)					
S5: Model = lmer(text_score ~ bcr_pause_duration_s5 * genre + (1|participant) + (1|prompt), data, REML = F)					
	Revision				
Text quality (S1)	Total * narrative	-.10	.13	-.75	.46
Text quality (S2)	Total * narrative	-.19	.15	-1.21	.23
Text quality (S3)	Total * narrative	-.16	.14	-1.13	.26
Text quality (S4)	Total * narrative	-.12	.15	-.80	.42
Text quality (S5)	Total * narrative	-.19	.15	-1.26	.21
S1: Model = lmer(text_score ~ log(total_revision_s1) * genre + (1|participant), data, REML = F)					
S2: Model = lmer(text_score ~ log(total_revision_s2) * genre + (1|participant), data, REML = F)					
S3: Model = lmer(text_score ~ log(total_revision_s3) * genre + (1|participant), data, REML = F)					
S4: Model = lmer(text_score ~ log(total_revision_s4) * genre + (1|participant), data, REML = F)					
S5: Model = lmer(text_score ~ log(total_revision_s5) * genre + (1|participant), data, REML = F)					
Text quality (S1)	Below word * narrative	-.12	.14	-.87	.39
Text quality (S2)	Below word * narrative	-.15	.14	-1.06	.29
Text quality (S3)	Below word * narrative	-.15	.14	-1.16	.25
Text quality (S4)	Below word * narrative	-.21	.14	-1.54	.13
Text quality (S5)	Below word * narrative	-.24	.15	-1.58	.12
S1: Model = lmer(text_score ~ log(below_word_revision_s1) * genre + (1|participant), data, REML = F)					
S2: Model = lmer(text_score ~ log(below_word_revision_s2) * genre + (1|participant), data, REML = F)					
S3: Model = lmer(text_score ~ log(below_word_revision_s3) * genre + (1|participant), data, REML = F)					
S4: Model = lmer(text_score ~ log(below_word_revision_s4) * genre + (1|participant) + (1|prompt), data, REML = F)					
S5: Model = lmer(text_score ~ log(below_word_revision_s5) * genre + (1|participant), data,					

REML = F)

Text quality (S1)	Word * narrative	< -.01	.22	< -.01	1.00
Text quality (S2)	Word * narrative	-.07	.19	-.34	.74
Text quality (S3)	Word * narrative	-.34	.18	-1.86	.07
Text quality (S4)	Word * narrative	.13	.17	.74	.46
Text quality (S5)	Word * narrative	-.12	.20	-.61	.55

S1: Model = lmer(text_score ~ log(word_revision_s1) * genre + (1|participant), data, REML = F)
S2: Model = lmer(text_score ~ log(word_revision_s2) * genre + (1|participant), data, REML = F)
S3: Model = lmer(text_score ~ log(word_revision_s3) * genre + (1|participant), data, REML = F)
S4: Model = lmer(text_score ~ log(word_revision_s4) * genre + (1|participant), data, REML = F)
S5: Model = lmer(text_score ~ log(word_revision_s5) * genre + (1|participant), data, REML = F)

Text quality (S1)	Below clause * narrative	-.20	.23	-.85	.40
Text quality (S2)	Below clause * narrative	-.30	.22	-1.33	.19
Text quality (S3)	Below clause * narrative	-.04	.22	-.17	.87
Text quality (S4)	Below clause * narrative	-.06	.25	.23	.82
Text quality (S5)	Below clause * narrative	-.16	.24	-.67	.51

S1: Model = lmer(text_score ~ log(below_clause_revision_s1) * genre + (1|participant), data, REML = F)
S2: Model = lmer(text_score ~ log(below_clause_revision_s2) * genre + (1|participant), data, REML = F)
S3: Model = lmer(text_score ~ log(below_clause_revision_s3) * genre + (1|participant), data, REML = F)
S4: Model = lmer(text_score ~ log(below_clause_revision_s4) * genre + (1|participant), data, REML = F)
S5: Model = lmer(text_score ~ log(below_clause_revision_s5) * genre + (1|participant), data, REML = F)

Text quality (S1)	Clause and above * narrative	-.20	.55	-.37	.71
Text quality (S2)	Clause and above * narrative	.24	.45	.54	.59
Text quality (S3)	Clause and above * narrative	.36	.39	.92	.36
Text quality (S4)	Clause and above * narrative	.04	.43	.10	.92
Text quality (S5)	Clause and above * narrative	-.08	.41	-.20	.84

S1: Model = lmer(text_score ~ log(clause_and_above_revision_s1) * genre + (1|participant), data, REML = F)
S2: Model = lmer(text_score ~ log(clause_and_above_revision_s2) * genre + (1|participant), data, REML = F)
S3: Model = lmer(text_score ~ log(clause_and_above_revision_s3) * genre + (1|participant), data, REML = F)
S4: Model = lmer(text_score ~ log(clause_and_above_revision_s4) * genre + (1|participant), data, REML = F)
S5: Model = lmer(text_score ~ log(clause_and_above_revision_s5) * genre + (1|participant), data, REML = F)

Text quality (S1)	Pre-contextual * narrative	-.12	.13	-.88	.38
Text quality (S2)	Pre-contextual * narrative	-.17	.15	-1.17	.25
Text quality (S3)	Pre-contextual * narrative	-.12	.13	-.90	.37

Text quality (S4)	Pre-contextual * narrative	-.15	.14	-1.10	.27
Text quality (S5)	Pre-contextual * narrative	-.16	.12	-1.32	.19

S1: Model = lmer(text_score ~ log(precontextual_revision_s1) * genre + (1|participant), data, REML = F)
S2: Model = lmer(text_score ~ log(precontextual_revision_s2) * genre + (1|participant), data, REML = F)
S3: Model = lmer(text_score ~ log(precontextual_revision_s3) * genre + (1|participant), data, REML = F)
S4: Model = lmer(text_score ~ log(precontextual_revision_s4) * genre + (1|participant), data, REML = F)
S5: Model = lmer(text_score ~ log(precontextual_revision_s5) * genre + (1|participant), data, REML = F)

Text quality (S1)	Contextual * narrative	.39	.29	1.36	.18
Text quality (S2)	Contextual * narrative	-.10	.23	-.44	.66
Text quality (S3)	Contextual * narrative	-.13	.19	-.73	.47
Text quality (S4)	Contextual * narrative	.07	.18	.42	.68
Text quality (S5)	Contextual * narrative	.09	.16	.56	.58

S1: Model = lmer(text_score ~ log(contextual_revision_s1) * genre + (1|participant) + (1|prompt), data, REML = F)
S2: Model = lmer(text_score ~ log(contextual_revision_s2) * genre + (1|participant), data, REML = F)
S3: Model = lmer(text_score ~ log(contextual_revision_s3) * genre + (1|participant), data, REML = F)
S4: Model = lmer(text_score ~ log(contextual_revision_s4) * genre + (1|participant) + (1|prompt), data, REML = F)
S5: Model = lmer(text_score ~ log(contextual_revision_s5) * genre + (1|participant) + (1|prompt), data, REML = F)

Text quality (S1)	Pinyin * narrative	-.09	.14	-.64	.52
Text quality (S2)	Pinyin * narrative	-.14	.15	-.96	.34
Text quality (S3)	Pinyin * narrative	-.13	.14	-.95	.34
Text quality (S4)	Pinyin * narrative	-.23	.14	-1.64	.10
Text quality (S5)	Pinyin * narrative	-.22	.14	-1.52	.13

S1: Model = lmer(text_score ~ log(pinyin_revision_s1) * genre + (1|participant), data, REML = F)
S2: Model = lmer(text_score ~ log(pinyin_revision_s2) * genre + (1|participant), data, REML = F)
S3: Model = lmer(text_score ~ log(pinyin_revision_s3) * genre + (1|participant), data, REML = F)
S4: Model = lmer(text_score ~ log(pinyin_revision_s4) * genre + (1|participant), data, REML = F)
S5: Model = lmer(text_score ~ log(pinyin_revision_s5) * genre + (1|participant), data, REML = F)

Text quality (S1)	Character * narrative	-.05	.18	-.30	.76
Text quality (S2)	Character * narrative	-.10	.16	-.63	.53
Text quality (S3)	Character * narrative	-.11	.15	-.77	.44
Text quality (S4)	Character * narrative	.05	.15	.35	.73
Text quality (S5)	Character * narrative	-.12	.16	.76	.45

S1: Model = lmer(text_score ~ log(character_revision_s1) * genre + (1|participant) + (1|prompt), data, REML = F)

S2: Model = lmer(text_score ~ log(character_revision_s2) * genre + (1|participant), data, REML = F)

S3: Model = lmer(text_score ~ log(character_revision_s3) * genre + (1|participant), data, REML = F)

S4: Model = lmer(text_score ~ log(character_revision_s4) * genre + (1|participant), data, REML = F)

S5: Model = lmer(text_score ~ log(character_revision_s5) * genre + (1|participant), data, REML = F)

Note: Ww = Within a word, Bpc = Between Pinyin and character(s), Bw = Between words, Bc = Between clauses, Bs = Between sentences, Bcr = Between contextual revision

Table 20 Non-significant results from linear mixed-effects regressions examining the effects of interaction between writing behaviours and genre by stage on L1 Chinese text quality (N = 32)

Dependent variable	Predictor: behaviour index * genre	*E*	*SE*	*t*	*p*
	Speed fluency				
Text quality (S1)	Production rate * narrative	< -.01	< .01	-1.18	.24
Text quality (S2)	Production rate * narrative	< -.01	< .01	-.60	.55
Text quality (S3)	Production rate * narrative	< -.01	< .01	-1.08	.28
Text quality (S4)	Production rate * narrative	.01	< .01	1.27	.21
Text quality (S5)	Production rate * narrative	-.01	.01	-1.12	.27
S1: Model = lmer(text_score ~ production_rate_s1 * genre + (1|participant), data, REML = F)					
S2: Model = lmer(text_score ~ production_rate_s2 * genre + (1|participant), data, REML = F)					
S3: Model = lmer(text_score ~ production_rate_s3 * genre + (1|participant), data, REML = F)					
S4: Model = lmer(text_score ~ production_rate_s4 * genre + (1|participant), data, REML = F)					
S5: Model = lmer(text_score ~ production_rate_s5 * genre + (1+genre|participant), data, REML = F)					
Text quality (S1)	P-burst length * narrative	-.03	.03	-.87	.38
Text quality (S2)	P-burst length * narrative	.03	.03	1.04	.30
Text quality (S3)	P-burst length * narrative	-.03	.03	-.99	.33
Text quality (S4)	P-burst length * narrative	< .01	.03	.16	.87
Text quality (S5)	P-burst length * narrative	-.03	.03	-1.19	.24
S1: Model = lmer(text_score ~pburst_length_s1 * genre + (1|participant), data, REML = F)					
S2: Model = lmer(text_score ~pburst_length_s2 * genre + (1|participant), data, REML = F)					
S3: Model = lmer(text_score ~pburst_length_s3 * genre + (1|participant), data, REML = F)					
S4: Model = lmer(text_score ~pburst_length_s4 * genre + (1|participant), data, REML = F)					
S5: Model = lmer(text_score ~pburst_length_s5 * genre + (1+genre|participant), data, REML = F)					
	Pause frequency				
Text quality (S1)	Total pause * narrative	.05	.07	.74	.46
Text quality (S2)	Total pause * narrative	.11	.07	1.46	.15
Text quality (S3)	Total pause * narrative	-.06	.08	-.74	.46
Text quality (S4)	Total pause * narrative	< -.01	.07	-.07	.94
Text quality (S5)	Total pause * narrative	-.08	.08	-1.03	.31
S1: Model = lmer(text_score ~ total_pause_frequency_s1 * genre + (1|participant), data, REML = F)					
S2: Model = lmer(text_score ~ total_pause_frequency_s2 * genre + (1+genre|participant), data, REML = F)					
S3: Model = lmer(text_score ~ total_pause_frequency_s3 * genre + (1|participant), data, REML = F)					
S4: Model = lmer(text_score ~ total_pause_frequency_s4 * genre + (1+genre|participant), data, REML = F)					

S5: Model = lmer(text_score ~ total_pause_frequency_s5 * genre + (1+genre|participant), data, REML = F)

Text quality (S1)	Ww pause * narrative	.15	.63	.24	.81
Text quality (S2)	Ww pause * narrative	.83	.59	1.42	.16
Text quality (S3)	Ww pause * narrative	.89	.61	1.44	.15
Text quality (S4)	Ww pause * narrative	.56	.66	.85	.40
Text quality (S5)	Ww pause * narrative	.16	.68	.23	.82

S1: Model = lmer(text_score ~ ww_pause_frequency_s1 * genre + (1|participant), data, REML = F)

S2: Model = lmer(text_score ~ ww_pause_frequency_s2 * genre + (1+genre|participant), data, REML = F)

S3: Model = lmer(text_score ~ ww_pause_frequency_s3 * genre + (1|participant), data, REML = F)

S4: Model = lmer(text_score ~ ww_pause_frequency_s4 * genre + (1|participant), data, REML = F)

S5: Model = lmer(text_score ~ ww_pause_frequency_s5 * genre + (1+genre|participant), data, REML = F)

Text quality (S2)	Bpc pause * narrative	.29	.29	1.03	.31
Text quality (S3)	Bpc pause * narrative	.22	.33	.66	.51
Text quality (S4)	Bpc pause * narrative	-.08	.32	.24	.81
Text quality (S5)	Bpc pause * narrative	-.05	.41	.11	.91

S2: Model = lmer(text_score ~ bpc_pause_frequency_s2 * genre + (1+genre|participant), data, REML = F)

S3: Model = lmer(text_score ~ bpc_pause_frequency_s3 * genre + (1|participant), data, REML = F)

S4: Model = lmer(text_score ~ bpc_pause_frequency_s4 * genre + (1|participant), data, REML = F)

S5: Model = lmer(text_score ~ bpc_pause_frequency_s5 * genre + (1+genre|participant), data, REML = F)

Text quality (S1)	Bw pause * narrative	.03	.12	.22	.83
Text quality (S2)	Bw pause * narrative	-.08	.10	-.77	.45
Text quality (S3)	Bw pause * narrative	-.07	.11	-.66	.51
Text quality (S4)	Bw pause * narrative	-.08	.09	-.83	.41
Text quality (S5)	Bw pause * narrative	-.02	.10	-.22	.83

S1: Model = lmer(text_score ~ bw_pause_frequency_s1 * genre + (1|participant), data, REML = F)

S2: Model = lmer(text_score ~ bw_pause_frequency_s2 * genre + (1+genre|participant), data, REML = F)

S3: Model = lmer(text_score ~ bw_pause_frequency_s3 * genre + (1+genre|participant), data, REML = F)

S4: Model = lmer(text_score ~ bw_pause_frequency_s4 * genre + (1|participant), data, REML = F)

S5: Model = lmer(text_score ~ bw_pause_frequency_s5 * genre + (1+genre|participant), data, REML = F)

Text quality (S1)	Bc pause * narrative	.13	.20	.64	.53
Text quality (S2)	Bc pause * narrative	.20	.16	1.24	.22
Text quality (S3)	Bc pause * narrative	-.16	.15	-1.05	.30
Text quality (S4)	Bc pause * narrative	-.13	.14	-.89	.37

S1: Model = lmer(text_score ~ bc_pause_frequency_s1 * genre + (1+genre|participant), data, REML = F)

S2: Model = lmer(text_score ~ bc_pause_frequency_s2 * genre + (1+genre|participant), data, REML = F)

S3: Model = lmer(text_score ~ bc_pause_frequency_s3 * genre + (1|participant), data, REML = F)

S4: Model = lmer(text_score ~ bc_pause_frequency_s4 * genre + (1+genre|participant), data, REML = F)

Text quality (S1)	Bs pause * narrative	-.19	.17	-1.14	.26
Text quality (S3)	Bs pause * narrative	-.11	.20	-.54	.59
Text quality (S4)	Bs pause * narrative	.06	.18	.36	.72
Text quality (S5)	Bs pause * narrative	-.08	.19	-.42	.67

S1: Model = lmer(text_score ~ bs_pause_frequency_s1 * genre + (1|participant), data, REML = F)

S3: Model = lmer(text_score ~ bs_pause_frequency_s3 * genre + (1|participant), data, REML = F)

S4: Model = lmer(text_score ~ bs_pause_frequency_s4 * genre + (1|participant), data, REML = F)

S5: Model = lmer(text_score ~ bs_pause_frequency_s5 * genre + (1+genre|participant), data, REML = F)

Text quality (S2)	Bcr pause * narrative	.20	.24	.85	.40
Text quality (S3)	Bcr pause * narrative	.34	.24	1.39	.17
Text quality (S5)	Bcr pause * narrative	.14	.13	1.14	.26

S2: Model = lmer(text_score ~ bcr_pause_frequency_s2 * genre + (1+genre|participant), data, REML = F)

S3: Model = lmer(text_score ~ bcr_pause_frequency_s3 * genre + (1|participant), data, REML = F)

S5: Model = lmer(text_score ~ bcr_pause_frequency_s5 * genre + (1|participant), data, REML = F)

Pause duration

Text quality (S1)	Total pause * narrative	-.03	.04	-.77	.44
Text quality (S2)	Total pause * narrative	-.07	.06	-1.15	.25
Text quality (S3)	Total pause * narrative	.02	.09	.26	.79
Text quality (S4)	Total pause * narrative	.05	.08	.67	.50
Text quality (S5)	Total pause * narrative	< -.01	.05	-.03	.98

S1: Model = lmer(text_score ~ total_pause_duration_s1 * genre + (1+genre|participant), data, REML = F)

S2: Model = lmer(text_score ~ total_pause_duration_s2 * genre + (1|participant), data, REML = F)

S3: Model = lmer(text_score ~ total_pause_duration_s3 * genre + (1|participant), data, REML = F)

S4: Model = lmer(text_score ~ total_pause_duration_s4 * genre + (1|participant), data, REML = F)

S5: Model = lmer(text_score ~ total_pause_duration_s5 * genre + (1|participant), data, REML = F)

Text quality (S1)	Ww pause * narrative	.02	.10	.19	.85
Text quality (S2)	Ww pause * narrative	.02	.25	.08	.94
Text quality (S3)	Ww pause * narrative	.03	.06	.53	.60

Text quality (S4)	Ww pause * narrative	.48	.20	2.36	.03[7]
Text quality (S5)	Ww pause * narrative	.20	.23	.87	.39

S1: Model = lmer(text_score ~ ww_pause_duration_s1 * genre + (1|participant), data, REML = F)
S2: Model = lmer(text_score ~ ww_pause_duration_s2 * genre + (1|participant), data, REML = F)
S3: Model = lmer(text_score ~ ww_pause_duration_s3 * genre + (1|participant), data, REML = F)
S4: Model = lmer(text_score ~ ww_pause_duration_s4 * genre + (1|participant), data, REML = F)
S5: Model = lmer(text_score ~ ww_pause_duration_s5 * genre + (1|participant), data, REML = F)

Text quality (S1)	Bpc pause * narrative	-.08	.11	-.70	.49
Text quality (S2)	Bpc pause * narrative	.11	.11	1.01	.32
Text quality (S3)	Bpc pause * narrative	.04	.15	.28	.78
Text quality (S4)	Bpc pause * narrative	-.04	.19	-.22	.83
Text quality (S5)	Bpc pause * narrative	-.09	.06	-1.42	.16

S1: Model = lmer(text_score ~ bpc_pause_duration_s1 * genre + (1|participant), data, REML = F)
S2: Model = lmer(text_score ~ bpc_pause_duration_s2 * genre + (1|participant), data, REML = F)
S3: Model = lmer(text_score ~ bpc_pause_duration_s3 * genre + (1|participant), data, REML = F)
S4: Model = lmer(text_score ~ bpc_pause_duration_s4 * genre + (1+genre|participant), data, REML = F)
S5: Model = lmer(text_score ~ bpc_pause_duration_s5 * genre + (1|participant) + (1|prompt), data, REML = F)

Text quality (S1)	Bw pause * narrative	-.06	.07	-.91	.37
Text quality (S2)	Bw pause * narrative	-.14	.07	-1.95	.05
Text quality (S3)	Bw pause * narrative	.07	.08	.86	.39
Text quality (S4)	Bw pause * narrative	-.01	.07	-.08	.94
Text quality (S5)	Bw pause * narrative	.09	.07	1.20	.23

S1: Model = lmer(text_score ~ bw_pause_duration_s1 * genre + (1+genre|participant), data, REML = F)
S2: Model = lmer(text_score ~ bw_pause_duration_s2 * genre + (1+genre|participant), data, REML = F)
S3: Model = lmer(text_score ~ bw_pause_duration_s3 * genre + (1|participant), data, REML = F)
S4: Model = lmer(text_score ~ bw_pause_duration_s4 * genre + (1+genre|participant), data, REML = F)
S5: Model = lmer(text_score ~ bw_pause_duration_s5 * genre + (1|participant), data, REML = F)

Text quality (S1)	Bc pause * narrative	.05	.03	1.91	.06
Text quality (S2)	Bc pause * narrative	.06	.04	1.63	.11
Text quality (S3)	Bc pause * narrative	.07	.08	.86	.39
Text quality (S4)	Bc pause * narrative	.03	.03	.91	.36
Text quality (S5)	Bc pause * narrative	-.02	.03	-.69	.49

S1: Model = lmer(text_score ~ bc_pause_duration_s1 * genre + (1+genre|participant), data, REML = F)
S2: Model = lmer(text_score ~ bc_pause_duration_s2 * genre + (1+genre|participant), data, REML = F)
S3: Model = lmer(text_score ~ bc_pause_duration_s3 * genre + (1|participant), data, REML = F)
S4: Model = lmer(text_score ~ bc_pause_duration_s4 * genre + (1|participant), data, REML = F)

[7] The model is singular fit, and 84% of the data for within-word pause duration in Stage 4 contains missing value.

S5: Model = lmer(text_score ~ bc_pause_duration_s5 * genre + (1|participant), data, REML = F)

Text quality (S1)	Bs pause * narrative	< .01	< .01	-.70	.49
Text quality (S2)	Bs pause * narrative	-.01	.01	-.98	.33
Text quality (S3)	Bs pause * narrative	.01	.02	.48	.63
Text quality (S4)	Bs pause * narrative	.01	.02	.62	.54
Text quality (S5)	Bs pause * narrative	< -.01	.01	-.38	.70

S1: Model = lmer(text_score ~ bs_pause_duration_s1 * genre + (1+genre|participant), data, REML = F)

S2: Model = lmer(text_score ~ bs_pause_duration_s2 * genre + (1+genre|participant) + (1|prompt), data, REML = F)

S3: Model = lmer(text_score ~ bs_pause_duration_s3 * genre + (1+genre|participant), data, REML = F)

S4: Model = lmer(text_score ~ bs_pause_duration_s4 * genre + (1+genre|participant), data, REML = F)

S5: Model = lmer(text_score ~ bs_pause_duration_s5 * genre + (1+genre|participant), data, REML = F)

Text quality (S1)	Bcr pause * narrative	-.02	.05	-.47	.64
Text quality (S2)	Bcr pause * narrative	.03	.05	.74	.47
Text quality (S3)	Bcr pause * narrative	-.04	.04	-1.04	.31
Text quality (S4)	Bcr pause * narrative	-.03	.03	-1.06	.29
Text quality (S5)	Bcr pause * narrative	< -.01	.01	-.28	.78

S1: Model = lmer(text_score ~ bcr_pause_duration_s1 * genre + (1|participant), data, REML = F)

S2: Model = lmer(text_score ~ bcr_pause_duration_s2 * genre + (1|participant), data, REML = F)

S3: Model = lmer(text_score ~ bcr_pause_duration_s3 * genre + (1|participant), data, REML = F)

S4: Model = lmer(text_score ~ bcr_pause_duration_s4 * genre + (1|participant), data, REML = F)

S5: Model = lmer(text_score ~ bcr_pause_duration_s5 * genre + (1+genre|participant), data, REML = F)

Revision

Text quality (S1)	Total * narrative	.30	.17	1.82	.07
Text quality (S2)	Total * narrative	.20	.22	.90	.37
Text quality (S3)	Total * narrative	-.04	.23	-.16	.88
Text quality (S4)	Total * narrative	.18	.22	.83	.41
Text quality (S5)	Total * narrative	-.12	.22	-.56	.58

S1: Model = lmer(text_score ~ log(total_revision_s1) * genre + (1|participant), data, REML = F)

S2: Model = lmer(text_score ~ log(total_revision_s2) * genre + (1+genre|participant), data, REML = F)

S3: Model = lmer(text_score ~ log(total_revision_s3) * genre + (1+genre|participant), data, REML = F)

S4: Model = lmer(text_score ~ log(total_revision_s4) * genre + (1|participant), data, REML = F)

S5: Model = lmer(text_score ~ log(total_revision_s5) * genre + (1+genre|participant), data, REML = F)

Text quality (S1)	Below word * narrative	.33	.17	1.90	.06
Text quality (S2)	Below word * narrative	.13	.21	.65	.52
Text quality (S3)	Below word * narrative	.09	.21	.43	.67

Text quality (S4)	Below word * narrative	.19	.20	.93	.35
Text quality (S5)	Below word * narrative	-.11	.19	.56	.58
S1: Model = lmer(text_score ~ log(below_word_revision_s1) * genre + (1+genre|participant), data, REML = F)					
S2: Model = lmer(text_score ~ log(below_word_revision_s2) * genre + (1+genre|participant), data, REML = F)					
S3: Model = lmer(text_score ~ log(below_word_revision_s3) * genre + (1|participant), data, REML = F)					
S4: Model = lmer(text_score ~ log(below_word_revision_s4) * genre + (1+genre|participant) + (1|prompt), data, REML = F)					
S5: Model = lmer(text_score ~ log(below_word_revision_s5) * genre + (1+genre|participant), data, REML = F)					
Text quality (S1)	Word * narrative	.30	.21	1.46	.15
Text quality (S2)	Word * narrative	.26	.22	1.18	.24
Text quality (S3)	Word * narrative	-.15	.22	-.67	.51
Text quality (S4)	Word * narrative	.11	.23	.48	.63
Text quality (S5)	Word * narrative	.01	.27	.03	.97
S1: Model = lmer(text_score ~ log(word_revision_s1) * genre + (1|participant), data, REML = F)					
S2: Model = lmer(text_score ~ log(word_revision_s2) * genre + (1|participant), data, REML = F)					
S3: Model = lmer(text_score ~ log(word_revision_s3) * genre + (1+genre|participant), data, REML = F)					
S4: Model = lmer(text_score ~ log(word_revision_s4) * genre + (1|participant), data, REML = F)					
S5: Model = lmer(text_score ~ log(word_revision_s5) * genre + (1|participant), data, REML = F)					
Text quality (S1)	Below clause * narrative	.31	.20	1.55	.13
Text quality (S2)	Below clause * narrative	-.20	.22	-.90	.37
Text quality (S3)	Below clause * narrative	-.07	.22	-.35	.73
Text quality (S4)	Below clause * narrative	.13	.22	.58	.56
Text quality (S5)	Below clause * narrative	-.09	.21	-.43	.67
S1: Model = lmer(text_score ~ log(below_clause_revision_s1) * genre + (1+genre|participant), data, REML = F)					
S2: Model = lmer(text_score ~ log(below_clause_revision_s2) * genre + (1+genre|participant), data, REML = F)					
S3: Model = lmer(text_score ~ log(below_clause_revision_s3) * genre + (1|participant), data, REML = F)					
S4: Model = lmer(text_score ~ log(below_clause_revision_s4) * genre + (1|participant), data, REML = F)					
S5: Model = lmer(text_score ~ log(below_clause_revision_s5) * genre + (1|participant), data, REML = F)					
Text quality (S1)	Clause and above * narrative	.39	.34	1.16	.25
Text quality (S2)	Clause and above * narrative	.60	.33	1.81	.07
Text quality (S3)	Clause and above * narrative	.09	.32	.29	.77
Text quality (S4)	Clause and above * narrative	.43	.34	1.28	.20
Text quality (S5)	Clause and above * narrative	.01	.35	.03	.98
S1: Model = lmer(text_score ~ log(clause_and_above_revision_s1) * genre + (1|participant), data,					

REML = F)					
S2: Model = lmer(text_score ~ log(clause_and_above_revision_s2) * genre + (1|participant), data, REML = F)					
S3: Model = lmer(text_score ~ log(clause_and_above_revision_s3) * genre + (1+genre|participant), data, REML = F)					
S4: Model = lmer(text_score ~ log(clause_and_above_revision_s4) * genre + (1+genre|participant), data, REML = F)					
S5: Model = lmer(text_score ~ log(clause_and_above_revision_s5) * genre + (1|participant), data, REML = F)					
Text quality (S1)	Pre-contextual * narrative	.26	.16	1.65	.10
Text quality (S2)	Pre-contextual * narrative	.15	.22	.69	.49
Text quality (S3)	Pre-contextual * narrative	-.04	.21	-.19	.85
Text quality (S4)	Pre-contextual * narrative	.18	.21	.85	.40
Text quality (S5)	Pre-contextual * narrative	-.10	.17	-.55	.58
S1: Model = lmer(text_score ~ log(precontextual_revision_s1) * genre + (1+genre|participant), data, REML = F)					
S2: Model = lmer(text_score ~ log(precontextual_revision_s2) * genre + (1+genre|participant), data, REML = F)					
S3: Model = lmer(text_score ~ log(precontextual_revision_s3) * genre + (1+genre|participant), data, REML = F)					
S4: Model = lmer(text_score ~ log(precontextual_revision_s4) * genre + (1|participant), data, REML = F)					
S5: Model = lmer(text_score ~ log(precontextual_revision_s5) * genre + (1|participant), data, REML = F)					
Text quality (S1)	Contextual * narrative	.27	.22	1.22	.22
Text quality (S2)	Contextual * narrative	.25	.20	1.32	.19
Text quality (S3)	Contextual * narrative	.10	.19	.53	.60
Text quality (S4)	Contextual * narrative	.01	.23	.04	.97
Text quality (S5)	Contextual * narrative	-.02	.19	-.12	.93
S1: Model = lmer(text_score ~ log(contextual_revision_s1) * genre + (1|participant) + (1|prompt), data, REML = F)					
S2: Model = lmer(text_score ~ log(contextual_revision_s2) * genre + (1+genre|participant), data, REML = F)					
S3: Model = lmer(text_score ~ log(contextual_revision_s3) * genre + (1+genre|participant), data, REML = F)					
S4: Model = lmer(text_score ~ log(contextual_revision_s4) * genre + (1+genre|participant), data, REML = F)					
S5: Model = lmer(text_score ~ log(contextual_revision_s5) * genre + (1+genre|participant) + (1|prompt), data, REML = F)					
Text quality (S1)	Pinyin * narrative	.29	.17	1.67	.10
Text quality (S2)	Pinyin * narrative	.11	.21	.53	.60
Text quality (S3)	Pinyin * narrative	-.01	.19	-.05	.96
Text quality (S4)	Pinyin * narrative	.16	.20	.84	.41
Text quality (S5)	Pinyin * narrative	-.12	.18	-.71	.48

S1: Model = lmer(text_score ~ log(pinyin_revision_s1) * genre + (1|participant), data, REML = F)
S2: Model = lmer(text_score ~ log(pinyin_revision_s2) * genre + (1+genre|participant), data, REML = F)
S3: Model = lmer(text_score ~ log(pinyin_revision_s3) * genre + (1+genre|participant), data, REML = F)
S4: Model = lmer(text_score ~ log(pinyin_revision_s4) * genre + (1+genre|participant), data, REML = F)
S5: Model = lmer(text_score ~ log(pinyin_revision_s5) * genre + (1+genre|participant), data, REML = F)

Text quality (S2)	Character * narrative	.18	.18	.97	.34
Text quality (S3)	Character * narrative	-.06	.20	-.33	.75
Text quality (S4)	Character * narrative	.11	.19	.61	.55
Text quality (S5)	Character * narrative	-.05	.20	-.26	.80

S2: Model = lmer(text_score ~ log(character_revision_s2) * genre + (1|participant), data, REML = F)
S3: Model = lmer(text_score ~ log(character_revision_s3) * genre + (1+genre|participant), data, REML = F)
S4: Model = lmer(text_score ~ log(character_revision_s4) * genre + (1|participant), data, REML = F)
S5: Model = lmer(text_score ~ log(character_revision_s5) * genre + (1|participant), data, REML = F)

Note: Ww = Within a word, Bpc = Between Pinyin and character(s), Bw = Between words, Bc = Between clauses, Bs = Between sentences, Bcr = Between contextual revision

Table 21 Non-significant results from linear mixed-effects regressions examining the effects of interaction between writing behaviours and L2 proficiency by stage on text quality in L2 Chinese (N = 32)

Dependent variable	Predictor	*E*	*SE*	*t*	*p*
	Speed fluency				
Text quality (S1)	Production rate * proficiency	< .01	< .01	.24	.81
Text quality (S2)	Production rate * proficiency	< -.01	< .01	-.20	.85
Text quality (S3)	Production rate * proficiency	< -.01	< .01	-.70	.49
Text quality (S4)	Production rate * proficiency	< .01	< .01	.53	.60
Text quality (S5)	Production rate * proficiency	< -.01	< .01	-1.04	.30
S1: Model = lmer(text_score ~ production_rate_s1 * cproficiency + (1|participant), data, REML = F)					
S2: Model = lmer(text_score ~ production_rate_s2 * cproficiency + (1|participant) + (1|prompt), data, REML = F)					
S3: Model = lmer(text_score ~ production_rate_s3 * cproficiency + (1|participant) + (1|prompt), data, REML = F)					
S4: Model = lmer(text_score ~ production_rate_s4 * cproficiency + (1|participant) + (1|prompt), data, REML = F)					
S5: Model = lmer(text_score ~ production_rate_s5 * cproficiency + (1|participant) + (1|prompt), data, REML = F)					
Text quality (S1)	P-burst length * proficiency	< -.01	< .01	-.30	.77
Text quality (S2)	P-burst length * proficiency	< -.01	< .01	-.33	.74
Text quality (S3)	P-burst length * proficiency	< -.01	< .01	-.83	.41
Text quality (S4)	P-burst length * proficiency	-.01	< .01	-1.34	.18
Text quality (S5)	P-burst length * proficiency	< -.01	< .01	-.21	.84
S1: Model = lmer(text_score ~ pburst_length_s1 * cproficiency + (1|participant) + (1|prompt), data, REML = F)					
S2: Model = lmer(text_score ~ pburst_length_s2 * cproficiency + (1|participant) + (1|prompt), data, REML = F)					
S3: Model = lmer(text_score ~ pburst_length_s3 * cproficiency + (1|participant) + (1|prompt), data, REML = F)					
S4: Model = lmer(text_score ~ pburst_length_s4 * cproficiency + (1|participant), data, REML = F)					
S5: Model = lmer(text_score ~ pburst_length_s5 * cproficiency + (1|participant) + (1|prompt), data, REML = F)					
	Pause frequency				
Text quality (S1)	Total pause * proficiency	< .01	< .01	.52	.60
Text quality (S2)	Total pause * proficiency	< .01	.01	.48	.63
Text quality (S3)	Total pause * proficiency	< -.01	.01	-.37	.71
Text quality (S4)	Total pause * proficiency	< -.01	< .01	-.35	.73
S1: Model = lmer(text_score ~ total_pause_frequency_s1 * cproficiency + (1|participant) + (1|prompt), data, REML = F)					

S2: Model = lmer(text_score ~ total_pause_frequency_s2 * cproficiency + (1|participant), data, REML = F)

S3: Model = lmer(text_score ~ total_pause_frequency_s3 * cproficiency + (1|participant), data, REML = F)

S4: Model = lmer(text_score ~ total_pause_frequency_s4 * cproficiency + (1|participant), data, REML = F)

Text quality (S1)	Ww pause * proficiency	.02	.03	.75	.46
Text quality (S2)	Ww pause * proficiency	-.01	.02	-.32	.75
Text quality (S3)	Ww pause * proficiency	-.02	.02	-.97	.33
Text quality (S4)	Ww pause * proficiency	.02	.04	.64	.52
Text quality (S5)	Ww pause * proficiency	< -.01	.02	-.08	.94

S1: Model = lmer(text_score ~ ww_pause_frequency_s1 * cproficiency + (1|participant) + (1|prompt), data, REML = F)

S2: Model = lmer(text_score ~ ww_pause_frequency_s2 * cproficiency + (1|participant) + (1|prompt), data, REML = F)

S3: Model = lmer(text_score ~ ww_pause_frequency_s3 * cproficiency + (1|participant) + (1|prompt), data, REML = F)

S4: Model = lmer(text_score ~ ww_pause_frequency_s4 * cproficiency + (1|participant) + (1|prompt), data, REML = F)

S5: Model = lmer(text_score ~ ww_pause_frequency_s5 * cproficiency + (1|participant) + (1|prompt), data, REML = F)

Text quality (S1)	Bpc pause * proficiency	.01	.01	.85	.40
Text quality (S2)	Bpc pause * proficiency	-.02	.01	-1.49	.14
Text quality (S3)	Bpc pause * proficiency	.02	.01	1.44	.15
Text quality (S4)	Bpc pause * proficiency	.01	.02	.65	.52
Text quality (S5)	Bpc pause * proficiency	-.03	.02	-1.82	.07

S1: Model = lmer(text_score ~ bpc_pause_frequency_s1 * cproficiency + (1|participant) + (1|prompt), data, REML = F)

S2: Model = lmer(text_score ~ bpc_pause_frequency_s2 * cproficiency + (1|participant) + (1|prompt), data, REML = F)

S3: Model = lmer(text_score ~ bpc_pause_frequency_s3 * cproficiency + (1|participant) + (1|prompt), data, REML = F)

S4: Model = lmer(text_score ~ bpc_pause_frequency_s4 * cproficiency + (1|participant) + (1|prompt), data, REML = F)

S5: Model = lmer(text_score ~ bpc_pause_frequency_s5 * cproficiency + (1|participant) + (1|prompt), data, REML = F)

Text quality (S1)	Bw pause * proficiency	.01	.01	.90	.37
Text quality (S2)	Bw pause * proficiency	.01	.01	1.27	.21
Text quality (S3)	Bw pause * proficiency	< .01	.01	.34	.74
Text quality (S4)	Bw pause * proficiency	< -.01	.01	-.04	.97
Text quality (S5)	Bw pause * proficiency	< -.01	.01	-.72	.47

S1: Model = lmer(text_score ~ bw_pause_frequency_s1 * cproficiency + (1|participant) + (1|prompt), data, REML = F)

S2: Model = lmer(text_score ~ bw_pause_frequency_s2 * cproficiency + (1|participant) +

(1|prompt), data, REML = F)

S3: Model = lmer(text_score ~ bw_pause_frequency_s3 * cproficiency + (1|participant) + (1|prompt), data, REML = F)

S4: Model = lmer(text_score ~ bw_pause_frequency_s4 * cproficiency + (1|participant) + (1|prompt), data, REML = F)

S5: Model = lmer(text_score ~ bw_pause_frequency_s5 * cproficiency + (1|participant) + (1|prompt), data, REML = F)

Text quality (S1)	Bc pause * proficiency	-.02	.02	-1.10	.27
Text quality (S2)	Bc pause * proficiency	< .01	.01	.16	.87
Text quality (S3)	Bc pause * proficiency	-.01	.02	-.35	.73
Text quality (S4)	Bc pause * proficiency	< -.01	.02	-.26	.79
Text quality (S5)	Bc pause * proficiency	-.02	.02	-1.15	.26

S1: Model = lmer(text_score ~ bc_pause_frequency_s1 * cproficiency + (1|participant) + (1|prompt), data, REML = F)

S2: Model = lmer(text_score ~ bc_pause_frequency_s2 * cproficiency + (1|participant), data, REML = F)

S3: Model = lmer(text_score ~ bc_pause_frequency_s3 * cproficiency + (1|participant) + (1|prompt), data, REML = F)

S4: Model = lmer(text_score ~ bc_pause_frequency_s4 * cproficiency + (1|participant), data, REML = F)

S5: Model = lmer(text_score ~ bc_pause_frequency_s5 * cproficiency + (1|participant) + (1|prompt), data, REML = F)

Text quality (S1)	Bs pause * proficiency	-.01	.02	-.78	.43
Text quality (S2)	Bs pause * proficiency	< -.01	.02	-.21	.84
Text quality (S3)	Bs pause * proficiency	-.03	.02	-1.88	.06
Text quality (S4)	Bs pause * proficiency	-.02	.02	-1.34	.19
Text quality (S5)	Bs pause * proficiency	-.02	.01	-1.60	.11

S1: Model = lmer(text_score ~ bs_pause_frequency_s1 * cproficiency + (1|participant) + (1|prompt), data, REML = F)

S2: Model = lmer(text_score ~ bs_pause_frequency_s2 * cproficiency + (1|participant) + (1|prompt), data, REML = F)

S3: Model = lmer(text_score ~ bs_pause_frequency_s3 * cproficiency + (1|participant) + (1|prompt), data, REML = F)

S4: Model = lmer(text_score ~ bs_pause_frequency_s4 * cproficiency + (1|participant) + (1|prompt), data, REML = F)

S5: Model = lmer(text_score ~ bs_pause_frequency_s5 * cproficiency + (1|participant) + (1|prompt), data, REML = F)

Text quality (S1)	Bcr pause * proficiency	.02	.03	.82	.41
Text quality (S2)	Bcr pause * proficiency	.01	.02	.78	.44
Text quality (S3)	Bcr pause * proficiency	< -.01	.01	-.04	.97
Text quality (S4)	Bcr pause * proficiency	< .01	.01	.13	.90
Text quality (S5)	Bcr pause * proficiency	-.01	.01	-.88	.38

S1: Model = lmer(text_score ~ bcr_pause_frequency_s1 * cproficiency + (1|participant) + (1|prompt), data, REML = F)

S2: Model = lmer(text_score ~ bcr_pause_frequency_s2 * cproficiency + (1|participant) + (1|prompt), data, REML = F)

S3: Model = lmer(text_score ~ bcr_pause_frequency_s3 * cproficiency + (1|participant) + (1|prompt), data, REML = F)

S4: Model = lmer(text_score ~ bcr_pause_frequency_s4 * cproficiency + (1|participant) + (1|prompt), data, REML = F)

S5: Model = lmer(text_score ~ bcr_pause_frequency_s5 * cproficiency + (1|participant) + (1|prompt), data, REML = F)

Pause duration

Text quality (S1)	Total pause * proficiency	< -.01	< .01	< -.01	1.00
Text quality (S2)	Total pause * proficiency	< .01	.01	.04	.97
Text quality (S3)	Total pause * proficiency	.01	< .01	1.75	.08
Text quality (S4)	Total pause * proficiency	< -.01	< .01	-.16	.87
Text quality (S5)	Total pause * proficiency	< .01	< .01	.16	.88

S1: Model = lmer(text_score ~ total_pause_duration_s1 * cproficiency + (1|participant) + (1|prompt), data, REML = F)

S2: Model = lmer(text_score ~ total_pause_duration_s2 * cproficiency + (1|participant) + (1|prompt), data, REML = F)

S3: Model = lmer(text_score ~ total_pause_duration_s3 * cproficiency + (1|participant) + (1|prompt), data, REML = F)

S4: Model = lmer(text_score ~ total_pause_duration_s4 * cproficiency + (1|participant) + (1|prompt), data, REML = F)

S5: Model = lmer(text_score ~ total_pause_duration_s4 * cproficiency + (1|participant) + (1|prompt), data, REML = F)

Text quality (S1)	Ww pause * proficiency	< -.01	< .01	-.96	.34
Text quality (S2)	Ww pause * proficiency	< -.01	< .01	-.71	.48
Text quality (S3)	Ww pause * proficiency	< -.01	< .01	-.81	.42
Text quality (S4)	Ww pause * proficiency	< .01	< .01	.69	.50
Text quality (S5)	Ww pause * proficiency	< -.01	< .01	-1.56	.13

S1: Model = lmer(text_score ~ ww_pause_duration_s1 * cproficiency + (1|participant) + (1|prompt), data, REML = F)

S2: Model = lmer(text_score ~ ww_pause_duration_s2 * cproficiency + (1|participant) + (1|prompt), data, REML = F)

S3: Model = lmer(text_score ~ ww_pause_duration_s3 * cproficiency + (1|participant) + (1|prompt), data, REML = F)

S4: Model = lmer(text_score ~ ww_pause_duration_s4 * cproficiency + (1|participant), data, REML = F)

S5: Model = lmer(text_score ~ ww_pause_duration_s5 * cproficiency + (1|participant), data, REML = F)

Text quality (S1)	Bpc pause * proficiency	< .01	< .01	.40	.69
Text quality (S2)	Bpc pause * proficiency	< .01	< .01	.84	.41
Text quality (S3)	Bpc pause * proficiency	.01	< .01	1.86	.07
Text quality (S4)	Bpc pause * proficiency	-.01	.01	-1.42	.16
Text quality (S5)	Bpc pause * proficiency	< -.01	< .01	-.15	.88

S1: Model = lmer(text_score ~ bpc_pause_duration_s1 * cproficiency + (1|participant) + (1|prompt), data, REML = F)
S2: Model = lmer(text_score ~ bpc_pause_duration_s2 * cproficiency + (1|participant) + (1|prompt), data, REML = F)
S3: Model = lmer(text_score ~ bpc_pause_duration_s3 * cproficiency + (1|participant) + (1|prompt), data, REML = F)
S4: Model = lmer(text_score ~ bpc_pause_duration_s4 * cproficiency + (1|participant) + (1|prompt), data, REML = F)
S5: Model = lmer(text_score ~ bpc_pause_duration_s5 * cproficiency + (1|participant) + (1|prompt), data, REML = F)

Text quality (S1)	Bw pause * proficiency	< .01	< .01	.55	.59
Text quality (S2)	Bw pause * proficiency	< .01	< .01	.19	.85
Text quality (S3)	Bw pause * proficiency	< .01	< .01	1.16	.25
Text quality (S4)	Bw pause * proficiency	.01	< .01	1.27	.21
Text quality (S5)	Bw pause * proficiency	< .01	< .01	1.27	.21

S1: Model = lmer(text_score ~ bw_pause_duration_s1 * cproficiency + (1|participant) + (1|prompt), data, REML = F)
S2: Model = lmer(text_score ~ bw_pause_duration_s2 * cproficiency + (1|participant) + (1|prompt), data, REML = F)
S3: Model = lmer(text_score ~ bw_pause_duration_s3 * cproficiency + (1|participant) + (1|prompt), data, REML = F)
S4: Model = lmer(text_score ~ bw_pause_duration_s4 * cproficiency + (1|participant) + (1|prompt), data, REML = F)
S5: Model = lmer(text_score ~ bw_pause_duration_s5 * cproficiency + (1|participant) + (1|prompt), data, REML = F)

Text quality (S1)	Bc pause * proficiency	< .01	< .01	1.35	.18
Text quality (S2)	Bc pause * proficiency	< -.01	< .01	-.53	.60
Text quality (S3)	Bc pause * proficiency	< -.01	< .01	-.76	.45
Text quality (S4)	Bc pause * proficiency	< .01	< .01	.01	.99
Text quality (S5)	Bc pause * proficiency	< .01	< .01	1.09	.28

S1: Model = lmer(text_score ~ bc_pause_duration_s1 * cproficiency + (1|participant) + (1|prompt), data, REML = F)
S2: Model = lmer(text_score ~ bc_pause_duration_s2 * cproficiency + (1|participant) + (1|prompt), data, REML = F)
S3: Model = lmer(text_score ~ bc_pause_duration_s3 * cproficiency + (1|participant) + (1|prompt), data, REML = F)
S4: Model = lmer(text_score ~ bc_pause_duration_s4 * cproficiency + (1|participant), data, REML = F)
S5: Model = lmer(text_score ~ bc_pause_duration_s5 * cproficiency + (1|participant), data, REML = F)

Text quality (S1)	Bs pause * proficiency	< .01	< .01	.96	.34
Text quality (S2)	Bs pause * proficiency	< -.01	< .01	-.54	.59
Text quality (S4)	Bs pause * proficiency	< .01	< .01	.89	.38
Text quality (S5)	Bs pause * proficiency	< .01	< .01	.57	.57

S1: Model = lmer(text_score ~ bs_pause_duration_s1 * cproficiency + (1|participant) + (1|prompt), data, REML = F)

S2: Model = lmer(text_score ~ bs_pause_duration_s2 * cproficiency + (1|participant) + (1|prompt), data, REML = F)

S4: Model = lmer(text_score ~ bs_pause_duration_s4 * cproficiency + (1|participant) + (1|prompt), data, REML = F)

S5: Model = lmer(text_score ~ bs_pause_duration_s4 * cproficiency + (1|participant), data, REML = F)

Text quality (S1)	Bcr pause * proficiency	< .01	< .01	.71	.49
Text quality (S2)	Bcr pause * proficiency	< -.01	< .01	-.58	.56
Text quality (S3)	Bcr pause * proficiency	< -.01	< .01	-.50	.62
Text quality (S4)	Bcr pause * proficiency	< .01	< .01	.20	.84
Text quality (S5)	Bcr pause * proficiency	< .01	< .01	.86	.39

S1: Model = lmer(text_score ~ bcr_pause_duration_s1 * cproficiency + (1|participant) + (1|prompt), data, REML = F)

S2: Model = lmer(text_score ~ bcr_pause_duration_s2 * cproficiency + (1|participant) + (1|prompt), data, REML = F)

S3: Model = lmer(text_score ~ bcr_pause_duration_s3 * cproficiency + (1|participant) + (1|prompt), data, REML = F)

S4: Model = lmer(text_score ~ bcr_pause_duration_s4 * cproficiency + (1|participant) + (1|prompt), data, REML = F)

S5: Model = lmer(text_score ~ bcr_pause_duration_s5 * cproficiency + (1|participant) + (1|prompt), data, REML = F)

Revision

Text quality (S1)	Total * proficiency	.02	.01	1.40	.16
Text quality (S2)	Total * proficiency	.02	.02	1.11	.27
Text quality (S3)	Total * proficiency	.03	.02	1.86	.07
Text quality (S4)	Total * proficiency	.03	.02	1.59	.11
Text quality (S5)	Total * proficiency	-.01	.01	-.61	.55

S1: Model = lmer(text_score ~ log(total_revision_s1) * cproficiency + (1|participant) + (1|prompt), data, REML = F)

S2: Model = lmer(text_score ~ log(total_revision_s2) * cproficiency + (1|participant) + (1|prompt), data, REML = F)

S3: Model = lmer(text_score ~ log(total_revision_s3) * cproficiency + (1|participant) + (1|prompt), data, REML = F)

S4: Model = lmer(text_score ~ log(total_revision_s4) * cproficiency + (1|participant) + (1|prompt), data, REML = F)

S5: Model = lmer(text_score ~ log(total_revision_s5) * cproficiency + (1|participant) + (1|prompt), data, REML = F)

Text quality (S1)	Below word * proficiency	.01	.01	.80	.43
Text quality (S2)	Below word * proficiency	< .01	.01	.23	.82
Text quality (S3)	Below word * proficiency	.02	.02	1.26	.21
Text quality (S4)	Below word * proficiency	.01	.01	.43	.67
Text quality (S5)	Below word * proficiency	< -.01	.01	-.22	.83

S1: Model = lmer(text_score ~ log(below_word_revision_s1) * cproficiency + (1|participant) + (1|prompt), data, REML = F)

S2: Model = lmer(text_score ~ log(below_word_revision_s2) * cproficiency + (1|participant) + (1|prompt), data, REML = F)

S3: Model = lmer(text_score ~ log(below_word_revision_s3) * cproficiency + (1|participant) + (1|prompt), data, REML = F)

S4: Model = lmer(text_score ~ log(below_word_revision_s4) * cproficiency + (1|participant) + (1|prompt), data, REML = F)

S5: Model = lmer(text_score ~ log(below_word_revision_s5) * cproficiency + (1|participant) + (1|prompt), data, REML = F)

Text quality (S2)	Word * proficiency	.03	.02	1.70	.09
Text quality (S3)	Word * proficiency	.02	.02	1.13	.26
Text quality (S4)	Word * proficiency	.01	.01	.51	.61
Text quality (S5)	Word * proficiency	-.01	.02	-.40	.69

S2: Model = lmer(text_score ~ log(word_revision_s2) * cproficiency + (1|participant) + (1|prompt), data, REML = F)

S3: Model = lmer(text_score ~ log(word_revision_s3) * cproficiency + (1|participant) + (1|prompt), data, REML = F)

S4: Model = lmer(text_score ~ log(word_revision_s4) * cproficiency + (1|participant) + (1|prompt), data, REML = F)

S5: Model = lmer(text_score ~ log(word_revision_s5) * cproficiency + (1|participant) + (1|prompt), data, REML = F)

Text quality (S1)	Below clause * proficiency	-.01	.02	-.42	.68
Text quality (S2)	Below clause * proficiency	< -.01	.02	-.18	.86
Text quality (S3)	Below clause * proficiency	.02	.02	.92	.36
Text quality (S5)	Below clause * proficiency	< .01	.02	.24	.81

S1: Model = lmer(text_score ~ log(below_clause_revision_s1) * cproficiency + (1|participant) + (1|prompt), data, REML = F)

S2: Model = lmer(text_score ~ log(below_clause_revision_s2) * cproficiency + (1|participant) + (1|prompt), data, REML = F)

S3: Model = lmer(text_score ~ log(below_clause_revision_s3) * cproficiency + (1|participant) + (1|prompt), data, REML = F)

S5: Model = lmer(text_score ~ log(below_clause_revision_s5) * cproficiency + (1|participant) + (1|prompt), data, REML = F)

Text quality (S1)	Clause and above * proficiency	-.07	.04	-1.73	.09
Text quality (S2)	Clause and above * proficiency	-.03	.03	-.98	.33
Text quality (S3)	Clause and above * proficiency	.02	.03	.62	.54
Text quality (S5)	Clause and above * proficiency	-.02	.04	-.41	.68

S1: Model = lmer(text_score ~ log(clause_and_above_revision_s1) * cproficiency + (1|participant) + (1|prompt), data, REML = F)

S2: Model = lmer(text_score ~ log(clause_and_above_revision_s2) * cproficiency + (1|participant) + (1|prompt), data, REML = F)

S3: Model = lmer(text_score ~ log(clause_and_above_revision_s3) * cproficiency + (1|participant), data, REML = F)

S4: Model = lmer(text_score ~ log(clause_and_above_revision_s4) * cproficiency + (1|participant) + (1|prompt), data, REML = F)

Text quality (S1)	Pre-contextual * proficiency	.01	.01	.92	.36
Text quality (S2)	Pre-contextual * proficiency	.01	.02	.86	.39
Text quality (S3)	Pre-contextual * proficiency	.02	.02	1.39	.17
Text quality (S4)	Pre-contextual * proficiency	.01	.01	.55	.58
Text quality (S5)	Pre-contextual * proficiency	< -.01	.01	-.32	.75

S1: Model = lmer(text_score ~ log(precontextual_revision_s1) * cproficiency + (1|participant) + (1|prompt), data, REML = F)

S2: Model = lmer(text_score ~ log(precontextual_revision_s2) * cproficiency + (1|participant) + (1|prompt), data, REML = F)

S3: Model = lmer(text_score ~ log(precontextual_revision_s3) * cproficiency + (1|participant) + (1|prompt), data, REML = F)

S4: Model = lmer(text_score ~ log(precontextual_revision_s4) * cproficiency + (1|participant) + (1|prompt), data, REML = F)

S5: Model = lmer(text_score ~ log(precontextual_revision_s5) * cproficiency + (1|participant) + (1|prompt), data, REML = F)

Text quality (S1)	Contextual * proficiency	.04	.03	1.61	.11
Text quality (S2)	Contextual * proficiency	.01	.01	.51	.61
Text quality (S3)	Contextual * proficiency	.02	.02	1.21	.23
Text quality (S4)	Contextual * proficiency	.02	.02	1.38	.17
Text quality (S5)	Contextual * proficiency	-.01	.02	-.85	.40

S1: Model = lmer(text_score ~ log(contextual_revision_s1) * cproficiency + (1|participant) + (1|prompt), data, REML = F)

S2: Model = lmer(text_score ~ log(contextual_revision_s2) * cproficiency + (1|participant) + (1|prompt), data, REML = F)

S3: Model = lmer(text_score ~ log(contextual_revision_s3) * cproficiency + (1|participant) + (1|prompt), data, REML = F)

S4: Model = lmer(text_score ~ log(contextual_revision_s4) * cproficiency + (1|participant) + (1|prompt), data, REML = F)

S5: Model = lmer(text_score ~ log(contextual_revision_s5) * cproficiency + (1|participant) + (1|prompt), data, REML = F)

Text quality (S1)	Pinyin * proficiency	.01	.01	.88	.38
Text quality (S2)	Pinyin * proficiency	< .01	.02	.17	.87
Text quality (S3)	Pinyin * proficiency	.01	.01	.93	.36
Text quality (S4)	Pinyin * proficiency	.01	.02	.84	.40
Text quality (S5)	Pinyin * proficiency	< -.01	.01	-.27	.79

S1: Model = lmer(text_score ~ log(pinyin_revision_s1) * cproficiency + (1|participant) + (1|prompt), data, REML = F)

S2: Model = lmer(text_score ~ log(pinyin_revision_s2) * cproficiency + (1|participant) + (1|prompt), data, REML = F)

S3: Model = lmer(text_score ~ log(pinyin_revision_s3) * cproficiency + (1|participant) + (1|prompt), data, REML = F)

S4: Model = lmer(text_score ~ log(pinyin_revision_s4) * cproficiency + (1|participant) +

(1|prompt), data, REML = F)					
S5: Model = lmer(text_score ~ log(pinyin_revision_s5) * cproficiency + (1|participant) + (1|prompt), data, REML = F)					
Text quality (S1)	Character * proficiency	.03	.02	1.51	.13
Text quality (S2)	Character * proficiency	.02	.02	.94	.35
Text quality (S3)	Character * proficiency	.02	.02	1.14	.26
Text quality (S4)	Character * proficiency	.03	.02	1.67	.10
Text quality (S5)	Character * proficiency	-.02	.02	-1.02	.31
S1: Model = lmer(text_score ~ log(character_revision_s1) * cproficiency + (1|participant) + (1|prompt), data, REML = F)					
S2: Model = lmer(text_score ~ log(character_revision_s2) * cproficiency + (1|participant) + (1|prompt), data, REML = F)					
S3: Model = lmer(text_score ~ log(character_revision_s3) * cproficiency + (1|participant) + (1|prompt), data, REML = F)					
S4: Model = lmer(text_score ~ log(character_revision_s4) * cproficiency + (1|participant) + (1|prompt), data, REML = F)					
S5: Model = lmer(text_score ~ log(character_revision_s5) * cproficiency + (1|participant) + (1|prompt), data, REML = F)					

Note: Ww = Within a word, Bpc = Between Pinyin and character(s), Bw = Between words, Bc = Between clauses, Bs = Between sentences, Bcr = Between contextual revision