INVESTIGATING TEST-TAKERS' COGNITIVE PROCESSES WHILE COMPLETING AN INTEGRATED READING-TO-WRITE TASK: EVIDENCE FROM EYE-TRACKING, STIMULATED RECALL AND QUESTIONNAIRE

PUCHENG WANG

THESIS SUBMITTED TO THE DEPARTMENT OF LINGUISTICS AND ENGLISH LANGUAGE, LANCASTER UNIVERSITY FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

SEPTEMBER 2018

DECLARATION

This thesis is my own work and has not been submitted for the award of a higher degree elsewhere.

Pucheng Wang
September 2018

ABSTRACT

The TBEM-8 (Test for Business English Majors, band 8) is a newly-developed, nationwide test of business English proficiency administered to business English undergraduates in China at the end of their final year. One notable feature of the test is that it includes a reading-to-write task in which test-takers read texts in English and Chinese and then use this information to write an essay on a business-related topic. Although the test has been operational for several years, there is currently little validity evidence to support claims about the cognitive processing which takes place during this reading-to-write task. This presents a threat to the quality of inferences drawn from test scores.

The present research examined test-takers' cognitive processes while completing the TBEM-8 reading-to-write task, aiming to gain further insights into cognitive processing on this integrated task type. Two separate studies were conducted. In Study I, 16 participants completed this task while their eye movements were tracked by a Tobii TX300 eye-tracker. These eye traces then formed the stimuli for a stimulated recall session to elicit cognitive processes; in Study II, another 172 participants responded to a reading-to-write process questionnaire after completing the task. This questionnaire was developed by Chan (2013) and adapted for the TBEM-8 reading-to-write task. A pilot study was also conducted to finalise the main study questionnaire, in which 40 items were grouped to reflect the cognitive processes that writers are hypothesised to undergo.

The results showed that test-takers engaged in a wide range of cognitive processes specified in Shaw and Weir's (2007) model of writing and Spivey's (1990, 1997, 2001) discourse synthesis model during task completion, thus justifying the current use of it in the TBEM-8 test. Text interpretation and selecting were the two most frequently reported processes according to participants' stimulated recalls, and macro-planning and translating were the two least reported processes. A high level of agreement was found in participants' responses to the reading-to-write process questionnaire, with more than 70 percent of participants choosing either "agree" or "strongly agree" in 28 items, and only four items achieving an agreement rate below 60 percent.

The correlation analysis between the use of cognitive processes/eye-tracking measures

and test-takers' performance on the TBEM-8 reading-to-write task yielded no statistically significant results (at the 0.05 or 0.01 levels), except for a moderate positive correlation (ρ=.499, p=.049) between the participants' max visit duration on Source 5 (key concepts and expressions) and their reading-to-write performance, and one (ρ=.432, p=.098) between the counts of text interpretation-2 process (reading source materials) and the task performance if the p-value was set to 0.1.

This study demonstrated the usefulness of combining eye-tracking, stimulated recall and questionnaire methods for generating insights into the complexity of cognitive processing on an integrated reading-to-write task. Findings from the analysis of all sources of data were triangulated and discussed, providing a solid basis for the conclusions drawn about test-takers' cognitive processing during task completion. Also, a model of reading-to-write process was proposed to illustrate how different categories of cognitive processes examined in this study interact with each other for successful task completion.

ACKNOWLEDGEMENTS

At this moment when I finish writing the thesis, I would like to thank all those who have helped me in numerous ways throughout this journey of obtaining my PhD.

First of all, I would like to express my deep gratitude to my supervisor, Dr. Luke Harding, for his patient guidance, valuable suggestions, enthusiastic encouragement, and willingness to give his time so generously through all the time of researching and writing up this thesis. It was a real privilege and honour for me to share of his immense knowledge in language testing and of his extraordinary human qualities.

I am also extremely indebted to Dr. Tineke Brunfaut, for her insightful comments on the confirmation panel document, useful critiques of this research work, and for providing me with precious opportunity to work as research assistants in her projects, during which I learned basic skills to conduct eye-tracking and stimulated recall studies. I would also like to thank Dr. Baoquan Liu, my Master's supervisor, and Dr. Ying Xu, who provided me tremendous help in data collection and analysis, massive inspirations and support during this study.

Special thanks should also be given to Wood Marjorie, the former postgraduate programme coordinator, who always helped me with programme information, documents and various personal issues, and Dr. Bimali Indrarathne, Dr. Marije Michel (my post-confirmation panellist), for their support and encouragement during my study. My sincere gratitude also goes to Dr. Anuchit Toomaneejinda, who helped me with the formatting of the final draft and has been a wonderful companion throughout the PhD journey.

Last, but not least, I would like to thank my family, in particular my parents, Jinping Wang and Meijing Chen, and my girlfriend, Liujing Tang, for their unconditional support, encouragement and love, without which I would not have come this far.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1 INTRODUCTION

1.1 Background of the study

The past two decades have witnessed a growing interest among language testers in integrated writing tasks. It is generally considered that writing is unlikely to be done separately from other skills, instead, it tends to be dependent on gathering information from outside sources (Esmaeili, 2002; Flower, Stein, Ackerman, Kantz, McCormick, & Peck, 1990). Compared with independent writing tasks, which have often been criticised for decontextualising writing activities and under-representing the writing construct, integrated writing tasks have been proposed as a promising task type in writing assessment (Hamp-Lyons and Kroll, 1997; Plakans, 2008; Weigle, 2004). For example, the TOEFL iBT test (Test of English as a Foreign Language Internet-based Test) includes both integrated and independent writing tasks in its writing component. The rationale for this combination is that the concurrent use of these two types of writing tasks may enhance, to some extent, the authenticity and validity of a writing test which is designed for a specific academic purpose (Cumming, Kantor, Baba, Erdosy, Eouanzoui, & James, 2005, 2006).

Typically, source materials are provided in an integrated writing task. Test-takers are required to comprehend these sources (either in written or oral format), extract relevant information, and/or synthesise personal ideas in their own writing. By providing an accurate simulation of real tasks in the target language use domain, integrated writing tasks may better contextualise writing activities, thus enhancing the connection between test-takers' performance and real language use. Furthermore, the background information presented in the sources can help to mitigate the negative effects imposed on test-takers who are unfamiliar with the writing topics assigned (Jennings, Fox, Graves, & Shohamy, 1999; Lee and Anderson, 2007). Impact studies of integrated writing tasks have also demonstrated that this kind of test can improve, to some extent, washback on teaching and learning of writing (Feak and Dobson, 1996; Weigle, 2004).

Despite these advantages, there have been several constraints on using integrated writing tasks for assessment purposes. One of the most fundamental constraints is what psychometricians call "task dependencies" (Cumming, 2013). In an integrated writing task, test-takers' performance is dependent on variables such as the ability to read and/or listen besides the ability to write. The presence of reading or listening input may pose a threat to the performance of test-takers who lack adequate comprehension abilities, thus compromising the validity of measurements of writing abilities. This leads to another major limitation of integrated writing tasks, that is, they "require threshold levels of abilities for competent performance, producing results for examinees that may not compare neatly across different ability levels" (Cumming, 2013, p. 2). A further challenge is the scoring of integrated writing tasks in that the constructs of these tasks remain ill-defined and are amorphous due to the various genres of this task type; in addition, textual borrowing may make it difficult to distinguish the text produced by test-takers from source materials (Shi, 2004; Yu, 2013).

As discussed above, promises and perils coexist in integrated writing assessment, calling for more research efforts to improve our understanding of this task type. An urgent need now is to refine the constructs of integrated writing tasks, thus setting the groundwork for building a comprehensive framework for systematically researching integrated writing assessment.

1.2 The TBEM-8 reading-to-write task

The integrated writing task to be investigated in this study is the TBEM-8 (Test for Business English Majors, Band 8) reading-to-write task. The TBEM is a criterion-referenced English language test administered to undergraduate students majoring in business English in China. As the four-year undergraduate programme is divided by the teaching syllabus into the foundation stage (the first two years) and the advanced stage (the last two years), correspondingly, the TBEM test battery consists of TBEM-4 and TBEM-8, which assess students' business English proficiency at the end of these two stages. The TBEM-8 reading-to-write task is one of two tasks (the other one is a data commentary task) in the TBEM-8 writing component, the purpose of which is to measure the writing proficiency of students to examine whether they meet the required levels of writing abilities as specified in the teaching syllabus at the end of the advanced stage.

One notable feature of the TBEM-8 reading-to-write task is that it includes both English and Chinese source materials; test-takers are required to read and integrate the information in these sources into an essay on a business-related topic (see Section 3.3.2 for more details about this task). Although the task has been operational for several years, there is currently little validity evidence to support claims about the cognitive processing which takes place while completing this task. This presents a threat to the quality of inferences drawn from test scores. Also, there is a parallel need to explore the best methods for eliciting data on cognitive processing in integrated writing tasks.

1.3 The current study

As Kunnan (1988) claimed, the central location of intense language assessment research has been validation; in order to establish the validity of score interpretations on a certain test, validation evidence must be collected related to different aspects of validity.

The cognitive aspect of the validity of a writing test refers to "how closely it represents the cognitive processing involved in writing contexts beyond the test itself" (Shaw and Weir, 2007, p. 34). This study makes use of Shaw and Weir's socio-cognitive framework (see Section 2.4.1 for more details about the rationale for choosing this framework) for validating writing tests. In doing so, the study aims to establish cognitive validity evidence for the TBEM-8 reading-to-write task by examining test-takers' cognitive processes during task completion on an archetypal task, thus clarifying the construct inherent in this task. The findings will provide further insights into the usefulness of Shaw and Weir's (2007) framework in validating integrated writing tests (its application is currently limited to writing-only tests), and, most importantly, into an understanding of reading-to-write processes.

Two separate studies were conducted in this research study (see Chapter 3 for details about the methodology). In Study I, 16 participants completed the TBEM-8 reading-to-write task while their eye movements were tracked. These eye traces then formed the stimuli for a stimulated recall to elicit cognitive processes. Findings from Study I fed into revisions of a reading-to-write process questionnaire developed by Chan (2013), and the revised questionnaire was then administered in Study II to another 172 participants after they completed the TBEM-8 reading-to-write task.

1.4 Overview of the thesis

The first chapter has presented the background, topic, and aims of the study. This section provides an overview of the thesis.

Chapter Two is a literature review and contains the background for the current study. It considers three broad areas: first, in order to gain an understanding of the reading-to-write processes, relevant models of writing and reading processes, as well as a discourse synthesis model are reviewed; second, an overview of studies on integrated writing tasks is provided in terms of four different topics: comparison studies between independent and integrated writing tasks, discourse features of the written products in integrated writing tasks, and processes and the use of source texts in integrated writing tasks; and third, several validation frameworks are reviewed, and the cognitive processes to be examined, essential to this study, are proposed and relevant research on these processes is discussed. This chapter ends with two sets of research questions related to the studies (see Section 2.5 for these questions).

An overall design of the study is presented in Chapter Three. It starts with a discussion of the methodological underpinning to each research method chosen in this research study, to explain: first, why the combination of eye-tracking and stimulated recall methods is useful in Study I; and second, how the reading-to-write process questionnaire was used (Study II) to complement the data collected in Study I. After this discussion, details such as participants, data collection procedures, and methods of data analysis for each independent study are presented.

Chapter Four looks at the findings from the analysis of eye-tracking data collected in the eye-tracking and stimulated recall study (Study I). First, the results of a correlation analysis are presented to demonstrate the relationships between test-takers' performance on the TBEM-8 reading-to-write task and the IELTS test; second, a heat-map output is presented to show the overall distribution of test-takers' attention throughout task completion; third, four eye-tracking measures - *time to first fixation*, *total visit duration*, *visit count*, and *visit duration* - illustrate in detail how test-takers engage with the source materials during task completion; and finally, the results of a set of correlation analyses between the eye-tracking measures and test-takers' scores on the TBEM-8 reading-to-write task are presented.

Chapter Five presents the findings from the analysis of stimulated recall data collected in the eye-tracking and stimulated recall study (Study I). Detailed results of coding are presented with quotes from test-takers' verbal reports, to demonstrate in detail the way they applied each type of cognitive processes during task completion. Also, the relationships between the use of these cognitive processes and test-takers' performance on the task are also examined in this chapter.

Chapter Six reports on the results from the analysis of questionnaire data collected in Study II. It begins with the results of an internal consistency analysis, demonstrating whether each group of items in the reading-to-write process questionnaire reliably measured the same type of cognitive processes. Test-takers' agreement rates for each item are then presented in this chapter and it ends with the results of a set of Mann-Whitney U tests to investigate if higher- and lower-scoring participants responded differently to each item in the questionnaire.

In Chapter Seven, three topics are discussed. First, findings from all three sources (eye-tracking, stimulated recall, and questionnaire) are triangulated to answer the first set of research questions: test-takers' cognitive processing during task completion are discussed to look at how they fit in with previously published knowledge; second, findings are triangulated to address the second set of research questions, with a discussion specifically of how test-takers engaged with source materials in the TBEM-

8 reading-to-write task; and third, a reading-to-write process model is proposed to illustrate how the proposed cognitive processes interact with each other.

The final chapter, Chapter Eight, provides a summary of the current study. Implications of the findings are considered for the TBEM-8 reading-to-write task, for the reading-to-write processes in general, and for the methods used to examine reading-to-write processes. Limitations of this study and an agenda for future research are presented, and, finally, a summary of the thesis concludes this chapter.

CHAPTER 2 LITERATURE REVIEW

2.1 Introduction

This chapter provides background from the research literature relevant to the current study. First, in order to gain insights into the reading-to-write process, Section 2.2 reviews relevant writing models and a reading model developed by Khalifa and Weir (2009). Section 2.3 then reviews studies on integrated writing tasks. In section 2.4, several validation frameworks are reviewed, and a set of cognitive processes to be investigated in this study are proposed and described. Section 2.5 proposes the research questions to be addressed in this study and a summary of this chapter is provided in Section 2.6.

2.2 Cognitive processes involved in integrating reading with writing

Findings in recent literature have suggested that integrated reading-to-write tasks tap into a differing set of literacy skills which go beyond those normally required by traditional independent writing tasks (Chan, Wu and Weir, 2014; Chan, 2017; Gebril and Plakans, 2013; Grabe, 2003; Plakans, 2009a, 2009b; Weir, Vidakovic and Galaczi, 2013). If reading-to-write skill differs from reading or writing skills in isolation, there is a need to model the processes involved in reading-to-write tasks. However, in the existing literature, this type of task has not been systematically defined, and reading-to-write processes are not well understood, although a number of models of writing have been proposed and refined over time.

2.2.1 Models of writing process

As writing is an indispensable part of any type of integrated writing tasks, reviewing relevant literature on writing models is necessary to achieve a basic understanding of the foundation of reading-to-write processes.

Before the 1960s, writing was often conceptualised as a process of transcribing speech and was regarded as "decontextualised" (Ellis, 1994, p. 188) and product-oriented as the final texts were often seen as "autonomous objects" in that different writing components were combined in accordance with a "system of rules" (Hyland, 2002, p. 6). Writing is now viewed as essentially a communicative act. Therefore, a written text is seen as discourse because the writer tries to involve the reader in the context by using linguistic patterns which are influenced by various social constraints, for example, writers' content knowledge and writing goals, and writers' relationship with readers. Any writing model needs to take these contextual elements into careful consideration in understanding writing as a social act.

In addition, writing is now seen as a cognitive activity and a number of researchers have proposed writing models that describe cognitive processing activities involved in writing (Bereiter and Scardamalia, 1987; Field, 2004; Grabe and Kaplan, 1996; Hayes, 1996, 2012; Hayes and Flower, 1980; Kellogg, 1994, 1996; Shaw and Weir, 2007). While these models may not provide a completely accurate picture of the writing process, they are useful for considering the possible factors that may influence the process. It should also be noted that the models listed above draw mainly on L1 research, but are still of importance to our understanding of L2 writing processes, since the literature on cognitive processing in L2 writing is relatively scarce, and L1 models of writing proficiency are commonly used as metrics in examining L2 writing (Shaw and Weir, 2007).

I. Hayes and Flower (1980)

Hayes and Flower provided an influential model of the writing process in 1980. They described the writing process in terms of three interactive components, first, the *task environment*, which includes the writing assignment and the text produced so far; second, the writer's *long-term memory*, which includes knowledge of topic, knowledge of audience and stored writing plans; third, a number of *cognitive processes*, including planning, translating ideas into texts and reviewing (see Figure 2.1 shown on the next page). It is the third part that demonstrates the mental process of writing as a cognitive activity, which has been influential to the subsequent writing research in this respect. Also, Hayes and Flower proposed that writing is not a linear process, but involves multiple recursions of *planning, translating* and *reviewing*. This conceptualization largely fixed the terminology of writing processes in the literature (Scardamalia and Bereiter, 1996). Although the Hayes-Flower (1980) model provided some useful insights into the writing process, the model has been criticised as it does not fully reflect the way in which writing processes vary with different task types, and does not distinguish skilled from unskilled writing (Shaw and Weir, 2007).

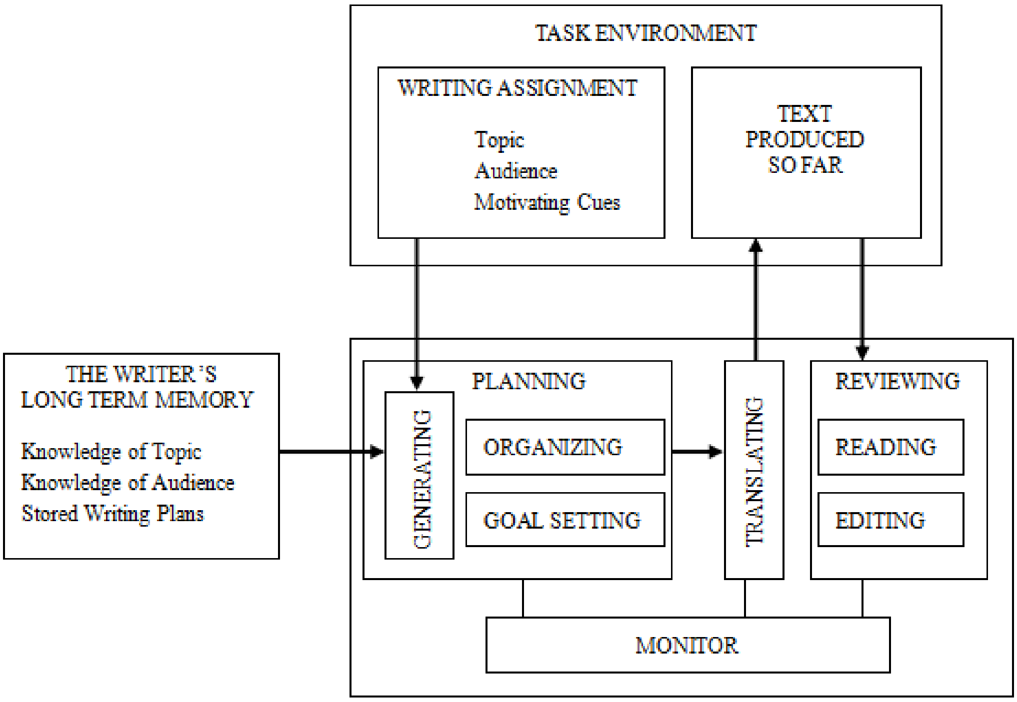

Figure 2.1: The Hayes-Flower (1980, p. 11) writing model

II. Hayes (1996)

Hayes' (1996) model (see Figure 2.2) is an updated version of the Hayes and Flower (1980) model. It looks at the writing process as consisting of two essential components: the *task environment* and the *individual*. The *task environment* is divided into the social environment and the physical environment. The social environment includes audience (real or imagined) for one's writing, and the possible collaborators during the process of writing; the physical environment consists of the text written so far, which affects and shapes the text to be produced, and the composing medium, for example, handwriting or using word processors.

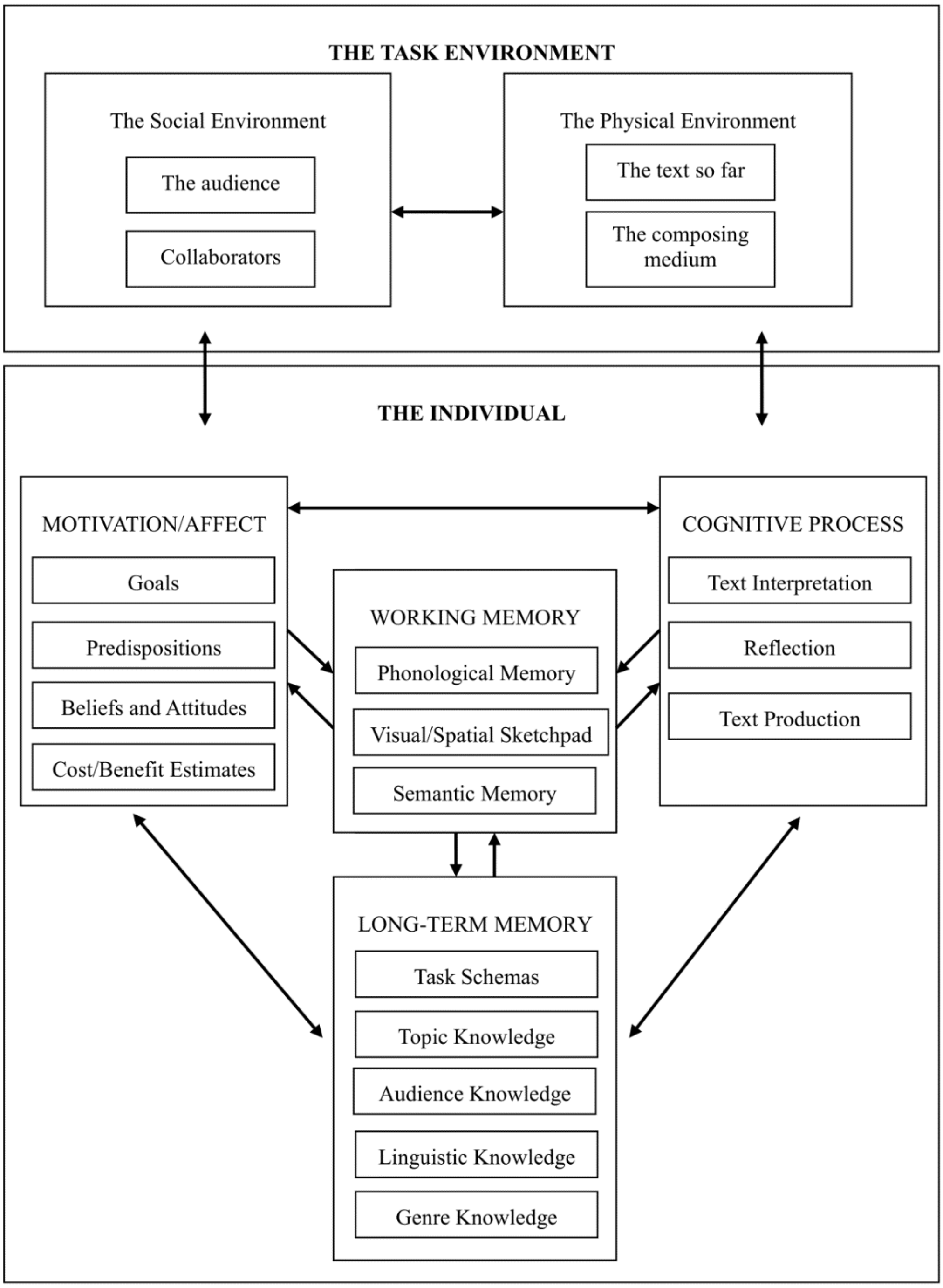

Figure 2.2: The Hayes (1996, p. 4) model

The central focus of the Hayes (1996) model is the *individual* part, which involves an interaction among four components: working memory, motivation and affect, cognitive processes and long-term memory. **Working memory** in this model is mainly based on Baddeley's (1986) conception of working memory, and is composed of three parts: phonological memory, which stores information of speech; the visual/spatial sketchpad, storing visually or spatially coded information such as written words or graphs; and semantic memory which stores conceptual information. **Motivation and affect** play important roles in Hayes' model. Specifically, a writer's goals, predispositions, beliefs and attitudes and cost/benefit estimates may influence how the writer is going to write and how much effort they are going to put in the writing activity (Weigle, 2002). Regarding the component of **cognitive processes**, the three major processes proposed in the 1980 model (planning, translating and reviewing) were replaced by three more general process categories: text interpretation, reflection and text production. Text interpretation, including listening, reading and scanning graphics, is the process during which "internal representations are created from linguistic and graphic input" (Weigle, 2002, p. 25). The reviewing process in the 1980 model was no longer considered as a separate process, but was included in the text interpretation process; reflection, taking place of planning, involves problem-solving and decision-making processes, through which writers achieve writing goals; translation was replaced by text production, in which new linguistic output is created by consulting writing plans or text produced so far. (4) The last component in the *individual* part is **long-term memory**, in which writers' knowledge and background information relevant to the writing task is stored. The Hayes' (1996) model attempted to illustrate the complex interactive nature of the four previously mentioned components in the writing process, however, no claims are made as to precisely the way in which these components interact, other than a claim of the theoretical relationships among them at a very general level.

Another contribution of Hayes' (1996) work is that he emphasised the significance of reading as a central process in writing, which aligns well with the proposition of the present study. Hayes pointed out three types of reading that are important in writing. The first of these is **reading to evaluate**, in which writers read the text that has been produced to detect any possible problems and make potential revisions; this type of reading is more commonly known as monitoring in recent literature on L2 writing assessment, as will be explained later in this chapter. The other two types of reading involved in writing are **reading instructions** and **reading source texts**. If writers create a representation of the task based on a misunderstanding of the task instructions, they may not be able to respond to the task appropriately. Similarly, since some writing tasks (for example, the reading-to-write tasks) are based upon source texts, a writer's ability to understand the source texts will almost certainly impact on their performance on the task.

III. Hayes (2012)

The most recent writing model of Hayes (2012), shown in Figure 2.3 on the next page, differs from the two previous models in a number of ways. One major change is in the writing processes component that, based on Chenoweth and Hayes' (2001, 2003) work, Hayes proposed that texts are produced through the interaction of four cognitive processes: a proposer, a translator, a reviser and a transcriber.

The **proposer** is a prelinguistic source that suggests a pool of ideas to be included in the text, which is then passed on to the translator; the proposer can take input from writing plans, from source materials, and even from writers' long-term memory and the text-written-so-far; ideas produced by the proposer are often in non-verbal form (Chenoweth and Hayes, 2001; Hayes and Berninger, 2014).

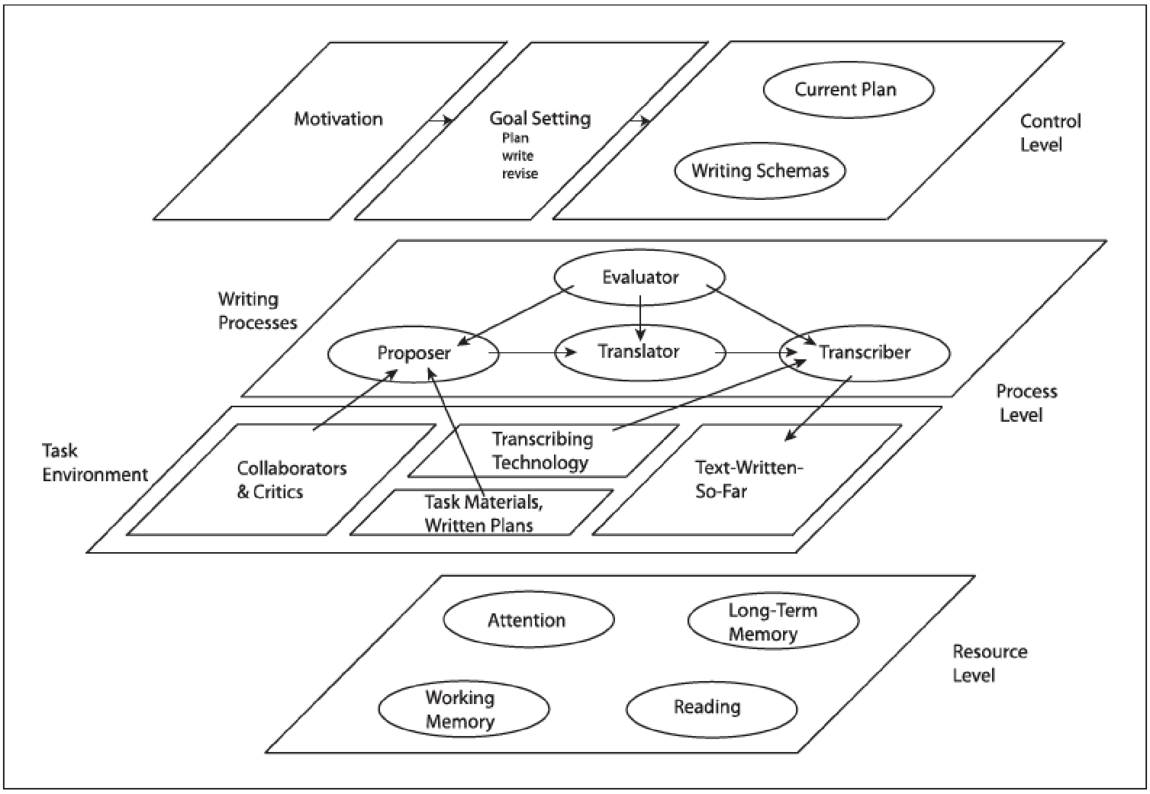

Figure 2.3: The Hayes (2012, p. 371) model

The **translator** receives ideas from the proposer, and converts them into grammatical strings of language, that is, translating non-verbal ideas into a verbal form of expression (Chenoweth and Hayes, 2003); to translate ideas, the translator draws mainly on writers' long-term memory, in particular the linguistic knowledge stored in it, and working memory resources. For L2 writers, translation appears to be the barrier that limits their writing fluency.

The **transcriber** then converts the linguistic strings produced by the translator into written text. It was believed in earlier work on the writing process that the transcription of adult writers was "so thoroughly automated that it would not have any significant impact on other writing processes and could safely be ignored" (Hayes, 2012, p. 371). However, more recent studies have discovered that the transcription played a critical role in the writing process (Berninger, Cartwright, Yates, Swanson and Abbott, 1994;

Bourdin and Fayol, 1994; Connelly, Gee and Walsh, 2007; Hayes and Chenoweth, 2006; Jones and Christensen, 1999). For example, Hayes and Chenoweth (2006) found that adult writers' transcribing process was slowed when verbal working memory was reduced, suggesting that transcription is very likely to compete with other processes in writing for cognitive sources and thus should be accounted for when modelling writing.

The **evaluator** examines the outputs of any of the above three processes and determines their adequacy to the task. For example, the evaluator may reject a proposed idea before it is translated to linguistic strings, or it may reject an already translated verbal form of expression before it is transcribed.

To sum up, the Hayes and Flower (1980) model and Hayes' (1996, 2012) models are considered to be significant in L1 writing research because they present the various factors that may influence writing, and, despite their age, contain features that are still meaningful in current literature on models of writing. The significance of writers' long-term memory and working memory in writing; the attempt to identify the interacting cognitive processes in writing; and the importance of text-written-so-far are all still considered useful ideas in modelling writing processes. Although these models are relatively complete in many aspects, they have one major shortcoming, that is, little attention paid to linguistic knowledge, which is another essential component of writing. The Grabe and Kaplan (1996) model of writing can be used to fill in this gap.

IV. Grabe and Kaplan (1996)

Grabe and Kaplan attempted to examine the cognitive processing activities involved in L2 writing in 1996. Their model (see Figure 2.4), based upon a framework of communicative language use developed by Chapelle, Grabe and Berns (1993), is one of the few L2 writing models in the literature. It has two major components: "a *context* for language use and a representation of the language user's *verbal working memory*" (Grabe and Kaplan, 1996, p. 233). They proposed that 'goal setting' is a process of setting goals and purposes for writing based on the contextual situation (for example, task, text and topic), and activates three components in the 'verbal processing unit': language competence, world knowledge and on-line processing assembly (i.e., execution of writing processes).

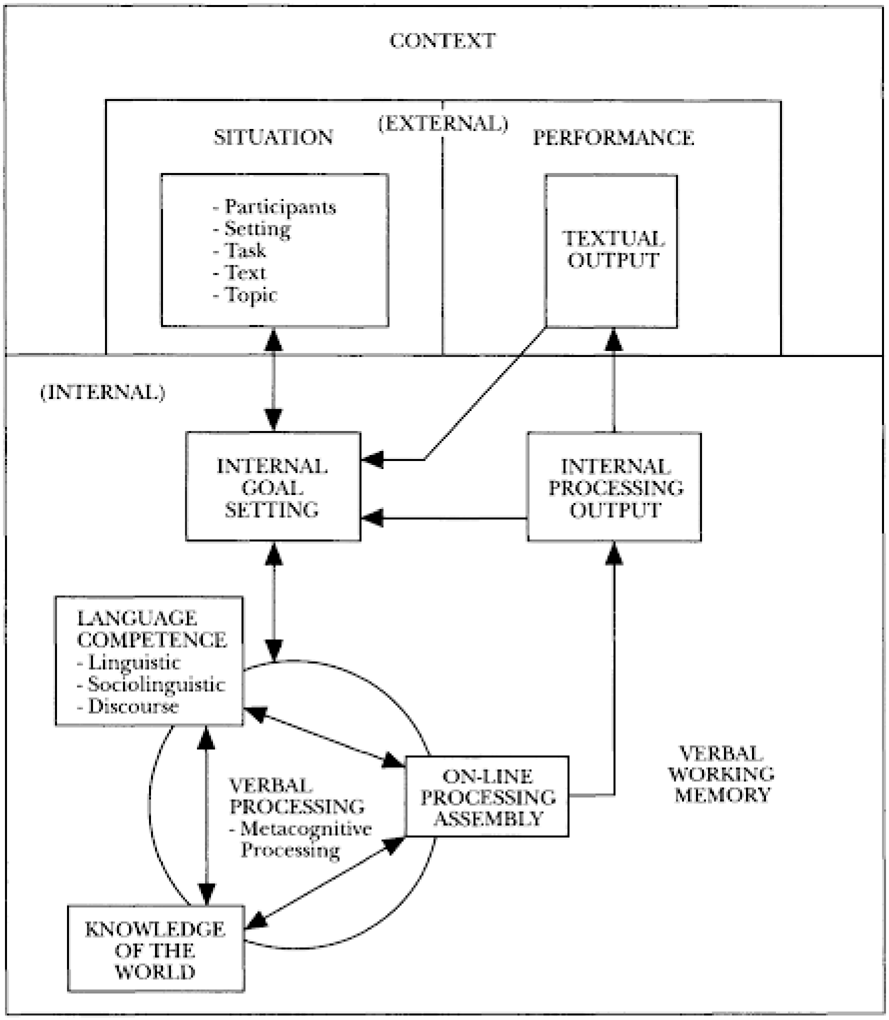

Figure 2.4: The Grabe and Kaplan (1996, p. 226) model

As discussed earlier, the Hayes' models paid little attention to writers' linguistic knowledge in writing; Grabe and Kaplan (1996) provided a list of three components of language competence relevant to writing, which includes linguistic knowledge and two other types of knowledge: sociolinguistic knowledge and discourse knowledge (see Table 2.1). Linguistic knowledge includes knowledge of the fundamental structural elements of language, which is regarded as a critical component of writing ability and the foundation for text construction (Grabe and Kaplan, 1996). Discourse knowledge refers to the knowledge of how cohesive text is constructed such as knowledge of semantic relations across clauses and knowledge of recognizing main topics. Sociolinguistic knowledge also plays a role in writing from the socio-cognitive perspective, for example, audience consideration and degree of formality.

Table 2.1: Taxonomy of language knowledge – adapted from Grabe and Kaplan (1996, p. 220-221) (Weigle, 2002)

I. **Linguistic knowledge**

- A. Knowledge of the written code
  1. Orthography
  2. Spelling
  3. Punctuation
  4. Formatting conventions (margins, paragraphing, spacing, etc.)
- B. Knowledge of phonology and morphology
  1. Sound/Letter correspondences
  2. Syllables (onset, rhyme/rhythm, coda)
  3. Morpheme structure (word-part knowledge)
- C. Vocabulary
  1. Interpersonal words and phrases
  2. Academic and pedagogical words and phrases
  3. Formal and technical words and phrases
  4. Topic-specific words and phrases
  5. Non-literal and metaphoric language
- D. Syntactic/Structural knowledge
  1. Basic syntactic patterns

2. Preferred formal writing structures (appropriate style)
3. Tropes and figures of expression
4. Metaphors/Similes

E. Awareness of differences across languages

F. Awareness of relative proficiency in different languages and registers

II. **Discourse Knowledge**

A. Knowledge of intrasentential and intersentential marking devices (cohesion, syntactic parallelism)

B. Knowledge of informational structuring (topic/comment, given/new, theme/rheme, adjacency pairs)

C. Knowledge of semantic relations across clauses

D. Knowledge of recognizing main topics

E. Knowledge of genre structure and genre constraints

F. Knowledge of organizing schemes (top-level discourse structure)

G. Knowledge of inferencing (bridging, elaborating)

H. Knowledge of differences in features of discourse structuring across languages and cultures

I. Awareness of different proficiency levels of discourse skills in different languages

III. **Sociolinguistic knowledge**

A. Functional uses of written language

B. Application and interpretable violation of Gricean maxims (Grice, 1975)

C. Register and situational parameters
1. Age of writer
2. Language used by writer (L1, L2, . . .)
3. Proficiency in language used
4. Audience considerations
5. Relatives status of interactants (power/politeness)
6. Degree of formality (deference/solidarity)
7. Degree of distance (detachment/involvement)
8. Topic of interaction
9. Means of writing (pen/pencil, computer, dictation, shorthand)
10. Means of transmission (single page/book/read aloud/printed)

D. Awareness of sociolinguistic differences across languages and cultures

E. Self-awareness of roles of register and situational parameters

V. Bereiter and Scardamalia (1987)

Different from other attempts to modelling writing, Bereiter and Scardamalia (1987) proposed a two-model description of writing that identifies the distinction between knowledge telling and knowledge transforming, and they argued that novice writers tend to use the knowledge telling approach to writing, whereas advanced writers are prone to using the knowledge transforming approach.

Knowledge telling refers to a rather linear text generating process as it needs little planning activities ahead or revision processes. Bereiter and Scardamalia viewed this kind of writing as "natural" and "unproblematic", because nearly any fluent speaker of a language can do this if they have a grasp of the writing system (Weigle, 2002). Figure 2.5 shows the knowledge telling model. The writer starts from using a mental representation of the writing task to call up both content knowledge (what the writer knows about the topic) and discourse knowledge (the writer's knowledge about the type of discourse, for example, an argumentative essay or an expository essay). Topic and genre identifiers in the writing task are used to search one's memory for relevant content items by first constructing memory probes and then retrieving the content from these probes. The content items, that is, the writer's ideas, are subjected to tests of appropriateness (to check whether the ideas sound right or not, or whether or not the ideas support the writer's argument). If rejected, the writer goes back to the process of constructing memory probes again, while if accepted, these ideas are written down and then the cycle repeats itself based on the text written so far, rather than the previous mental representation, as a source of additional memory probes. The writing process of knowledge telling ends when the writer fails to find more appropriate ideas to write down.

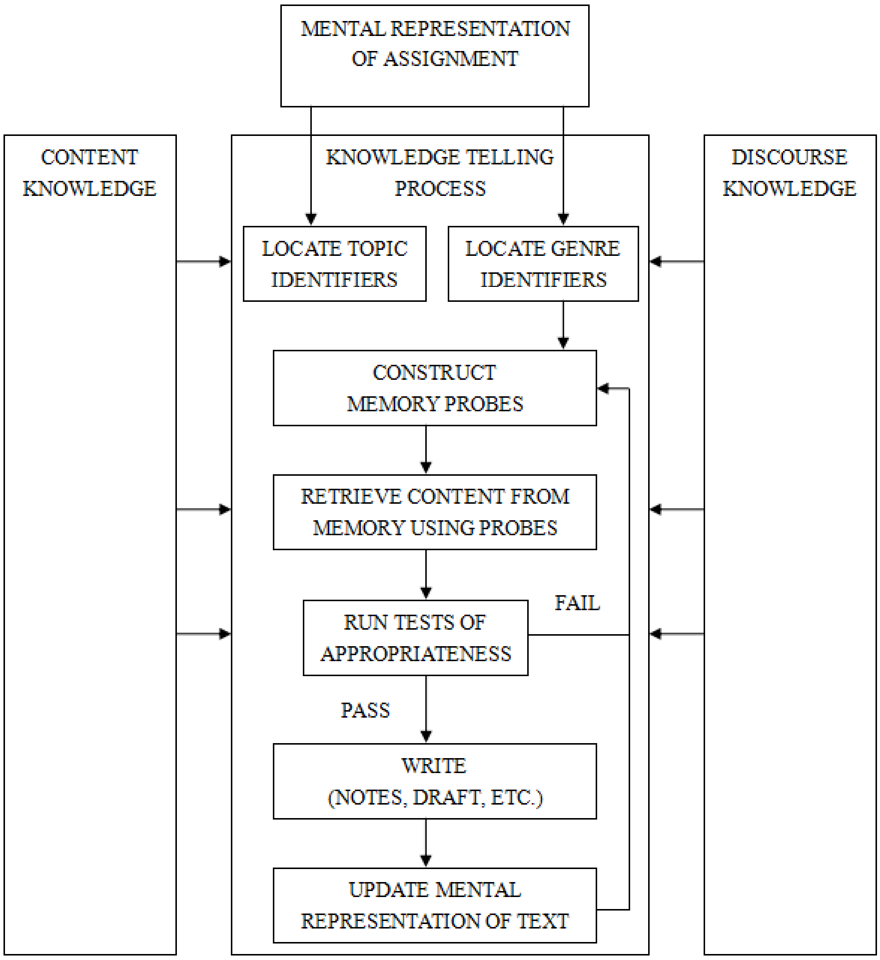

Figure 2.5: The knowledge-telling model (Bereiter and Scardamalia, 1987, p. 8)

On the other hand, advanced writers tend to use the **knowledge transforming** approach to writing, which asks for much more effort and skill. In knowledge transforming, the process of writing leads to generation of new ideas, and may enhance a writer's understanding of the subject knowledge or further develop their views on a particular topic. Figure 2.6 shows the knowledge transformation model. It starts with the problem analysis and goal setting, which then lead to two types of problem-solving activities: the content problem space and the rhetorical problem space. In the content problem space, the writer works on generating meaningful thoughts and knowledge, while in the rhetorical problem space, the issues of how to best achieve the goals of the writing task are dealt with. An attempt to address the content problem may lead the writer to a rhetorical problem, and vice versa. Bereiter and Scardamalia regraded this as “a two-way interaction between continuously developing knowledge and continuously developing text” (Bereiter and Scardamalia, 1987, p.12). Finally, the solutions to these two types of problem-solving activities become the inputs for knowledge-telling process, during which the actual text is generated.

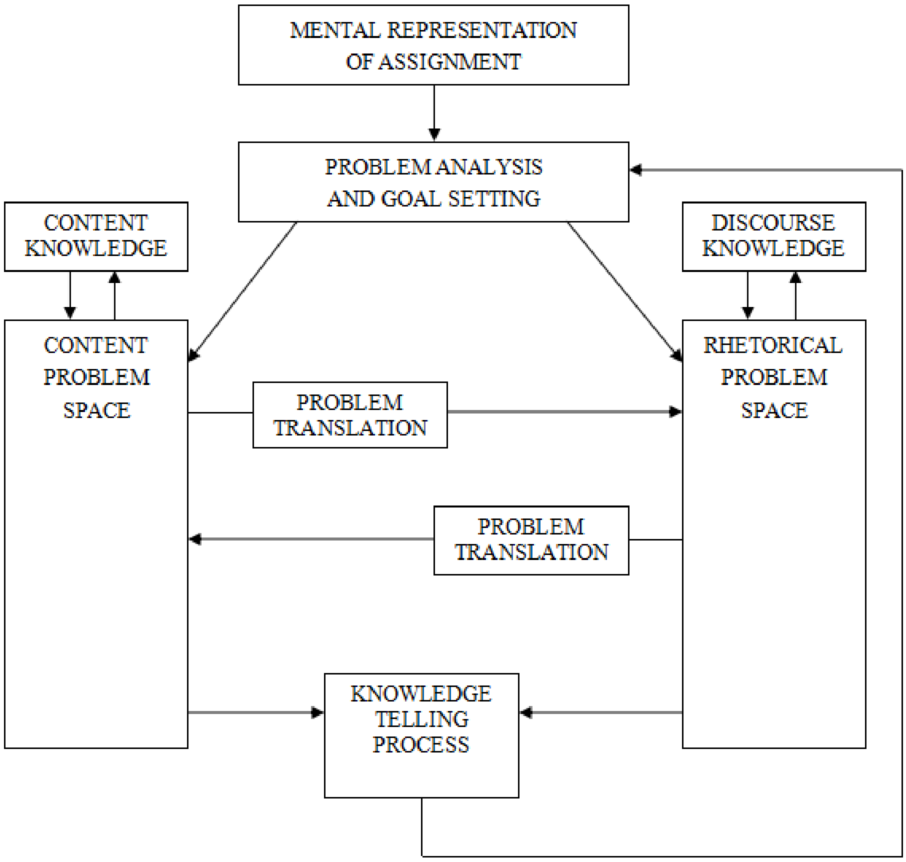

Figure 2.6: The knowledge-transforming model (Bereiter and Scardamalia, 1987, p. 12)

Bereiter and Scardamalia’s (1987) model provides an explanation for the distinction between the knowledge-telling writing approach typically used by unskilled writers, and the knowledge-transforming approach typically adopted by skilled writers. Although the model has limitations in a number of ways, for example, it did not provide an explanation for how a writer transits from knowledge telling to knowledge transforming, and did not account much for the interaction between the task and writers' cognitive processing activities, the distinction between knowledge telling and knowledge transforming has been a useful notion in writing assessment.

The above models (Bereiter and Scardamalia, 1987; Grabe and Kaplan, 1996; Hayes, 1996; Hayes, 2012; Hayes and Flower, 1980) that have been discussed so far are all important in shaping the current understanding of cognitive processes involved in writing. Despite these models' usefulness in modelling writing processes, they have several limitations in relation to the focus of the present study: (1) although they pointed out the cognitive processes which are affected by writers' own characteristics, task environment and socio-cognitive considerations, they did not provide enough explanation for the way in which the cognitive processes are influenced by these factors; (2) the cognitive processes proposed in these models are at a very general level, making them less effective to be used to define the processes and skills involved in a certain writing task, especially when looking at the cognitive validity of the task. The following models, which are based upon psycholinguistic theory, have the potential to address these two limitations.

VI. Kellogg (1996)

One of the most influential psycholinguistic models is Kellogg's (1996) model of working memory in writing, which distinguishes three systems of text production: formulation, execution and monitoring, and illustrates the significance of working memory in supporting different writing processes. Figure 2.7 shows the model. Each of the three systems involves two basic level processes: **formulation** consists of planning and translating; **execution** involves programming and executing; **monitoring** involves reading and editing. The flow of information is indicated by arrows in the figure between the six basic processes and the three systems. Thus, the output of planning will be input for translating, and the output of translating may then be sent to programming and then executing (handwriting, typing or dictating) in the execution system. It should be noted that outputs of formulation (planning and translating) may also feed into editing in the monitoring system prior to the execution. Potential corrections in each of the processes may thus be made before executing processes take place (Kellogg, 1996).

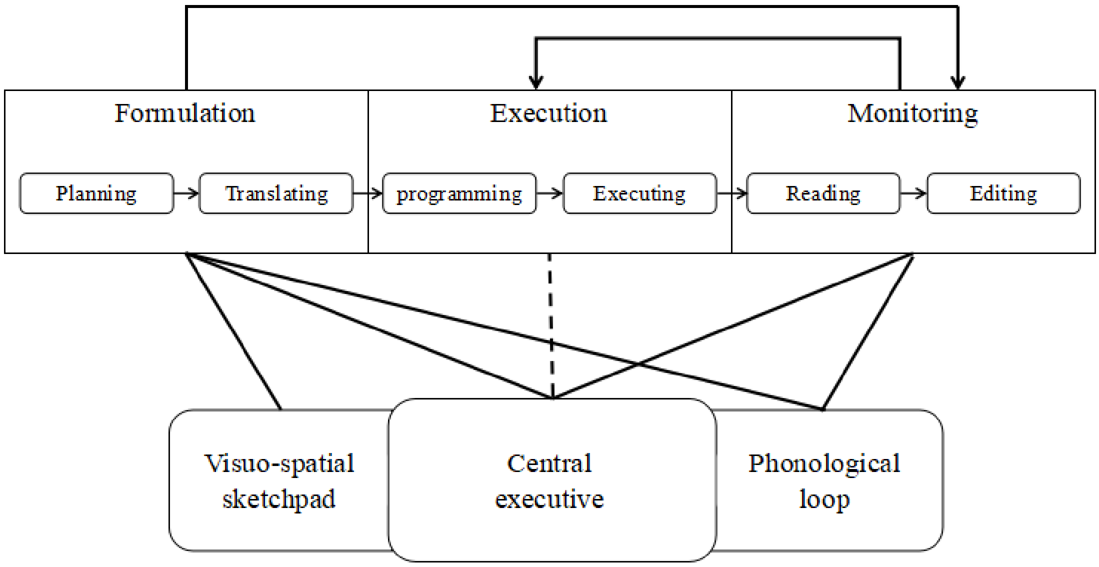

Figure 2.7: Kellogg's (1996) model of working memory in writing – adapted from Alamargot and Chanquoy (2001, p. 19)

For the role of working memory in writing, Kellogg proposed that each of the six basic-level processes draws upon different components of working memory, rather than seeing working memory as a unitary facility (Chan, 2013). The lines connecting each system of text production with components of working memory in Figure 2.7 indicate a demand by at least one basic process with a certain system (Kellogg, 1996). For example, the formulation system places major demands on the **executive control**, as well as the **visuospatial sketchpad** (stores and processes visual and spatial information) and the **phonological loop** (stores and processes auditory and verbal information); more specifically, the planning process in this system demands the resources of the spatial working memory and the central executive control, while translating theoretically demands the resources of the verbal working memory and the central executive control. Table 2.2 shows a detailed description of the resources of working memory used by the six basic-level processes of writing.

Table 2.2: The resources of working memory used by the individual writing process (Kellogg, Whiteford, Turner, Cahill & Mertens, 2013, p. 162)

Basic process	Working memory resource		
	Visuo-spatial sketchpad	Central Executive	Phonological loop
Planning	✓	✓	
Translating		✓	✓
Programming		✓	
Executing			
Reading		✓	✓
Editing		✓	

VII. Field (2004)

Field (2004) provided another influential account of the cognitive processes that a writer performs when engaged in the writing process. Much of his model is based upon Kellogg's (1996) model, which, as presented above, to some extent draws upon that of Hayes and Flower (1980). Also, it is influenced by Levelt's (1989) model of the speaking process. Field proposed that writing, as a productive skill, involves the processes of macro-planning, organisation, micro-planning, translation, execution, monitoring, and editing and revising.

Field's account of writing processes diverges from those of Kellogg and Levelt in one important aspect. Levelt proposed a stage of 'conceptualisation' in which a speaker selects the topic and retrieves their own world knowledge. Kellogg followed Levelt, identifying a similar stage termed as ‘planning’ in his model (see Figure 2.6), during which a writer generates and organises the ideas, and sets writing goals. Field, however, argues that writing differs from speaking because (1) it is not so time-constrainted as speaking in most real-life writing situations so that there is more conscious planning involved and (2) planning takes place at both text level (resulting in long-distance decisions about the readership, writing goals and genre, etc.) and utterance level (resulting in decisions about the text that is about to be produced). Therefore, Field divided ‘conceptualisation’ in Levelt’s (1989) model and ‘planning’ in Kellogg’s (1996) model into three processes: macro-planning, organisation and micro-planning. This is indeed consistent with the ‘planning’ process in the early Hayes and Flower (1980) model, which consists of subprocess of generating, organising and goal setting. The remaining cognitive processes in Field’s model are similar to those in Kellogg’s model. Also, Field explained the way in which higher- and lower-proficiency writers use these processes differently in the writing process.

The importance of the Field/Kellogg model is that it is not only more closely based on psycholinguistic theory, but it aims to provide a detailed account of the stages/processes that a writer may go through when producing a text, though these stages are “represented as interactive, with multiple possibilities of looping back” (Shaw and Weir, 2007, p.37).

VIII. Shaw and Weir (2007)

As discussed above, the Field/Kellogg model is significant in identifying different stages of processing, and the operations that take place within each stage, thus it provides a more accessible and detailed framework modelling writing processes than the other models presented so far. They are especially useful in language testing studies when there is need to identify which cognitive processes are relevant for test development and validity.

Based upon Field's (2004) and Kellogg's (1996) models, Shaw and Weir (2007) proposed five cognitive processes (see Figure 2.12 on page 48 for details) involved in writing: macro-planning, organisation, micro-planning, translation, and monitoring and revising. They are considered the most relevant to the investigation of the cognitive validity of a writing task. Shaw and Weir argue that a valid writing task should elicit from test-takers these five cognitive processes, which are commonly involved in real-life writing situations. In their work, they also evaluated how these core processes of writing have been elicited by the Cambridge English Language Assessment writing tests across different levels (Shaw and Weir, 2007). Their approach to evaluating the cognitive validity of a writing test through investigation of the five proposed processes has, to some extent, provided principles for validation research in L2 writing assessment. The approach and all the cognitive processes will be described in detail in Section 2.4.

In summary, the writing models discussed above (Bereiter and Scardamalia, 1987; Field, 2004; Grabe and Kaplan, 1996; Hayes, 1996, 2012; Hayes and Flower, 1980; Kellogg, 1994, 1996; Shaw and Weir, 2007) have shaped much of our current understanding of cognitive processing involved in writing. For example, writing is not a linear act, but involves multiple recursions of processes such as planning (at both macro- and micro-levels), organising, executing/translating, monitoring and revising; these processes are often overlapping with each other at particular writing stage, and looping back and forth for different purposes in writing. Also, writing should not be considered as an isolated act, but is greatly influenced by writers' internal traits such as working memory capacity and their store of long-term knowledge (e.g., linguistic, discourse and content knowledge), as well as external variables such as task settings and social variables (e.g., readership).

Despite their importance, the process of integrating reading with writing has largely been under-represented in these models. Although several models (e.g., Hayes, 1996) have pointed out the essential role that reading plays in writing process, issues such as what types of reading activities are involved during writing, when and how reading processes interact with other processes in writing remain largely unclear. With growing interest in using reading-to-write tasks to assess test-takers' language skills, there is an urgent need to look into how reading is integrated with writing to produce a meaningful text while performing such tasks. In the following subsection, literature on the integration of reading and writing processes will be reviewed.

2.2.2 Interaction of reading and writing

It is beyond the scope of this study to thoroughly review the existing models of reading though the literature on L1 and L2 reading processes is well established (see Khalifa and Weir, 2009 for a detailed review of different reading models). Instead, Khalifa and Weir's (2009) model of cognitive processing in reading is reviewed to gain some insights into the major stages and processes involved in reading.

I. Khalifa and Weir (2009)

Based on the socio-cognitive approach to validation of language tests (Weir, 2005; O'Sullivan and Weir, 2011), Khalifa and Weir (2009) proposed a cognitive processing model of reading, which integrates test-takers' cognitive and metacognitive processes with their language knowledge and general knowledge of the world to illustrate the mechanism of reading for comprehension.

Figure 2.8 (displayed on the next page) shows Khalifa and Weir's (2009) reading model. It has three major components: metacognitive activity (the left column), the central processing core (the middle column), and the knowledge base (the right column), all of which contain a variety of sub-processes (Brunfaut and McCray, 2015). Metacognitive activity involves a goal setter, a monitor, and a remediator. When setting reading goals (goal setter), a reader determines the type(s) of reading needed to fulfill certain reading purposes, for example, a careful reading approach may be adopted when the reader needs to comprehend the majority of information in a text. During reading, readers constantly check (monitor) their reading processes to decide if they are progressing consistently with the goals generated, and make remediations (remediator) when necessary.

The central processing core consists of a hierarchical system of eight cognitive processes that work together to carry out reading activities. They can be grouped into two categories of processing based on their demands on cognitive resources: (1) the lower-level processes, including word recognition, lexical access, syntactic parsing, and establishing propositional meaning; and (2) the higher-level processes, which include inferencing, building a mental model, creating a text level representation, and creating an intertextual representation. The difference between higher- and lower-level processes, according to is that the lower-level processes "can become strongly automatised and not subject to conscious processing" (Brunfaut and McCray, 2015, p.6). Skilled readers tend to have high automaticity of lower-level processes presented above and thus are able to use higher-level processes more freely while reading (Field, 2004).

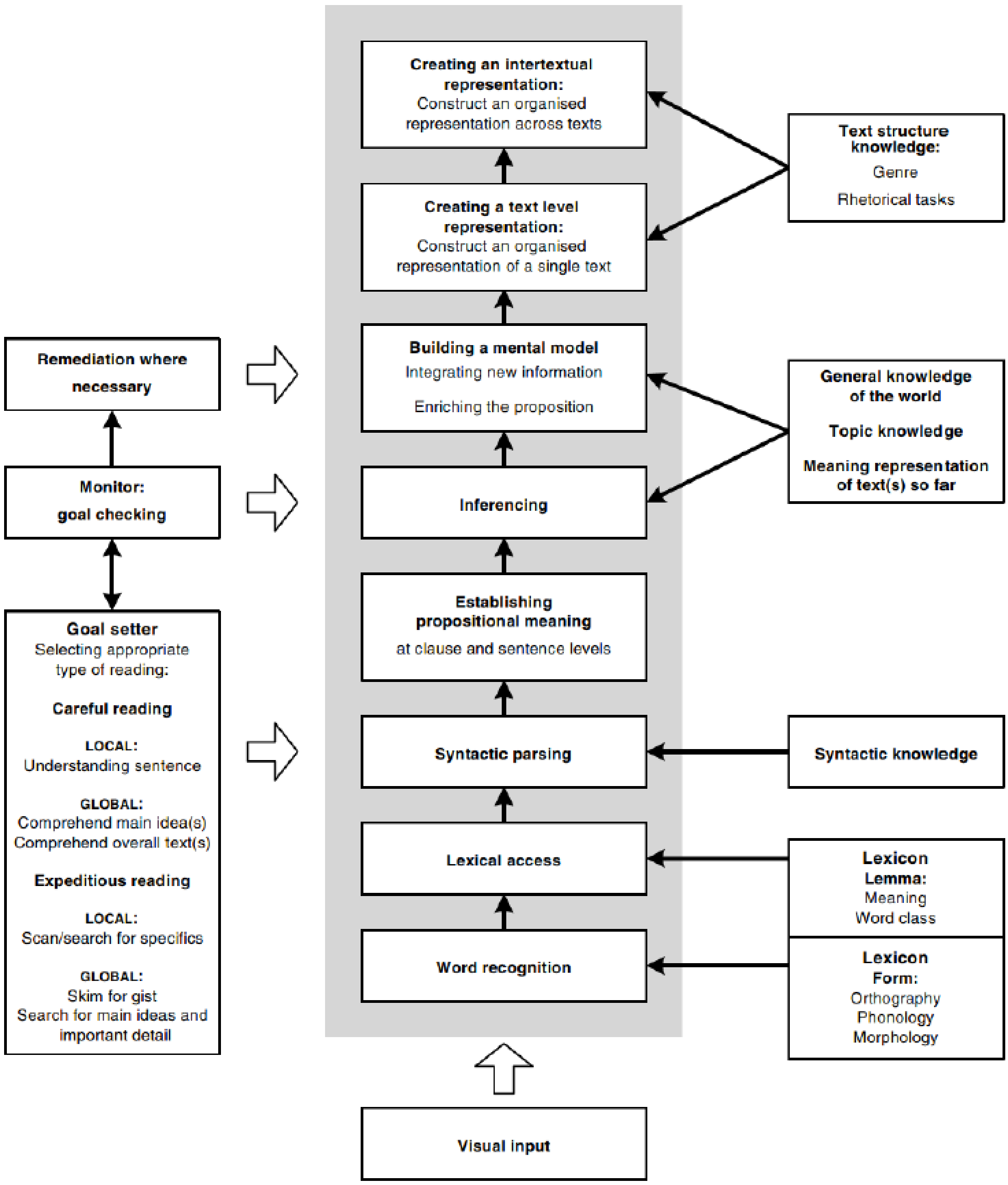

Figure 2.8: Khalifa and Weir's model (2009, p. 43) of cognitive processing in reading

While processing the text, readers may resort to a variety of knowledge sources, as represented in the right column (knowledge base) in Figure 2.8, linking to relevant cognitive processes in the central processing core. These knowledge sources include: lexical knowledge of a word's orthography, phonology, and morphology; lexical knowledge of a word's meaning and its word class; syntactic knowledge of the target language; general knowledge of the world, topic knowledge about the subject of the text being read, and knowledge of the meaning of the text produced so far; and finally, text structure knowledge, that is, knowledge of genre of the text or rhetorical tasks.

As presented above, Khalifa and Weir's (2009) model provided a detail account of the stages/processes that the reader may go through when reading for comprehension. However, like many other models of reading per se in the literature, it has little discussion about how these reading processes may fit into a model of writing.

II. Discourse synthesis process

Despite the lack of frameworks modelling reading-to-write processes, some studies have investigated writers' processes involved in writing from sources (see Section 2.3 for a review of these studies). One important notion that emerged from these studies is the concept of *discourse synthesis* (Ackerman, 1991; Greene, 1993; Lenski, 1998; Marsella, Hilgers and McLaren, 1992). Spivey and King (1989, p.11) defined discourse synthesis as follows:

> *some hybrid reading-to-write tasks involve discourse synthesis, a process in which readers (writers) read multiple texts on a topic and synthesize them. They select content from the composite offered by the sources – content that varies in its importance. They organize the content, often having to supply a new organizational structure. And they connect it by providing links between related ideas that may have been drawn from multiple sources.*

Research by Spivey and her colleagues (Mathison and Spivey, 1993; Spivey, 1984, 1990, 1997, 2001; Spivey and King, 1989) has shaped the notion of discourse synthesis in writing from external reading materials. The findings of these studies revealed that writers utilise a meaning-making process in reading-to-write tasks, by transforming a new representation of the meaning from source materials to their own text through three cognitive processes: (1) *organising* ideas as they read and write; (2) *selecting* relevant ideas or information from source texts; and (3) *connecting* ideas selected from different source texts and generating links between them. As a result, the reading process is more centrally situated in the discourse synthesis process, and reading and writing are integrated within each of the three processes. The findings indicated that reading-to-write tasks place higher cognitive demands on test-takers than reading or writing tasks alone, and organisation, selection and connection are "the very basis of reading, writing, and learning in almost any domain knowledge" (Spivey, 1997, p. 191). The discourse synthesis process will be discussed in more detail in Section 2.4.

In summary, in order to achieve an understanding of the reading-to-write process, Section 2.2 has reviewed several writing models (Bereiter and Scardamalia, 1987; Field, 2004; Grabe and Kaplan, 1996; Hayes, 1996, 2012; Hayes and Flower, 1980; Kellogg, 1994, 1996; Shaw and Weir, 2007) that are influential in the current literature; a cognitive processing model of reading proposed by Khalifa and Weir (2009) was also outlined in order to shed some light on the reading process; finally, the notion of discourse synthesis (Spivey and King, 1989) was looked at to provide insights into the interactive nature of reading and writing in reading-to-write tasks. In the next section, empirical research on integrated writing tasks will be reviewed.

2.3 Research on integrated writing tasks

Over the past decades, research studies on integrated writing tasks have blossomed, delving mainly into four lines of research: (1) studies that compared test-takers' performance on independent and integrated writing tasks; (2) studies that investigated discourse features of written products in integrated writing tasks; (3) studies that examined test-takers' processes while completing the integrated writing tasks; and (4) studies that looked into the source use in integrated writing tasks.

2.3.1 Comparison studies

One of the initial areas explored in comparison studies was the relationships between test-takers' performance on independent and integrated writing tasks. Several studies correlated scores from the two types of writing tasks and found that they were similar and achieved sizable correlations (Brown, Hilgers and Marsella, 1991; Gebril, 2006, 2009; Lee and Kantor, 2005; Watanabe, 2001).

In Watanabe's (2001) study, he investigated test-takers' (L2 writers at the University of Hawaii) scores on a source-based writing task (opinion-writing as follow-up on the reading input), which he correlated with an independent writing task, and discovered a medium positive correlation (r=.62) between these two tasks. In contrast, Gebril (2006) and Lee and Kantor (2005) found much higher correlations between independent and integrated writing tasks, with correlation coefficients of .93 and above. The differences between these findings may be due to the uses of different scoring criteria and the different proficiency levels of participants. It should be noted that, while these two types of writing tasks both seek to evaluate test-takers' written products, differences may still lie in the construct they are designed to elicit. Most importantly, integrated writing tasks include elicitation of test-takers' reading-writing skills as well as their ability to integrate source texts into their own writing.

Although correlation studies have shown the similarities between independent and integrated writing tasks, differences between these two task types have also been identified through investigations of discourse features in the written products (Cumming et al., 2005; Gebril and Plakans, 2009; Guo, Crossley and McNamara, 2013; Lewkowicz, 1994) and test-takers' processes during task completion (Ascensión, 2005;

Esmaeili, 2002; Plakans, 2008, 2009b; Yang, 2009).

In a large-scale study of prototype tasks piloted for the Internet-based Test of English as a Foreign Language (TOEFL iBT) writing section, Cumming et al. (2005) examined writing features of three tasks: a listening-writing task (writing in response to a listening passage), a reading-writing task (writing in response to a reading passage), and an independent writing task. They discovered significant differences across tasks in areas such as lexical sophistication (in terms of word length and different words produced), syntactic complexity (in terms of words per *T*-unit and clauses per *T*-unit) and argument structure (in terms of propositions, claims, data, warrants and oppositions). More specifically, they argued that test-takers in the integrated tasks, compared to the independent ones, tended "to write briefer compositions, to use longer words, to use a wide range of words, to write longer clauses and more clauses, to write less argumentatively oriented texts, to indicate sources of information other than oneself, and to paraphrase, repeat verbatim, or summarize source information more than to make declarations based on personal knowledge" (Cumming et al., 2005, p. 32).

Lewkowicz (1994) compared the essays produced by two groups of L2 writers (first-year undergraduates at the University of Hong Kong), one which was provided with background reading materials for the writing, and the other which had not been given the materials. She identified a difference in the number of points made in essays produced in the two types of tasks, with more points introduced in the integrated writing tasks. And she argued that although more points were included in the integrated task essays, the lengths of these essays were not longer, thus each point was less developed than those in the independent task essays. More recently, Guo et al. (2013) examined the written products that 240 L2 writers produced in two tasks: an independent writing task, and an integrated writing task (writing in response to a listening passage and a reading passage). Linguistic features related to lexical sophistication, syntactic complexity and cohesion were investigated. Results showed similarities in two features across tasks: essay length and past participle verb usage (in passive voice); other features did not show correlations across the two types of writing tasks.

In summary, the results of correlation studies have indicated that independent and integrated writing tasks are strongly correlated with each other, however, differences emerge when further investigations of discourse features and writing processes are performed. One obvious and critical difference between these two task types is the inclusion of source materials in integrated writing task, which elicits more discourse synthesis skills as described in Spivey's model (1990, 1997). These findings provide evidence for supporting the use of both types of writing tasks to measure the writing skills or for selecting the one that is most appropriate for the construct to be measured (Plakans, 2015).

2.3.2 Discourse features

In addition to the comparison studies presented above, another line of research has focused on features of the written products in integrated writing tasks across different proficiency levels (Baba, 2009; Cumming et al., 2005, 2006; Gebril and Plakans, 2009, 2013, 2016; Plakans and Gebril, 2017).

Earlier research on independent writing tasks investigated writing features in terms of fluency, syntactic complexity, grammatical accuracy, and vocabulary richness or sophistication (Jarvis, Grant, Bikowski, & Ferris 2003; Ortega, 2003; Sasaki, 2000). With respect to integrated writing tasks, several studies have shown that fluency consistently increases with proficiency levels (Cumming et al., 2005; Gebril and Plakans, 2009). In terms of syntactic complexity, Cumming et al. found significant differences in the number of words per T-unit (defined as the smallest unit of a sentence that can stand alone grammatically) across proficiency levels, but none when measuring the number of clauses per T-unit. In a similar study, Gebril and Plakans (2009) analysed the discourse features of 131 English essays written by Arabic speakers in a source-based writing task (an argumentative essay prompt with two short reading passages), and they found no significant differences in the number of T-units per sentence. Grammatical accuracy in written products has also been investigated (Cumming et al., 2005; Gebril and Plakans, 2009, 2013). In these studies, grammatical accuracy was rated and assigned a holistic score. It was found to differ significantly across proficiency levels, but in post-hoc comparison analysis, Gebril and Plakans discovered that it was only the lowest scoring group that held significant differences, while the upper levels did not differ significantly in grammatical accuracy (Plakans, 2015).

There is relatively little research on lexical diversity in integrated writing tasks (Baba, 2009; Cumming et al., 2005). Baba (2009) investigated the relationships between various aspects of lexical proficiency (including lexical diversity) and the quality of written products in a summary writing task, and identified a non-linear correlation between lexical diversity and the quality of summaries. She argued that this variation in lexical diversity may be because of participants' heavy reliance on source texts during writing, and due to this non-linear relationship, lexical diversity did not contribute much to the variability of scores. Cumming et al. also examined the lexical features in their 2005 study. Their analysis of lexical features was mainly on two measures: average word length and type/token ratio. Average word length is a typical index for measuring lexical sophistication, while type/token ratio measures lexical diversity. Their results showed that the two integrated tasks yielded higher average word length than independent tasks. Similarly, higher type/token ratio results were reported in integrated tasks. The findings also revealed statistically significant differences in the two lexical features across different proficiency levels.

A more recent study on discourse features was conducted by Plakans and Gebril (2017), in which they investigated rhetorical structure in a TOEFL integrated writing task (writing in response to a reading passage and a listening passage). Three features were analysed: organisational patterns, coherence and cohesion. The results showed that organisation quality increased across proficiency levels, but with no difference between upper levels. A similar pattern was found in coherence quality, which increased with score levels. However, cohesion features did not yield significant differences across different score levels.

This line of research contributes to our knowledge of integrated writing tasks, from the perspective of discourse features of the written products. Although discourse features are not the focus of the current study, it is helpful, to some extent, in understanding test-takers' thought processes while completing a reading-to-write task, that is, linguistic features of written composition might be somewhat connected to the integration of source materials, particularly with respect to the integration of more complex lexis.

2.3.3 Process studies

A third area of research on integrated writing tasks has explored test-takers' composing processes in responding to the tasks. These studies have concluded that integrated writing tasks involve a different set of processes (for example, discourse synthesis) which is distinct from those required to complete independent writing tasks (Chan, 2013; Plakans, 2008, 2009), and that reading skills play important roles in integrated reading-to-write tasks (Ascensión, 2005, 2008; Esmaeili, 2002; Plakans, 2009a).

In Esmaeili's (2002) study, he asked 34 ESL adult learners with intermediate levels of English proficiency to complete reading and writing tasks in two conditions: one includes both reading and writing tasks (thematically related to each other) and the other with unrelated tasks. Retrospective interviews were conducted and a checklist of the writing strategies used while writing was employed to understand the reading-writing process. The findings showed that these test-takers performed significantly better in the writing when reading and writing tasks were thematically related, concluding that reading played a critical role in test-takers' writing processes and "one can hardly view reading and writing as stand-alone skills" (p. 615). Ascensión (2005, 2008) conducted a validation study of two integrated writing tasks (a summary task and a reflective essay) through think-aloud protocols, which she coded on the basis of Spivey's (1984, 1987) discourse synthesis framework. The results confirmed the existence of a discourse synthesis process as an underlying construct in integrated writing tasks, and revealed that reflective essay tasks involved more cognitive operations than summary tasks did. Plakans' (2008) study, as mentioned earlier, compared test-takers' processes while completing an independent writing task and a source-based writing task (opinion-writing as follow-up on a reading passage) through think-aloud protocols and post-protocol interviews. She found that more pre-planning prior to composing was involved in writing-only tasks, but there was a greater difference in processes across writers in reading-to-write tasks.

Discourse synthesis, as discussed earlier, is arguably the most unique and essential process of the reading-to-write construct. This has been demonstrated in Plakans' (2009b) investigation of six L2 writers' discourse synthesis processes using think-aloud protocols. Her findings showed that several writers approached the tasks using discourse synthesis processes, with varying degrees of using organising, selecting and connecting subprocesses among these writers. Another influential study on discourse synthesis was conducted by Chan (2013), in which she developed a reading-to-write process questionnaire based upon previous models of reading and writing processes (this questionnaire will be described in detail in Section 3.2.2) and trialled it in a pilot study with 99 participants. Chan then used the validated questionnaire to investigate 219 students' cognitive processes while completing four reading-to-write tasks (opinion-writing as follow-up on the reading input) under real-life and test conditions. The results of exploratory factor analysis confirmed the underlying construct of different cognitive processes that Chan proposed as core processes in a reading-to-write task. Her findings also revealed that higher-scoring students reported more use of most of the specified cognitive processes (for example, task representation, connecting and generating) than lower-scoring students. This study will be discussed in greater detail in later sections as her reading-to-process questionnaire is also used in the current study to collect data on participants' cognitive processes.

Although there is relatively less research on test-takers' cognitive processes while completing integrated writing tasks, the above studies have laid a foundation for understanding this unique set of processes, though they used a restricted set of methods. The current study contributes to the knowledge of reading-to-write processes and more research on this aspect will be reviewed in Section 2.4.

2.3.4 Source use

A fourth line of research on integrated writing tasks has investigated the use of sources through examining test-takers' written products. Two topics in this area have received considerable attention: integration style and verbatim copying (Campbell, 1990; Cumming et al., 2005; Currie, 1998; Gebril and Plakans, 2009; Johns and Mayes, 1990;

Pennycook, 1996; Shi, 2004; Watanabe, 2001).

Watanabe (2001) identified two types of source use (explicit and implicit source use) in 47 reading-to-write responses, finding that writers tended to use quotation (explicit source use) most often, with some instances of partial paraphrasing and summarising (implicit source use). Similarly, Gebril and Plakans (2009) coded 145 English essays written by Arabic speakers and found that, overall, higher-scoring students used source texts more than lower-scoring students. Cumming et al. (2005) also discovered differences in source use across different score levels. The most proficient writers tended to summarise more than writers at other levels; writers at intermediate levels paraphrased and plagiarised more than writers at either high or low proficiency levels; and the least scoring writers tended to summarise, paraphrase and copy less than writers at all other levels. Cumming et al. (2005) explained that this may be due to the fact that low proficiency writers were not able to understand source texts well enough even to perform simple direct copying.

The other topic, verbatim use of source text, has been investigated extensively in L2 writing research (Asabi, Akbari and Graves, 2006; Currie, 1998; Johns and Mayes, 1990; Shi, 2004). In an early study on verbatim source use, Johns and Mayes (1990) examined direct copying in 80 writing response of L2 writers at two proficiency levels on a summary task (summary-writing of the reading input). The findings showed that the lower-proficiency writers tended to copy more directly, but there was no significant difference in "correct paraphrasing" between two groups. Interestingly, the higher-proficiency writers also combined idea units from the source texts more and were likely to distort some of these ideas. Shi (2004) compared the written products of two types of writing tasks (an opinion-writing task and a summary task based on the same reading input) produced by two groups of writers: native and non-native English writers. The findings revealed that L2 writers borrowed more from source texts than L1 writers, and that the summary task elicited more verbatim use of source texts than the opinion task. Similarly, Campbell's (1990) study examined the essays produced by native and non-native English-speaking university students in a source-based writing task (a book chapter was provided as the background reading text), and found that L2 writers cited the source texts considerably more than L1 writers.

These studies have provided a solid foundation of understanding source use in integrated writing tasks. It is clear that source use may vary across proficiency level, and that the type of text may influence the manner in which it is used. One topic, however, that has not received much attention is the role of multiple sources. As integrated tasks normally include more than one source text, how writers navigate across these texts remains under-researched. Also, most studies have investigated source use through examining test-takers' written products, very few studies looked at test-takers' online source use processes.

2.3.5 Methods of previous process studies

As the focus of this study is on reading-to-write processes, methods of previous research on processes are summarised here to shed some light on potential methods to investigate test-takers' processes while completing the TBEM-8 reading-to-write task.

It is important to note that the most significant obstacle to examining cognitive processes is that they cannot be observed directly. Previous process studies have investigated reading-to-write processes through two main approaches: self-report or observation. As presented earlier, most process studies used self-report methods in which participants are asked to report their cognitive processing activities either concurrently (for example, think-aloud protocols) or retrospectively (for example, retrospective interviews). One major concern of using concurrent self-report methods is the extent of reactivity and potential disruption imposed on test-takers' actual cognitive processes (Stratman and Hamp-Lyons, 1994). This issue needs to be considered carefully in particular when the test examined is highly demanding in cognitive resources (for example, reading-to-write tasks that involve integration of at least two skills). Retrospective self-reporting methods do not interfere with participants' actual processes, however, issues such as memory decay and over-reporting may also be detrimental to the accuracy of data collected (Harwood, 2009).

These two types of self-report methods rely largely on participants' perceptions of their cognitive processes, and on their ability to report or recall the processes (Smagorinsky, 1994). Meanwhile, as there are time costs in collecting and analysing think-aloud or interview protocols, a relatively small number of participants are usually involved in these studies (questionnaire is also a kind of self-reporting technique that can be used in large-scale studies). Other researchers have investigated test-takers' cognitive processes by using direct observation methods such as video recording (Bosher, 1998), and screen capture software (Chan, 2011). These studies allow participants to focus on their actual cognitive processing, with minimum interruption. However, observations are essentially an "etic" method (based on the researcher's interpretation of what he/she observes), and if it is not triangulated with participants' perceptions of their cognitive processing then important information may be lost.

In summary, there are pros and cons of using each method independently to investigate test-takers' reading-to-write processes during task completion, however, studies that used a combination of these methods have been scarce. More considerations about the methods to be used in the current study will be further discussed in the next chapter of methodology (see Section 3.2 for details).

2.4 Cognitive validity considerations for reading-to-write tasks

In order to establish cognitive validity evidence for the TBEM-8 reading-to-write task, several validation frameworks are first reviewed in Section 2.4.1 and a set of cognitive processes to be investigated in this study are proposed and described in Section 2.4.2.

2.4.1 Validation in language testing

Validity refers to the degree to which a test measures what it is supposed to measure (Cronbach, 1988; Lado, 1961), and is evaluated through observation of evidence pertaining to different categories of validity. Validity theory has undergone rapid developments in the past 50 years. One of the most important transition periods of validity theory in language testing was in early 1990s when Bachman first introduced Messick's (1989) unified (unitary) validity theory into the field, which ended the dominance of the early "Trinitarian doctrine", that is, content, construct and criterion validity (Guion, 1988), and a variety of validity classifications. Since then, the research horizon of validity has been significantly expanded, incorporating more practical studies such as test use and social consequences.

Despite its usefulness in test validation, the notion of validity as a unitary concept still has several problems, from both theoretical and practical perspectives, among which the most troublesome one is – how to develop a feasible and operable validation framework to validate language tests within the theory of this unified validity. An early attempt to address this issue was that of Kane (1992), in which he proposed an argument-based approach to test validation and developed the notion of the "interpretative argument" as providing a framework to gather and disseminate evidence for supporting intended score interpretation (Bachman, 2005). An interpretative argument consists of inferences and assumptions that need to be supported by relevant evidence; Kane (1992) put forward three criteria to evaluate the inferences made on the basis of an interpretative argument: (1) clarity of assumption; (2) coherence of argument; and (3) plausibility of assumptions.

Based on Kane's (1992) work, Kane, Crooks and Cohen (1999) provided a detailed explanation of an interpretative argument that links observations to interpretations. The argument is composed of four parts and each part is linked to the next by an inference (see Figure 2.9). The first inference, "scoring", is from a performance to an observed

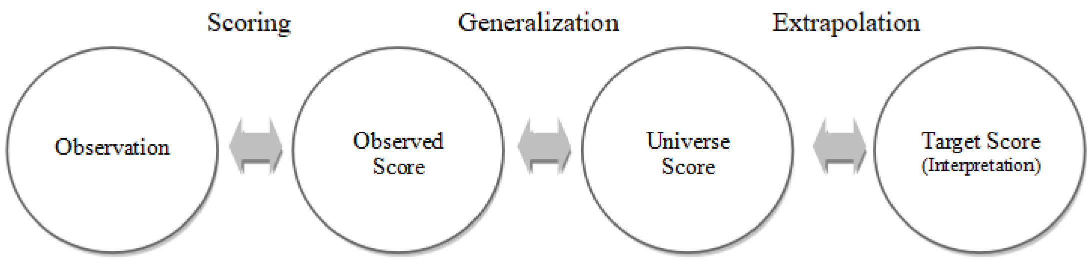

Figure 2.9: Links in an interpretative argument – adapted from Kane, Crooks and Cohen (1999, p.9) (Bachman, 2005)

score, and is based on two assumptions: (1) the scoring procedures are appropriate and consistent; and (2) the observed performance occurred under conditions consistent with the intended score interpretation. The second inference, "generalization", is from the observed score on a particular measure to a universe score, and is based on the assumptions of measurement theory, for example, generalisability theory. The third inference, "extrapolation", is from a universe score to a target score, which is, as Kane et. al described, an interpretation of what a test-taker knows or can do. This inference is based on the claims in an interpretative argument and the collected evidence supporting these claims (Bachman, 2005).

From a different perspective (test design and development), Mislevy, Steinberg and Almond (2002, 2003) also drew on the theory of evidentiary reasoning, and developed a set of steps and procedures called "evidence-centered design" (ECD) to guide test development. The key element in the ECD is what they referred to as an "evidentiary argument", which logically connects the claims (interpretations to be made) with the evidence that needs to be collected to support these claims. This evidentiary argument has some practical benefits in terms of guiding the development of tests and scoring rubrics, as well as facilitating the gathering of evidence for validity and generalisation (Mislevy et al., 2002).

Mislevy et al. (2003) claimed that their validity arguments are based on Toulmin's (2003) argument structure (see Figure 2.10), which consists of several essential elements as follows:

1. *Claim*: this is the interpretation to be made about what a test-taker knows or can do, based on the analysis of data.

2. *Data*: it consists of "information on which the claim is based" (Toulmin, 2003, p.90); in language testing, these are the responses of test-takers to certain tasks (for example, multiple choice questions); the link between the data and the claim represents an inference, which is justified through a warrant.

3. *Warrant*: warrants are propositions that help to justify the inference made from data to claim; the warrant is based on backing.

4. *Backing*: the backing consists of "other assurances, without which the warrants themselves would possess neither authority nor currency" (Toulmin, 2003, p. 96); in language testing, most of these backings come from theory, evidence collected during the validation process, or prior experience.

5. *Rebuttal*: rebuttals are counterclaims to an intended inference; in a validity argument, counterclaims correspond to potential sources of invalidity that may result in

“construct irrelevant variance” and “construct underrepresentation” (Messick, 1989).

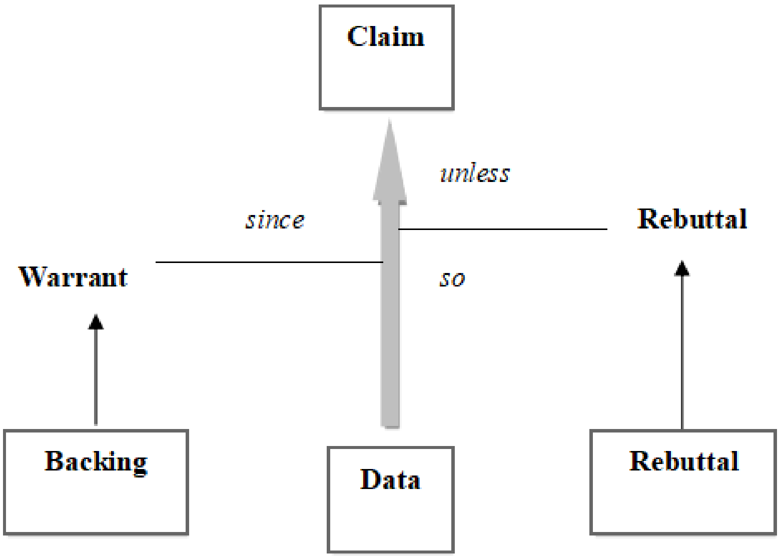

Figure 2.10: Toulmin diagram of the structure of arguments – adapted from Mislevy et al. (2003, p. 11) (Bachman, 2005)

These argument-based approaches to language test validation (Kane, 1992, 2001, 2002; Kane, Crooks and Cohen, 1999; Mislevy et al., 2002, 2003) provide a logic and set of procedures for articulating claims and for collecting evidence to support these claims, however, as Bachman stated (2005, p.4), “these argument-based approaches have focused primarily on claims about the interpretation of test scores, and have not, until very recently, begun to address issues of test use and the consequences of test use”. Based on the research of argument-based approaches to validation, Bachman (2005) put forward his structure of an assessment argument.

According to Bachman (see Figure 2.11), an assessment use argument consists of two parts: an assessment utilization argument, which links an interpretation to a decision, and an assessment validity argument, linking assessment performance to an interpretation. Thus the whole validation process can be divided into two stages, and in each stage, the arguments need to be justified by offering both backings and rebuttals to the inferences or decisions (for an updated version of the assessment use argument see Bachman and Palmer, 2010).

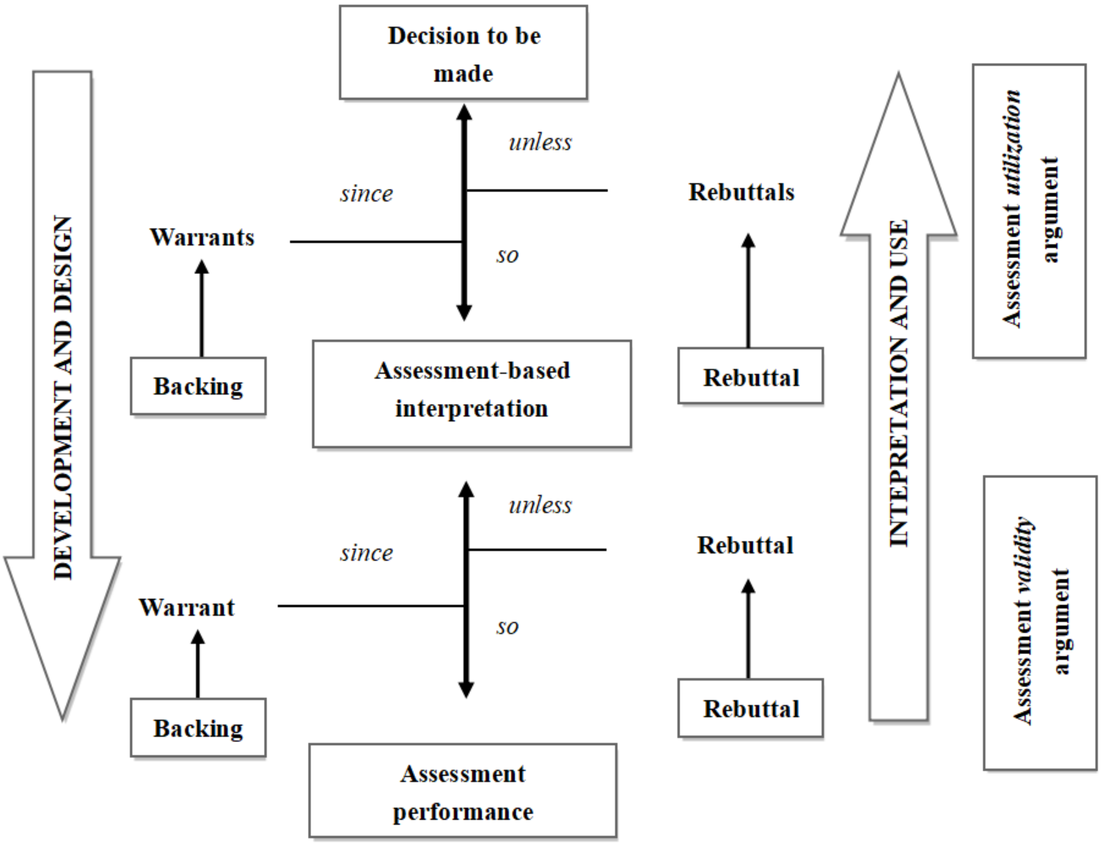

Figure 2.11: The structure of an assessment argument (Bachman, 2005, p. 25)

Although the argument-based approaches to validation provide a logic and scientific way to guide the validation and test development process, still, they are too general to be conducted with research instruments in practice, in other words, they lack operability. Weir (2005) made a notable attempt to address this issue by proposing an evidence-based "socio-cognitive validation framework". Figure 2.12 shows an updated version of this framework tailored towards the validation of writing tests. Shaw and Weir's (2007) framework consists of six aspects of validity: test-taker characteristics, cognitive/theory-based validity, context validity, scoring validity, consequential validity, and criterion-related validity. Arrows in this framework indicate the principal direction(s) of any hypothesised relationships, and the timeline runs from top to bottom, that is, "before the test is finalised, then administered and finally what happens after the test event" (Weir, 2005, p.43).

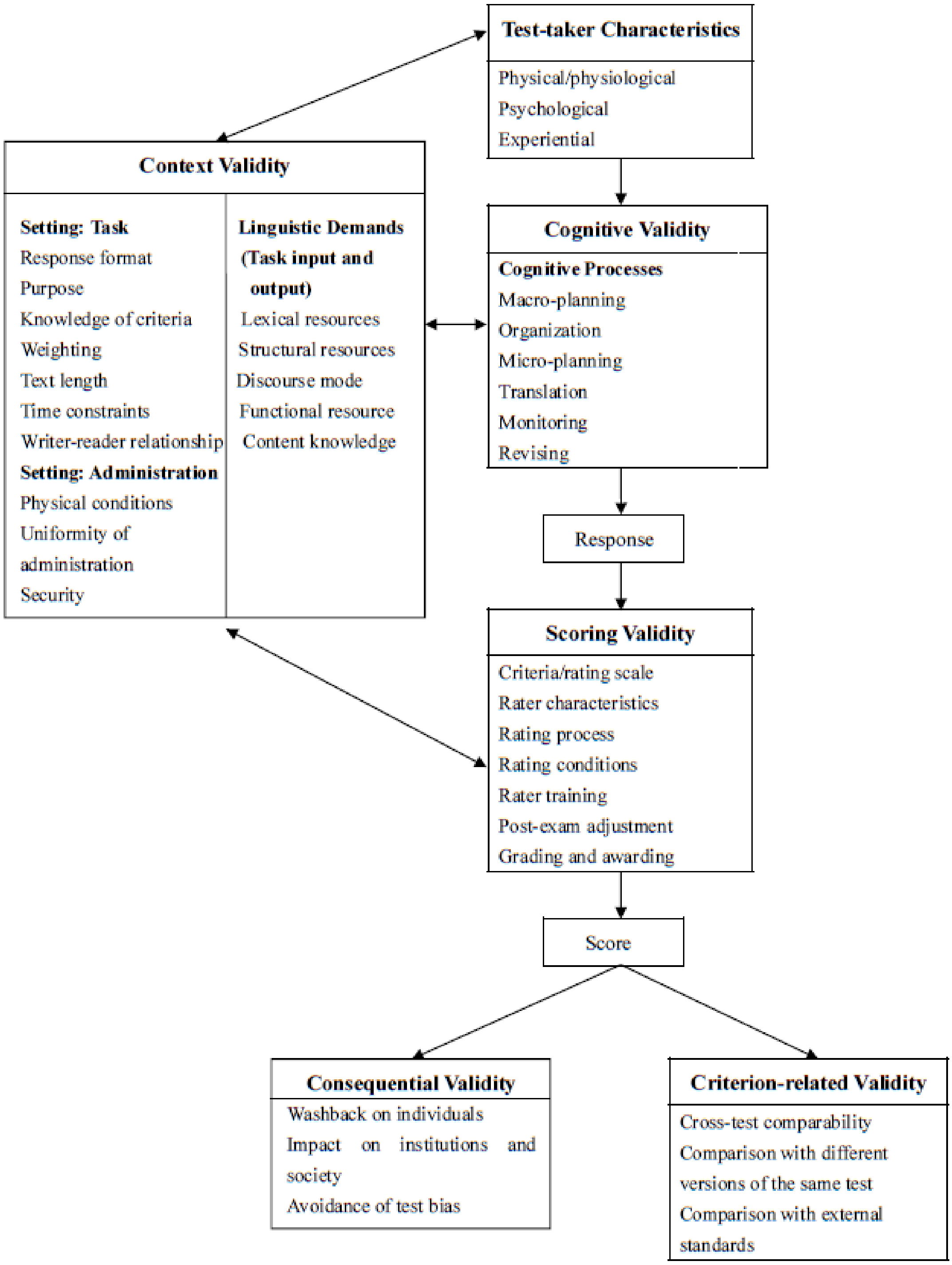

Figure 2.12: A socio-cognitive framework for validating writing tests (Shaw and Weir, 2007, p.4)

Test-taker characteristics can be divided into three main categories: (1) physical/physiological characteristics such as test-takers' age, sex, and partial sightedness; (2) psychological characteristics, for example, personality, memory, cognitive style; and (3) experiential characteristics, for example, education, and examination preparedness. Test-takers' characteristics are directly connected to the context and cognitive validity because these characteristics will directly impact on the way test-takers process the test task in a certain context.

Unlike previous validation frameworks, which see construct (content) validity as a uniform concept, Shaw and Weir's (2007) framework illustrated the abstract notion of 'construct' as consisting of context and cognitive components in order to provide stronger evidence for construct validity (Chan, 2013). Context validity considers the social and cultural contexts in which a test task is performed, for example, for a writing task (see Figure 2.12), context validity addresses the appropriateness of the task setting (for example, text length, time constraints, writer-reader relationship) and the actual test administration (for example, physical conditions and uniformity of administration), and the linguistic demands inherent in the successful performance of the task (Weir, 2005; Shaw and Weir, 2007). Cognitive validity looks at the extent to which the cognitive processing of a test-taker in completing, for example, a writing task resembles that of the test-taker in the target language situation. It involves collecting both *a priori* evidence on the cognitive processing activated by the task before the real test event, through methods such as think-aloud protocols, and *a posteriori* evidence on constructs measured through statistical analysis of scores after the task is performed.

Scoring validity is linked to both context and cognitive validity and accounts for all aspects of reliability (Weir, 2005). It considers "the extent to which test scores are based on appropriate criteria, exhibit consensual agreement in their marking, are as free as possible from measurement error, stable over time, consistent in terms of their content sampling and engender confidence as reliable decision-making indicators" (Shaw and Weir, 2007, p. 6). Criterion-related validity is primarily a quantitative concept and it accounts for the extent to which test scores correlate with a suitable external criterion of performance with established properties (Shaw and Weir, 2007). Consequential validity looks at a test's impact on institutions and society, its washback on individuals in classroom or workplace, and avoidance of test bias.

This socio-cognitive validation framework conceptualises the validation process in a "*temporal frame*" (Shaw and Weir, 2007), and thus it is easy to identify each type of validity evidence that needs to be collected at each stage in the test development, monitoring and evaluation cycle. Furthermore, this socio-cognitive approach improves the operability of validation work to a greater degree and it is a relatively cohesive validation framework that almost covers all the aspects which should be considered in a practical validation process. A number of language examination boards such as Cambridge English Language Assessment have used this framework to examine the extent to which the six aspects of validity in the framework have been operationalised in their tests of the four language skills, that is, listening, speaking, reading and writing.

Although the socio-cognitive validation framework has led to improvements in test design and validation, its current application has largely been limited to language tests assessing the writing-only skills. As reviewed earlier in this chapter, reading-to-write tasks have been increasingly used to assess test-takers' ability to write from sources, but the construct of this type of task remains under-theorised. Therefore, it is necessary to extend the use of this framework to the design and validation of reading-to-write tasks to gather more evidence to support their legitimacy.

This study aims to examine test-takers' cognitive processes while completing the TBEM-8 reading-to-write task and establish evidence of its cognitive validity. By reviewing literature on models of writing, reading and discourse synthesis, and on relevant studies of reading-to-write process in previous sections, a set of cognitive processes to be investigated in this study are proposed and described in detail in the following subsection.

2.4.2 Cognitive processes to be investigated in this study

In order to investigate the cognitive validity of a reading-to-write task, it is necessary to provide evidence that test-takers are engaging with the range of cognitive processes considered integral to real-world reading-to-write activities. Based on the relevant literature, this study proposes a set of ten categories of cognitive processes that test-takers are likely to use while completing the reading-to-write task. They are: text interpretation, task representation, macro-planning, organising, selecting, connecting, micro-planning, translating, monitoring, and revising.

I. Text interpretation

As presented in Hayes (1996) model, text interpretation is a process in which writers create "internal representations from linguistic and graphic inputs" (p. 13), and is mainly concerned with reading activities. This process is involved in almost any type of tasks. In a traditional independent writing task, text that needs to be interpreted includes the text in the task instructions and the text writers have written, while in an integrated reading-to-write task, text in the source materials is also added into the whole text, which may result in differences in writers' cognitive processing while completing the task.

Khalifa and Weir's (2009) reading model provides a useful classification of reading activities that test-takers perform in real-life situation. Two types of reading are identified in their model: careful and expeditious. Careful reading involves comprehending every part of the whole text, while expeditious reading involves selective, quick and efficient reading to access desired information from a text. Careful and expeditious reading can both be accessed at local and global levels.

Careful local reading is used to comprehend the meaning of sentence(s), during which lower-level processes such as "decoding at the word or phrase levels" and "establishing propositional meaning at the sentence level" are involved; careful global reading is used to comprehend main ideas or the majority information in the whole text, and higher-level processes such as "linking propositions in building a mental model" and "inferencing" are involved in this type of careful reading (Khalifa and Weir, 2009). Expeditious local reading is used to scan or search for specifics in the text, while expeditious global reading involves skimming for gist, or searching for main ideas and important details.

Studies of reading processes in language testing are mainly concerned with independent reading tasks (for example, reading comprehension). Most findings have revealed that independent reading tasks seem to be targeted at measuring careful local reading at the clause and sentence level rather than careful reading at the global level, and rarely at expeditious reading (Urquhart and Weir, 1998; Khalifa and Weir, 2009; Moore, Morton and Price, 2010). There is little research on reading processes involved in integrated reading-to-write tasks.

II. Task representation

As Flower et al. explained in their 1990 study, task representation is an interpretative process during which test-takers create an initial understanding of the task demands. Test-takers usually start with the process of task representation when responding to any type of task. In a writing task, they tend to create a representation of the task by reading through the task instruction, which contains information about the topic of the task, rhetorical functions expected, for example, describing and discussing, and contextual constraints such as time constraints and word length, and sometimes scoring criteria and information about the input materials (in a reading-to-write task).

Task representation is an important process because test-takers' performance is dependent on their understanding of the task. As discussed earlier in Hayes's (1996) writing model, if writers create a task representation based on a misunderstanding of the instructions, they may not be able to address the task appropriately. Flower et al. (1990) found that undergraduate students created different representations for the same reading-to-write task in terms of main sources of ideas, text features, organizational structure of the text, and strategies to use. Also, their results indicated that students with more experience in academic writing tended to create a more accurate task representation than students with less academic writing experience.

In studies of L2 writing, a number of researchers have investigated the process of task representation in completing a reading-to-write task. Ruiz-Funes (2001) examined the written products of 14 Spanish-as-a-foreign-language students who composed an essay discussing a literacy text, and found that writers approached the task differently, and resulted in various rhetorical styles. The more cognitively complex style, however, did not lead to a text with more syntactically complex structures. Allen (2004) followed an English-as-a-second-language student through a linguistic class assignment, finding that the student's representation of the task was greatly impacted by her prior experience in writing from external source materials. Similar to Allen's study, Wolfersberger (2007) found that four Chinese writers' representation of a classroom-based reading-to-write task were shaped by a variety of personal and contextual factors such as writers' background, prior experience, and interactions with course lecturers during the writing process.

Plakans (2010) compared ten writers' task representation process in completing an integrated reading-to-write task and an independent writing task through think-aloud protocols and interview, finding that some writers failed to spot the difference between these two types of tasks, and used the same independent writing process to compose essays. Also, her findings revealed that all writers followed "an initial circular process of reading and rereading the integrated prompt that consumed time and increased the complexity of understanding the instructions in task" (Plakans, 2010, p. 193), which was not found in the independent writing task.

III. Macro-planning

As discussed earlier in Hayes and Flower's (1980) model, they proposed that the process of planning involves generating, organising and goal setting. Based on Hayes and Flower's writing model, Field (2004) divided planning into three processes: macro-planning, organisation and micro-planning, to illustrate different purposes of the planning activities. Macro-planning is a process whereby writers plan for the writing goals and content, and identify major constraints of the task such as the target readership, genre and the level of formality required, on the basis of their understanding/representation of the task (Field, 2004; Shaw and Weir, 2007).

Scardamalia and Bereiter (1987) discovered that unskilled writers did not seem to use planning processes at the macro-level because they adopted a knowledge-telling approach to writing when they retrieved and listed ideas from their long-term memory in a rather linear way. In contrast, skilled writers tended to use a knowledge-transforming approach to writing, during which they put considerable effort into macro-planning to guide their writing. Similarly, Field (2004) argued that skilled writers paid much more attention to planning processes than less skilled writers did, and Eysenck and Keane (2005) claimed that it is the planning process that helps to distinguish between advanced writers and novice writers. There are also studies (for example, Hyland, 2002) which found that L2 writers are very likely to plan less during writing than L1 writers, and have more difficulty in setting goals for writing.

Burtis, Bereiter, Scardamalia and Tetroe (1983) attempted to look at what writers actually planned during writing through think-aloud protocols, and found that novice writers' planning protocols closely resembled the ideas put forward in their written product, while the protocols of advanced writers' planning process contained "provisional ideas, goal statement, comments and problem-solving attempts" (Burtis et al., 1983, p.154). Contextual features of the task seem to be another factor that impacts on writers' macro-planning process (Grabe and Kaplan, 1996; Shaw and Weir, 2007).

IV. Organising

In a traditional independent writing task, organising is a process in which writers organise the ideas to be put into their text by evaluating their priorities and relevance to topic of the task, while in an integrated reading-to-write task, as Spivey (1991) argued, writers not only order ideas in their own text, they also organise the relationships between ideas in the source texts to achieve an understanding of the text.

Field (2004) claimed that when writers are producing text, they often have an abstract provisional organisation of ideas in their mind, dependent on the task types. For example, if the task asks them to describe an event, they would be likely to have a sequential structure in their mind; if they are required to compare and contrast in a task, they would possibly have an advantage-versus-disadvantage structure in mind. These structures created in the writer's mind may or may not be same as those inherent in the source texts. Therefore, the writer may retain a similar structure in their text as the one presented in the source texts (Spivey, 1984), while they may also generate a new structure in order to absorb different ideas from multiple source materials.

Scardamalia and Bereiter (1987) discovered that writers who adopted a knowledge-telling approach to writing devoted little effort in ordering the ideas to be put in the text, but generated text in a rather linear fashion, that is, put down ideas in the order as they were retrieved from long-term memory. This is like dumping all the relevant knowledge in writers' mind at once, a process often found in writing-disabled students whose ability to plan is believed to be disrupted (Cherkes-Julkowski, Sharp and Stolzenberg, 1997). On the other hand, writers who adopted a knowledge-transforming approach to writing were actively engaged in organising processes as they transformed the ideas from their mind into the text by ordering and prioritising these ideas based on the evaluation of their relevance and importance to the writing goals.

It should be noted although the organising process has been extensively investigated in independent writing tasks, it is under-researched in integrated reading-to-write tasks. In Plakans' (2009b) study, she investigated test-takers' processes of discourse synthesis through think-aloud protocols and interviews, finding that some writers did spend time organising the relationships between ideas in the source texts to support their reading and guide their text production.

V. Selecting and connecting (generating)

Selecting and connecting are the other two important processes in Spivey's discourse synthesis model (Spivey, 1984, 1990, 1997; Spivey and King, 1989). Selecting is a process used when writers select relevant ideas or information from the source materials or their long-term memory to put into the text they are going to produce. Spivey (1991) argued that selecting plays an important role in meaning construction because the new meaning constructed is based on the ideas writers select from either internal or external sources.

Scardamalia and Bereiter (1987) found that advanced and immature writers employed the selecting process very differently. Immature writers select ideas simply by ranking information according to importance when recalling knowledge from memory, while advanced writers devote more cognitive effort in selecting content by resorting to a set of criteria, for example, the relevance to the writing goals, the appropriateness for intended readers, fitness to the overall structure.

Connecting is a process in which writers bring what they already know into the reading and create meaning-enhancing additions (Levin, 1988). In other words, writers combine the knowledge they retrieve from memory with the ideas they select from source materials, and generate either links between these ideas or new meaning (Kucer, 1985; Spivey, 1987). As they select and connect during reading, they are creating a pool of ideas from which to draw for the writing process (Stein, 1990). The output of selecting and connecting may ultimately become the basis of plans for the writing.

As discussed in Bereiter and Scardamalia's (1987) model, writers who adopt the knowledge-telling approach engage in a rather linear and straightforward writing process, during which the main activities are retrieving ideas from memory and putting them into the text. They are less likely to connect ideas in the source texts with their own knowledge when writing from sources. In contrast, skilled writers who use a knowledge-transforming approach tend to constantly connect ideas from their memory and source texts to generate ideas for the new text. These new ideas may be repetitive and vary in importance to the writing goals, thus the process of organising may be activated to address these issues.

VI. Micro-planning and translating

As Field (2004) argued, the processes of planning and organising not only take place at the macro-level, they may also be conducted at the micro-level, that is, at the sentence and paragraph level. During the micro-planning process, writers plan for the text that is about to be produced. At the paragraph level, writers plan for the goals, content and structure of a particular paragraph, possibly with constant reference back to the macro-plans established earlier (for example, the overall writing goal, genre and level of formality) as well as the text written so far. At the sentence level, writers plan for the structure and content of an upcoming sentence. It is believed that the actual text production process is based on writers' micro-plans rather than the macro-plans (Field, 2004).

The output of the micro-planning process is stored in writers' mind in the form of specific goals at the paragraph and sentence level, which then become the bases of the translating process, during which writers' abstract ideas are translated into concrete linguistic forms. Shaw and Weir claimed that it is through the translating process that "the writer moves from an internal 'private' representation, which is abstract and only understood by him or her, to its expression in the 'public' shared code of language" (2007, p. 39). They also argued that the language translated needs to be not only lexically and syntactically correct but also functionally appropriate. Field (2005) further pointed out that for L2 writers, the translating process may be so demanding in cognitive resources that the execution of other processes (for example, organising) is hindered.

Micro-planning and translating are two important processes when writers make micro-plans and carry out these plans to produce text. However, compared to other processes involved in writing, they may be more difficult to be investigated reliably as writers tend to be less aware of the use of these two processes. Previous studies that investigated micro-planning and translating processes have been performed solely under experimental settings by using methods such as directed verbal protocols; it is also believed that these two processes may not differ as much as other cognitive processes between an independent writing task and an integrated reading-to-write task (Chan, 2013).

VII. Monitoring and revising

Although the process of monitoring appears to be paid less attention in the models presented earlier, Field (2004) pointed out that writers actually engage in the monitoring process at different levels throughout the writing process. At a basic level, monitoring involves checking the mechanical accuracy of the text produced, for example, spelling, word use and syntax, while at a more advanced level, it involves monitoring higher-level features of the text produced such as development of arguments, relevance to and adequacy for the task set (Field, 2004, 2005; Shaw and Weir, 2007).

Because monitoring is a demanding process which requires high mental resources, it is subject to attentional constraints. Field (2004) argued that writers seem to focus on only one level of monitoring at a time; low-level monitoring is likely to happen during the text production process, while high-level monitoring may be reserved for a post-production stage, that is, when a certain amount of text had been produced. He (2005) further claimed that many L2 writers do not engage in monitoring processes because it is difficult for them to assess their writing qualities during translation due to its extra cognitive demands in retrieving linguistic knowledge.

The revising process is highly connected with the process of monitoring and may be conducted at any stage in writing. When revising, writers return to aspects of the text identified as unsatisfactory and make corrections or adjustment (Shaw and Weir, 2007). Although these aspects identified may not all be revised, "it is very unlikely that revising occurs without monitoring" (Chan, 2013, p.71). There are two levels of revising, each corresponding to one of the two levels of the monitoring process: at the basic level, writers make revisions of issues relating to textual features, for example, spelling and word use; at the advanced level, writers deal with issues such as the development of arguments, coherence and cohesion of the text.

Many studies have compared the use of revising processes between skilled and unskilled writers, finding that skilled writers are more proficient in revising than their counterparts (Flower and Hayes, 1980; Graham and Harris, 1996, 2000; Perl, 1979). Hayes and Flower (1980) found that fifteen percent of the protocols reported by skilled writers contributed to the revising process. Perl (1979) found that writers who adopted a knowledge-transforming approach engaged more often in revising writing goals and main ideas of the text. In contrast, novice writers devote much less attention to revising in writing, and, when revising, they are more likely to revise lower-level features of the text, for example, correcting spelling errors and making small changes in wording (MacArther, Graham and Harris, 2004). It seems that it is the level of revising process but not the number of revisions that distinguishes between skilled and unskilled writers (Scardamalia and Bereiter, 1987).

2.5 Research questions

Several research gaps emerged based on the review of literature presented above. Compared with the abundance of studies on independent writing tasks, there has been relatively less research on integrated writing tasks in the literature of L2 writing assessment. Among the studies that attempt to investigate the reading-to-write construct, the majority of them have focused on discourse features of test-takers' written products and score interpretations (Ascensión, 2008; Brown et al. 1991; Cumming et al., 2005, 2006; Gebril, 2009; Lewkowicz, 1994), and use of source texts (Campbell, 1990; Gebril and Plakans, 2009; Johns and Mayes, 1990; Watanabe, 2001). Relatively few studies have been conducted to examine the reading-to-write construct as a unique set of processes. Among these few attempts to explore the reading-to-write processes, most studies used think-aloud protocols, interview or questionnaire techniques to collect and analyse data, each of which is considered problematic when used alone.

Therefore, in order to address the gap in research, it was decided to combine eye-tracking, stimulated recall and questionnaire methods (these methods will be discussed in detail in Chapter 3) to examine test-takers' cognitive processing while completing the TBEM-8 reading-to-write task. This study contributes to the understanding of reading-to-write processes and establishes cognitive validity of the TBEM-8 reading-to-write task. Also, the usefulness of combining different research methods in integrated writing process studies is presented.

Two sets of research questions were proposed. The general aim of this study was translated into the first overarching research question:

RQ1. What cognitive processes do test-takers employ while completing the TBEM-8 reading-to-write task?

To gain further insights into test-takers' cognitive processing, one sub-question

(RQ1a) was formulated, exploring the nature of cognitive processing depending on test-takers' performance on the task:

RQ1a. Are there any relationships between the use of cognitive processes and test-takers' performance on the task?

The second overarching question aims to look at how test-takers engage with the source materials through an online investigation of their eye movements during task completion:

RQ2. To what extent do test-takers engage with the source materials in the TBEM-8 reading-to-write task?

Also, two sub-questions were proposed to further investigate test-takers' source use and explore the difference in eye-tracking measures across different performance levels:

RQ2a. Are there any difference in eye-tracking measures among different source materials?

RQ2b. Are there any relationships between eye-tracking measures and test-takers' performance on the task?

2.6 Summary

In this chapter, published literature in reading-to-write processes has been reviewed to provide a foundation for the thesis. First, in Section 2.2, relevant models of reading and writing processes, as well as a discourse synthesis model were presented to shed light on the reading-to-write process and identified several processes (for example, selecting and connecting) that are unique in reading-to-write tasks. Research on integrated writing tasks was then reviewed in Section 2.3 in terms of four different topics: comparison (between independent and integrated writing tasks) studies, discourse features, process studies, and source use; the focus of this study was pointed out as investigating test-takers' reading-to-write processes and methods of previous process studies were summarised to determine the research methods (see Chapter 3 for more discussion on the methods) used in this study. Section 2.4 revisited several validation frameworks in language testing and proposed a set of cognitive processes to be investigated to provide evidence for cognitive validity. Finally, based on the literature review, two sets of research questions to be addressed are proposed.

In the next chapter, methodological grounding for the research design of this study will be discussed in more detail, and the research methods for each separate study will be described.

CHAPTER 3 METHODOLOGY

3.1 Introduction

This chapter introduces the methodology for the present study from a macro- and micro-perspective. First, Section 3.2 provides some macro methodological grounding for the research design. The issues considered here relate to the study as a whole, illustrating the methodological underpinning to each chosen research method. Second, the micro plans for data collection and analyses of the two sources of data – from Study I, the eye-tracking and stimulated recall study, and Study II, the questionnaire study – are presented in Sections 3.3 and 3.4. Section 3.5 concludes this chapter.

3.2 Methodological grounding

The present study seeks to look into test-takers' cognitive processing while completing a reading-to-write task, which could potentially be investigated through a variety of methodological approaches used in relevant previous studies. However, as discussed in the literature review, these approaches (e.g., think-aloud protocols, questionnaires) have different drawbacks when used alone, and are liable to over-generalise the findings beyond the limitation of each method. Sections 3.2.1 and 3.2.2 below discuss the potential usefulness of these methods (eye-tracking, stimulated recall and questionnaires) in the present study, illustrating their strengths and drawbacks, and finalise the methods for data collection in Study I and II.

3.2.1 Combining eye-tracking and stimulated recall

The theoretical underpinning to the eye-tracking technique is that our eye movements can be used to make inferences about our cognitive processes (Peyrichoux and Robillard-Bastien, 2006). One of the main benefits of eye-tracking is that it is, to the best of the author's knowledge, the only method that can be used to objectively and accurately record and analyse individuals' visual behaviour, thereby allowing us to study a participant's eye movements when performing specific tasks (e.g., listening and reading). This gives insights into the cognitive processes underlying their looking behaviour and reveals things such as reading patterns throughout task completion.

Although eye-tracking adds detailed, quantitative data to understanding a participant' cognitive processes, the data cannot always be clearly interpreted without participants providing information about their behaviour (Hyrskykari, Ovaska, Majaranta, Räihä and Lehtinen, 2008). For example, a longer fixation does not necessarily mean the participant found a particular area interesting, but it may also mean that they found it hard to interpret (Cowen, Ball and Delin, 2002). Therefore, it is of importance to supplement eye-tracking data with additional qualitative data gained from participants on their experiences to facilitate interpretation.

Think-aloud methods have the potential to be combined with eye-tracking to add more qualitative information to the data. They are commonly used in second and foreign language testing research (Ascención, 2005, 2008; Green, 1998; Plakans, 2009b; Yoshida, 2007; Yu, Rea-Dickins, & Kiely, 2011). As a common source of data elicitation, they can be broadly categorised as either concurrent (on-line) or retrospective (off-line). The concurrent think-aloud (CTA) method allows a participant to verbalise their thoughts during task completion, while the retrospective think-aloud (RTA) method requires participants to report their thoughts either during specific breaks in the actual task, or immediately after they have completed a task (Leow and Morgan-Short, 2004).

Both methods are effective ways of gaining insights into participants' cognitive processes regarding task completion, however, each one has its own limitations and problems. In general, think-aloud protocols may not be sufficient since certain cognitive processes are unconscious, and participants may thus not be able to adequately report their thought processes. A serious critique of the CTA method is that it is more easily affected by *reactivity*, that is, "By thinking aloud, participants' internal processes may differ from what they would have been had they not performed the verbalisation" (Leow and Morgan-Short, 2004, p.38). As the cognitive workload increases, participants may be less likely to fully report meaningful information, or their natural behaviour (i.e., their linguistic and/or nonlinguistic output) may be more likely to be altered by the disruption imposed on the actual cognitive processes, thereby biasing results. Similarly, the RTA method is not a problem-free methodology as well. It must be used with care; as the participant is asked to recall the way they complete the task rather than provide real-time information while doing the task, certain processes may be forgotten or participants may intentionally or unintentionally fabricate information due to imperfect memories (Russo, 1979).

The combination of the CTA method with eye-tracking technique has proven to be less suitable in practice because participants may produce eye movements that they would not normally do if completing the task without thinking aloud in a normal environment (Kim, Dong, Kim and Lee, 2007). For example, they may fixate on certain areas of the screen while verbalising their cognitive processes. The RTA method is more appropriate to be used in process studies (particularly when participants have to perform tasks which require high cognitive demands, e.g., the TBEM-8 reading-to-write task) where quantitative eye-tracking data will be analysed.

Since memory decay and potential for fabrication are likely to happen when using the traditional RTA method, a variety of this method emerged, that is, cued RTA, or referred to as ‘stimulated recall’ in this study, which is “carried out with some degree of support for the recall” (Gass and Mackey, 2000, p. xi). Examples of commonly used support include showing participants a video playback so that they can watch themselves performing the original task, or “giving learners their L2 written product, so that they can follow the changes they made” (Gass and Mackey, 2000, p. xi). The stimulated recall method has proven to be able to get more detailed information from participants (Namahn, 2001), and also allows the participants to reflect upon their actions more actively that they may not be able to do through other methods. Using a video cue that features a participant’s eye movements (eye-movement recordings) has also been demonstrated effective at eliciting comments from participants (Brunfaut, 2016; Brunfaut and McCray, 2015; Holzknecht, Eberharter, Kremmel, Zehentner, McCray, Konrad, & Spöttl, 2017; McCray and Brunfaut, 2018; Yu, He, & Isaacs, 2017), as it shows in much detail the participants’ eye traces throughout task completion, which almost eliminates the risk of fabrication.

Therefore, based on the above discussion of the eye-tracking technique and two types of think-aloud methods, it was decided to combine eye-tracking and stimulated recalls to obtain data on test-takers’ cognitive processes while completing the TBEM-8 reading-to-write task (Study I). This can potentially balance the strengths and weaknesses of each individual method: the recordings of participants’ eye movements acted as stimuli for their recalls of cognitive processes employed during task completion, and the recalls in turn added more qualitative information to help the understanding of the eye-tracking data. In the next section, the potential to use another research method, questionnaires, will be discussed.

3.2.2 Reading-to-write process questionnaire

Although the combination of eye-tracking and stimulated recall methods may generate richer data from participants, it has certain drawbacks. First, it is very time-consuming to conduct such a study in practice. The research design is often intricate and operating an eye-tracker is a demanding task. The researcher needs to be well trained before carrying out an eye-tracking and stimulated recall experiment. The selection of ideal participants may be more of an art than a science as the eye-tracker works better on some people than others (this will be discussed in more detail in Section 3.3). Second, as it is time-consuming and demanding, it could only be applied to a relatively small number of participants, so that interpreting the results too broadly would be risky, and any conclusions drawn should be seen as tentative.

Therefore, a reading-to-write process questionnaire (see Appendix E for a full copy of the pilot questionnaire) was also utilised to elicit participants' cognitive processes, so as to offset, to some extent, the drawbacks of eye-tracking and stimulated recall methods, since it can "report the cognitive processes employed by a large number of participants in different conditions in a systematic and efficient way" (Chan, 2013, p.102). This questionnaire was developed by Chan (2013), and adapted according to the features of TBEM-8 reading-to-write task. In this questionnaire, 42 items were grouped into five hypothesised phases of academic writing, i.e, *conceptualisation, meaning and discourse construction, organisation, low-level monitoring and revising* and *high-level monitoring and revising*, which are mainly based upon Field's (2004, 2008, 2011, 2013) model of cognitive processing activities involved in writing, and Shaw and Weir's (2007) model of writing processes. In addition, other relevant cognitive models including Hayes and Flower's (1980) writing model, Spivey's (1984, 1990, 1997, 2001) discourse synthesis model, and Khalifa and Weir's (2009) reading model were studied to determine the reading-to-write cognitive processes that writers are hypothesised to undergo in each of the five academic writing phases presented above.

Table 3.1 shows the structure of the pilot reading-to-write process questionnaire. Seven categories of cognitive processes were identified in the questionnaire. They were *task representation, macro-planning, text interpretation, connecting and generating, organising, low-level editing* and *high-level editing*. The 42 items were organised in five stages: *while reading the task prompt, while reading the source materials, before writing, while writing the first draft* and *after writing the first draft*; the digit in front of the decimal point of an item number indicates which of the five stages this item is in. A 5-point Likert scale was used (5=strongly agree; 4=agree; 3=no view; 2=disagree; 1=strongly disagree), for example, **Item 2.1** below is one of the ten items in the second stage (*while reading the source materials*).

Item 2.1*: I read through the whole of each source text carefully.*
1. Strongly disagree 2. Disagree 3. No view 4. Agree 5. Strongly agree

At the end of each stage, there is one open-ended question eliciting more thoughts from participants about their thought processes, for example, below is the open-ended question in the first stage (*while reading the task prompt*).

What else did you do while reading the task prompt?

Table 3.1: Structure of the pilot questionnaire (42 items)

Phases of academic writing	Cognitive Processes	Stages					No. of items
		Reading task prompt	Reading source materials	Before writing	Writing the 1st draft	After writing the 1st draft	
Conceptualisation	Task representation	1.4	2.5		4.3		8
	Macro-planning	1.2 1.3 1.5	2.11		4.5		
Meaning and discourse construction	Text interpretation	1.1	2.1 2.2 2.3 2.4 2.6		4.4		10
	Connecting and generating		2.8 2.10		4.2		
Organisation	Organising		2.7 2.9	3.1 3.2 3.3	4.1		6
Low-level monitoring and revising	Low-level editing				4.11 4.13 4.14	5.6 5.8 5.9	6
High-level monitoring and revising	High-level editing				4.6 4.7 4.8 4.9 4.10 4.12	5.1 5.2 5.3 5.4 5.5 5.7	12

This preliminary questionnaire was first piloted with 77 participants. Revisions were made according to the results of several statistical analyses (see Section 3.4 for details of the pilot study). Also, as the pilot study was conducted almost the same time as the eye-tracking and stimulated recall study (Study I), some feedback from Study I fed into the revisions of the pilot questionnaire.

In summary, it was decided, based on the literature review and the discussion of each individual methodology above, to combine eye-tracking and stimulated recall techniques to obtain an in-depth look into test-takers' cognitive processes while completing the TBEM-8 reading-to-write task (Study I); furthermore, a reading-to-write process questionnaire was administered to a larger test-taking population (Study II), allowing for triangulation of the findings of each individual method. In the next two sections, the research methods for Study I and II will be introduced in detail.

3.3 Study I: eye-tracking and stimulated recall – methods

This section introduces the methods for Study I, the eye-tracking and stimulated recall study. In Section 3.3.1, the recruitment of participants and their background information are introduced. Section 3.3.2 provides a detailed description of the equipment (Tobii TX300 eye-tracker) and instrument (the TBEM-8 reading-to-write task) used in this study. Section 3.3.3 describes the procedures for data collection. Methods for data analysis are explained in Section 3.3.4.

3.3.1 Participants

A total of 20 students participated in the eye-tracking and stimulated recall study. The participants were all Master's students and enrolled in either Linguistics and English Language or Finance programmes at Lancaster University. They were all native Chinese learners of English and, at the time of data collection, the majority had been living in an English-speaking country for less than twelve months. It was believed that these participants were the most suitable ones, given the location of the eye-tracking equipment, that could be found to possibly represent the target population of the TBEM-8 reading-to-write task (see Section 3.3.2 for details).

The participants were recruited through e-mails sent by postgraduate coordinators of the Department of Linguistics and English Language and the Management School to potentially eligible students. In the e-mails, a brief description of the study was provided, and it was clearly stated that participation in the study was completely voluntary and would not affect any evaluation on participants' degree programme. Also, it was stated in the e-mails that participants would receive compensation either in the form of some chocolate (for those who participated in the eye-tracking screening but were not successful) or 15 pounds for their participation and time (for those who participated in the full study). After contacting the researcher, eligible participants received the participant information sheet (see Appendix A) explaining the detailed procedures of the experiment, how the data would be handled and were informed that they were free to withdraw at any time without giving a reason.

The 20 participants were invited to book a slot for taking part in the experiment on Doodle (an internet calendar) and they were all present at the eye-tracking laboratory at the determined time and date. Two of the 20 participants proved to be unsuitable for being eye-tracked through "scanpath" inspection (Holmqvist et al., 2011), during which a red ball appeared and moved across the eye-tracker screen, and the participants were asked to keep their eyes focused on the ball as it moved to assess how accurately their eye movements followed the path of the red ball. Specifically, one participant had somewhat downward eyelashes which can block the reflection of the light coming out of the eye-tracker onto the screen and affect the accuracy of the eye-tracking data. The other participant was wearing a pair of thick glasses, which may also hinder the reflection of the light. Data were then collected from the remaining 18 participants who had been successfully screened for eye-tracking suitability. Out of these 18 participants, two participants' data were excluded due to insufficient accuracy (weighted gaze samples < 50%; 50% means that at least one eye was found for the full recording) for further analyses. The final data set therefore included 16 participants: 11 were female (69%) and five were male (31%); their ages ranged from 21 to 28 years (*Mode*=23; *Mean*=22.6; *SD*=1.66).

14 participants sat the International English Language Testing System (IELTS) test within one and a half year before data collection, while the other two took the test two years earlier. Table 3.2 summarises their performance on IELTS overall and on Reading and Writing components respectively. According to the Common European Framework of Reference for Languages (CEFR), these participants' proficiency levels were between B2 and C1.

Table 3.2: Participants' IELTS test scores

IELTS/IELTS components	Mean	Median	Mode	Standard Deviation	Minimum	Maximum
Overall	7.16	7.00	7.50	0.35	6.50	7.50
Reading	8.00	8.00	8.50	0.58	7.00	9.00
Writing	6.25	6.00	6.00	0.55	5.50	7.00

3.3.2 Equipment and instrument

The participants' eye movements were recorded using a screen-based binocular tracking eye-tracker: Tobii TX300 (Tobii AB, Sweden), whose major technical specifications are presented below, followed by a description of the primary instrument used in this study, i.e., the TBEM-8 reading-to-write task.

I. Tobii TX300 eye-tracker

The Tobii TX300 eye-tracker uses dark pupil and corneal reflection techniques to detect eye movements. During tracking, the infrared illuminators (see Figure 3.1) emit light and create reflection patterns on the corneas of the subject's eyes. These reflection patterns, together with other data about the eyes are collected by image sensors, at a sampling rate of 300 Hz per second (collecting raw eye movement data points every 3.3 ms; this frequency is high as 50/60 Hz is more common for similar type of eye-trackers; the gaze accuracy is 0.4° at the 300 lux illumination level). Image processing algorithms are then executed to identify relevant features, including the exact positions of the eyes and the correct reflection patterns from the illuminators. Last, a mathematical model of the eye is used to calculate the position of the eyes in space and finally to determine the gaze point on the screen, that is, where the subject is looking.

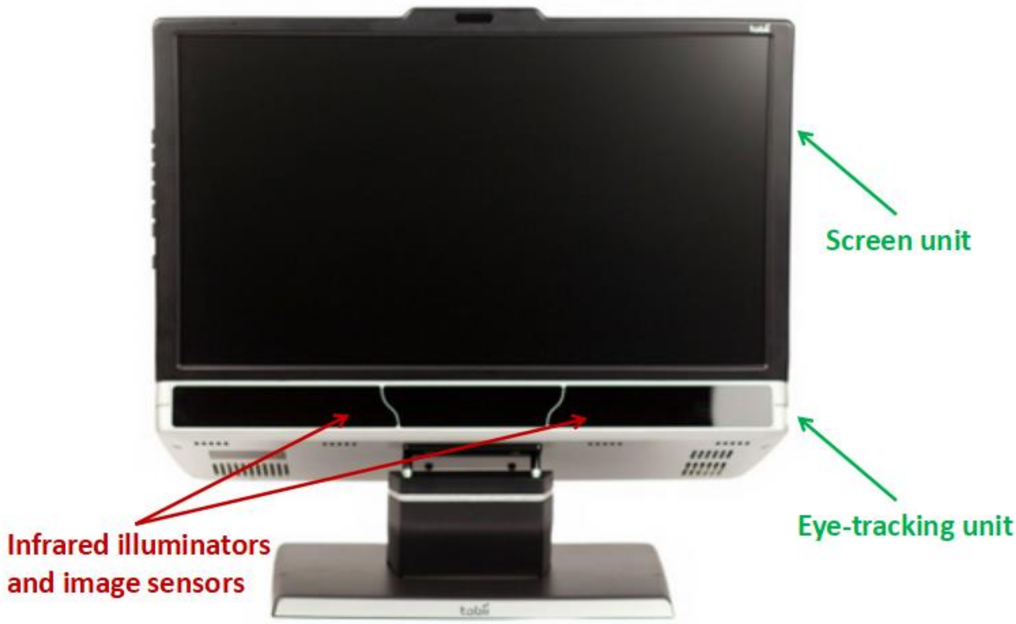

Figure 3.1: Tobii TX300 eye-tracker

The Tobii TX300 has a high tolerance for head movement. It allows the subject to move freely in front of the eye-tracker if their heads are positioned within an area of 37 cm (width) x 17 cm (height) at a distance of 65 cm from the screen (maximum head movement speed: 50 cm/s), and thus eye movements such as fixations and saccades can be studied without using a chinrest, a device for stabilising the head which may cause the subject to feel uncomfortable during the experiment. If the participant moves out of this area while being eye-tracked and then back into it, tracking is recovered almost instantly (time to tracking recovery after lost tracking: 10-165 ms). The infrared illuminators and image sensors, as shown in Figure 3.1, are located underneath an ordinary looking monitor (screen unit). They are both invisible to the human eye causing no disturbance to the subject in an experiment so that a participant would perform the task as if sitting in front of a normal computer screen. The freedom of head movement and unobtrusiveness allow participants to act more naturally and minimize their fatigue, particularly in a lengthy experiment such as the one reported in this study, which involved a reading-to-write task lasting about 40 minutes. In this way, the features of the specific eye-tracker used contribute to the validity of the claim that performance is authentic.

Figure 3.2 shows the layouts of the eye-tracking lab during the experiment in this study. As the data were collected from one participant at a time, two people were present in each session of data collection. In Figure 3.2 the individual depicted in green was the participant, who was seated in front of the Tobii TX300 eye-tracker. The distance between the participant's eyes and the eye-tracker screen was within a range of 50–80 cm, depending on participants' preferences for a comfortable position when working with a computer. As well as the main screen attached to the eye-tracker, there was another computer monitor on the same desk. It was used as the monitoring screen on which the participant's live eye movements were shown. This monitor was facing away from the participant in order to avoid any distraction that may be caused by the information shown on the screen. However the screen was angled in such a way that the researcher could monitor performance (the individual depicted in grey in Figure 3.2). The researcher sat around the corner in the lab, and monitored the participant's composing process primarily to deal with any issues which might have arisen during the experiment. A video camera was also placed behind the participant and used to record the stimulated recall session, during which an audio recorder (placed on the desk somewhere close to the participant) was used to create back-up audio recordings in case the video camera failed, or the video sound was not clear.

Figure 3.2: Layouts of the eye-tracking lab during the experiment

II. TBEM-8 reading-to-write task

One sample task of the TBEM-8 reading-to-write task (see Appendix C for the original task and Appendix D for an English version of it) was provided by the TBEM Testing Committee and investigated in this study. It should be noted that, as the TBEM-8 is still in an early stage of development, further tasks could not be supplied by the committee due to confidentiality and the small number of existing tasks, most of which are being used in live testing.

However, this sample task was considered a prototype task developed based on the writing test specifications of the TBEM-8, in which it is stated that the reading-to-write task is designed to assess test-takers' ability "to generalise and integrate information in the Chinese and English sources provided to write an English essay" (TBEM-8 testing committee, 2012, p.5). This sample task was, therefore, fundamentally indicative of the future tasks which would be developed. Other reading-to-write tasks which share features of the TBEM-8 task type (multiple language input, different types of sources, etc.) include the HKDSE English Language Paper 3 (Listening and Integrated Skills), in which test-takers are required to first complete a variety of listening tasks, and then to finish several integrated listening/reading and writing tasks of different levels of difficulty based on the same theme (Hong Kong Examinations and Assessment Authority, 2018).

The topic of the task concerned Steve Jobs' resignation from Apple. The task contained a set of instructions, and five source materials in the prompt. Source 1 (213 words) was a short passage in Chinese, which gave some background information of Steve Jobs and Apple; Source 2 (120 words) was a collection of English material including several video news headlines and two short excerpts from some internet news, all of which were on Steve Jobs' resignation; Source 3 (275 words) was another set of material in Chinese, and contained three short excerpts from some Chinese newspaper articles, which provided different views on Steve Job's resignation; Source 4, unlike other text materials, was a drawing of Steve Jobs' cartoon image, with a large Apple icon and some major Apple products beside it and also some additional text: "iRetire No more Jobs @ Apple" and "See Steve cook up one last announcement in his career"; Finally, Source 5 provided test-takers a list of ten words and expressions for reference while completing the task.

Instructions (117 words) were provided in English as follows:

> *In this task, you are required to write an essay of 250-280 words as an assignment for your professor of Strategic Management. You will be given 40 minutes to write an essay entitled* ***The Post Jobs' Era of Apple****. In the essay, you should describe the event, analyze the situation and comment on the impact of Jobs' resignation on Apple. Your essay is to be based on the source materials given below. But you should not simply copy and translate the source materials. Your essay will be judged according to how well you develop your ideas and how coherent your essay is. The task is to be completed on* ***Answer Sheet 3****.*

The instructions stated clearly (1) for whom this essay was to be written, so that the test-taker may be able to decide in what style the writing should be, for example, whether a colloquial style as might be used in an e-mail or an academic style similar to that used in an assignment for university course; (2) what content was expected in the writing (*describe the event, analyze the situation and comment on the impact of Jobs' resignation on Apple*); (3) how long the writing should be (*250-280 words*) and how much time (*40 minutes*) was given to complete the task; and (4) some indication of how the writing was to be scored (*how well you develop your ideas and how coherent your essay is*).

This task was displayed on the eye-tracker screen (23-inch TFT monitor; aspect ratio of 16:9; screen resolution of 1920×1080 pixels). Through a piloting process conducted with two participants, the task layouts were finalised and transformed for the eye-tracker screen in html format (see Figure 3.3). The task instructions and the first three source materials were presented down the left part of the screen and the other two source materials and the answer sheet (where participants wrote the essay) were presented on the right part of the screen. The font was legible, and its size was big

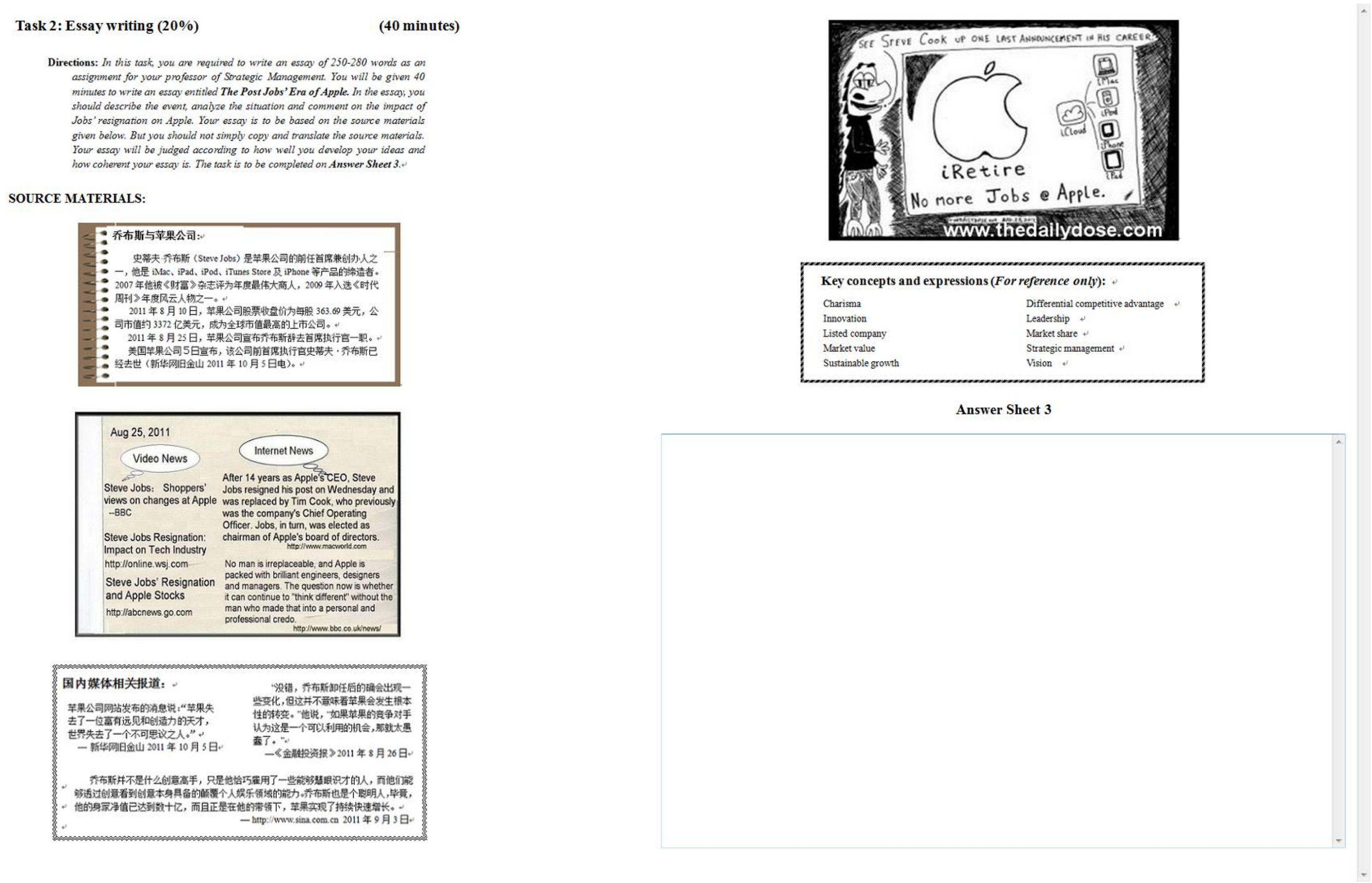

Figure 3.3: Layouts of the TBEM-8 reading-to-write task displayed on the eye-tracker screen

enough to be read, as reported by the pilot study participants. The answer sheet provided sufficient space (a maximum of 400 word in the Times New Roman with a font size of 13px/10pt) for participants to write on. Each part of the task was fixed on the screen, thus no scrolling was required, which made it possible for the eye-tracker to calculate eye movement data within each individual area on the screen.

3.3.3 Procedures for data collection

The data were collected over two sessions. During the first session (eye-tracking session), the participants completed the TBEM-8 reading-to-write task while their eye movements were being recorded by the Tobii TX300 eye-tracker. This was immediately followed by the second session (stimulated recall session), during which the participants were asked to verbalise their thoughts during task completion, using their eye traces recorded in the first session as stimuli for retrospection. Figure 3.4 shows the procedures for data collection. Factors that influenced this design included the nature of the TBEM-8 reading-to-write task and technical practicalities of the eye-tracking software (Brunfaut, 2016; Brunfaut and McCray, 2014, 2015). Due to the constraints of the data collection methods chosen, the data were collected from one participant at a time, and the session for each participant lasted about one and a half hours.

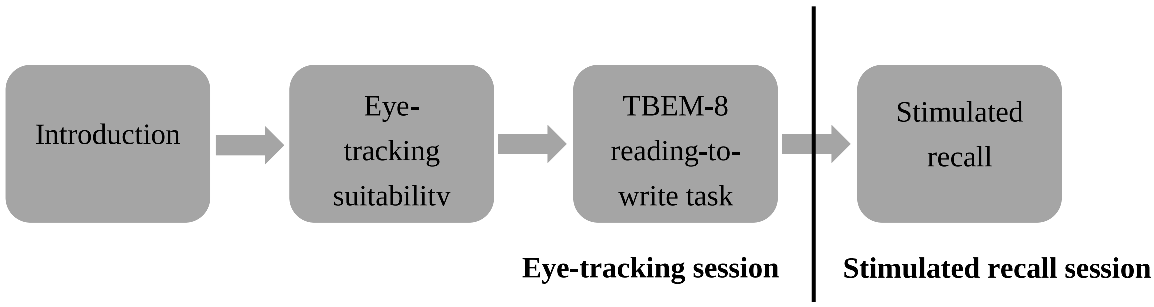

Figure 3.4: Flowchart of the data collection

As can be seen in Figure 3.4, the eye-tracking session started with an introduction to the experiment. The researcher explained to the 20 participants in detail what was expected of them, and then they were asked to sign an ethical consent form (see Appendix B), followed by an eye-tracking suitability test, which was to "determine whether the participant's eye-traces could sufficiently be captured by the hardware" (Brunfaut, 2016).

After the eye-tracking suitability test, the participant was instructed to find a comfortable seating position, which allows them to type easily on the keyboard without strain and look at the eye-tracker screen in a natural way. This is important because if the participant was sitting comfortably, their head movement was more likely to be within the range that the eye-tracker allows. Once a comfortable position was obtained, the participant was taken through a calibration procedure. During this procedure, the eye-tracker measures characteristics of the participant's eyes in order to collect eye traces as accurately as possible. The participant was instructed to keep their heads still during calibration and not to move their heads too much throughout the reading-to-write task completion afterwards (the eye-tracker allows some natural head movement, but too much movement could impact on the accuracy of the data collected). A 9-point built-in calibration procedure was used in this study to calibrate the eye-tracker.

Following successful calibrations, the participant proceeded to complete the TBEM-8 reading-to-write task, which was presented, as shown in Figure 3.3, on the eye-tracker screen. The participant's eye movements were simultaneously recorded as they completed the task. In order to maintain high eye-tracking accuracy, the eye-positions of the participants were monitored by the researcher throughout the task completion. This was achieved by looking at the "Track Status" window (see Figure 3.5) displayed on the monitoring screen (see Figure 3.2). The two white circles in the middle of the black box represent the participant's eyes. The circles do not need to be centered but they should not be too close to the edges. The multi-coloured side bar along the right edge represents optimal distance (the white triangle inside the bar should be neither too near the top nor too near the bottom). If the participant was found not to be staying within the acceptable boundaries, they would be instructed to adjust their position so that the eye-tracker could record their eye traces accurately throughout the task completion.

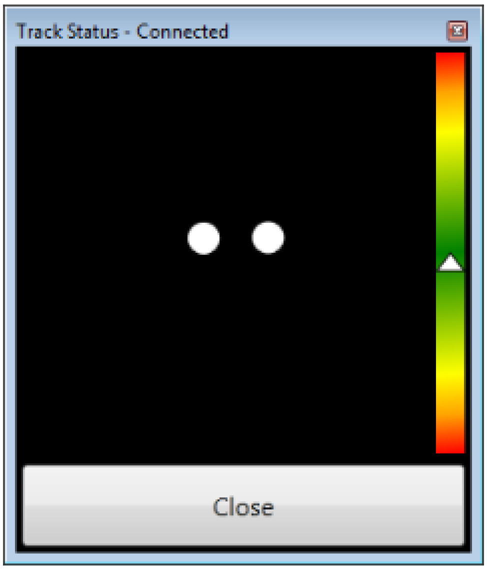

Figure 3.5: "Track Status" window in Tobii Studio

Immediately after completing the reading-to-write task, a stimulated recall session was conducted. During this session, the participant could move freely and was asked to recall their thought processes while reading and writing. The recordings of their eye traces were replayed for them to stimulate recall. This session was primarily led by the participant throughout though some researcher intervention of asking questions to clarify certain issues occurred at some point. The participant was allowed to pause or rewind the replay when they wanted to. As mentioned earlier, the stimulated recall sessions were audio- and video-recorded, while the transcription and the following data analysis were based on the video recordings because (1) in the pilot study the participant was found to point at the eye traces replayed on the screen when verbalising their thought processes; (2) the visual replay of the participant's eye traces, together with their recalls facilitate the interpretation of their thought processes while reading and writing. The audio recordings served as back-up in case the video sound was unclear or the device failed.

It should also be noted that the stimulated recall session was conducted in the participant's first language, Mandarin Chinese, therefore the participant was more likely to recall their thought processes more accurately and in more detail (Brunfaut, 2016; Brunfaut and McCray, 2015). After the participants finished reporting their thought processes, they were asked to fill in a background information questionnaire. The whole experiment was then over and the main data had been collected awaiting for further analysis.

3.3.4 Data analyses

To investigate the participants' eye movements and cognitive processing while completing the TBEM-8 reading-to-write task, heat-maps resulting from the recordings of the participants' eye traces were generated and four eye-tracking metrics were calculated. The stimulated recalls were coded to identify major types of cognitive processes. It was believed that the eye-tracking and stimulated recall analyses are complementary to each other (Brunfaut, 2016; Brunfaut and McCray, 2015; Holzknecht et al., 2017; McCray and Brunfaut, 2018), as the former would generate more quantitative data, for example, how long each participant spent on a particular source material while completing the task, while the latter can provide qualitative information on cognitive process employed by the participant during a particular period, thus balancing the strengths and weaknesses of each individual method and triangulating the findings. Before going into detail about the eye-tracking and stimulated recall analyses, the scoring of the participants' written products is first described below.

I. Scores on the TBEM-8 reading-to-write task

Two raters (both specialised in language testing and with experience in rating writing tasks) scored the 16 participants' essays using the TBEM-8 reading-to-write task rating scale, which is composed of two sub-scales: an analytic scale that has five differently weighted dimensions, i.e., register (0.8), organisation (1.0), coherence and cohesion (1.0), grammatical range and accuracy (1.0) and higher-order thinking (1.2); and a holistic scale. Thus, two types of scores were assigned by the raters: a set of analytic scores and a holistic score, which were then added up as the participants' final scores in line with the formal TBEM-8 scoring procedures.

The inter-rater reliability between the two raters was calculated. As Shapiro-Wilk tests of normality indicated that both variables' distributions were not statistically significantly different from normality ($p>.05$), Pearson product-moment correlations were adopted, achieving an acceptable correlation coefficient of $r=.79$ ($p=.00$**). Furthermore, Spearman's rank-order correlations between the participants' scores on the TBEM-8 reading-to-write task and their IELTS scores, particularly on the reading and writing components (normal distributions could not be found for some of the measures) were calculated. Results will be presented in the next chapter.

II. Eye-tracking analyses

In order to answer the second overarching research question (RQ2), that is, to what extent do test-takers engage with the source materials in the TBEM-8 reading-to-write task, heat-maps and four eye-tracking metrics were investigated. They are mainly based on the participants' eye fixation data, which in practice are determined by fixation filters (the algorithm for detecting fixations). In other words, fixation filters are responsible for what constitutes a fixation, and thus affect how eye-tracking measures such as fixation count, location and duration, are calculated. The fixation filter adopted in this study is the Tobii I-VT filter with its default settings (*max gap length* 75 ms; *eye selection* average; *noise reduction* disabled; *window length* 20 ms; *velocity threshold* 30 degrees/second; *max time between fixations* 75 ms; *max angle between fixations* 0.5 degrees; *minimum fixation duration* 60 ms), which is a velocity-based filter and considered to be functioning well on high-speed eye-trackers such as the Tobii TX300 (Brunfaut and McCray, 2015; Holmqvist et al., 2011; Olsen, 2012).

Absolute duration heat-maps (*radius* 50 pixels; *Scale max value* 3.00; *Opacity* 100%) were created using the fixation filter described above. In this type of heat-map, different colours are used to display the accumulated fixation duration on different locations in the image and thus can be used, for example, to measure the amount of time dedicated to a particular area of the stimulus. Red usually indicates the longest fixation duration and green the least, with varying levels in between. By contrast, areas of the stimulus which the subject does not look at remain transparent. Figure 3.6 shows an example of heat-map visualisation. It can be seen in the figure that, during the first one minute of recording, the participants spent the majority of time on the area of instructions while other parts of the task received limited amounts of attention. The heat-maps, therefore, are useful in understanding the overall distribution of a subject's attention on different areas of the input throughout the experiment.

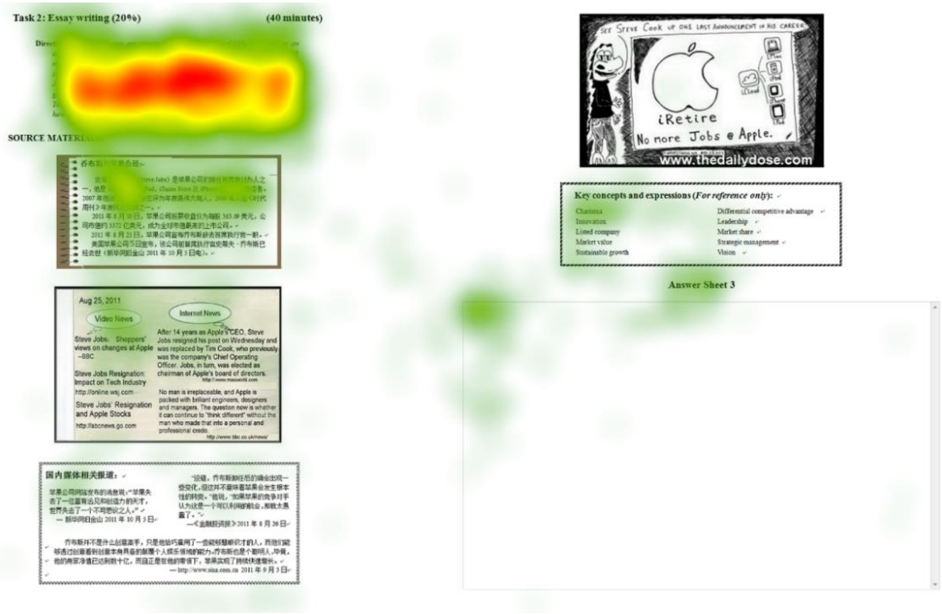

Figure 3.6: Heat-map of the TBEM-8 reading-to-write task for the first one minute of recording

A total of four metrics were also investigated in the eye-tracking analyses. Before the data analysis, the eye-tracker screen was divided into seven AOIs (areas of interest) corresponding to the seven parts of the TBEM-8 reading-to-write task (see Figure 3.7), which include the task instructions, the source materials one to five and the answer sheet where the participant wrote the essay. Having identified AOIs, the eye-tracker software can analyse fixation data within each individual area. Below are the four eye-tracking metrics examined in this study:

1. *Time to first fixation*, which measures how long it takes before a participant fixates on an AOI for the first time.

2. *Total visit duration*, which measures the duration of all visits within an AOI.

3. *Visit count*, which measures the number of times a participant visits an AOI.

4. *Visit duration*, which measures the duration of each individual visit within an

AOI.

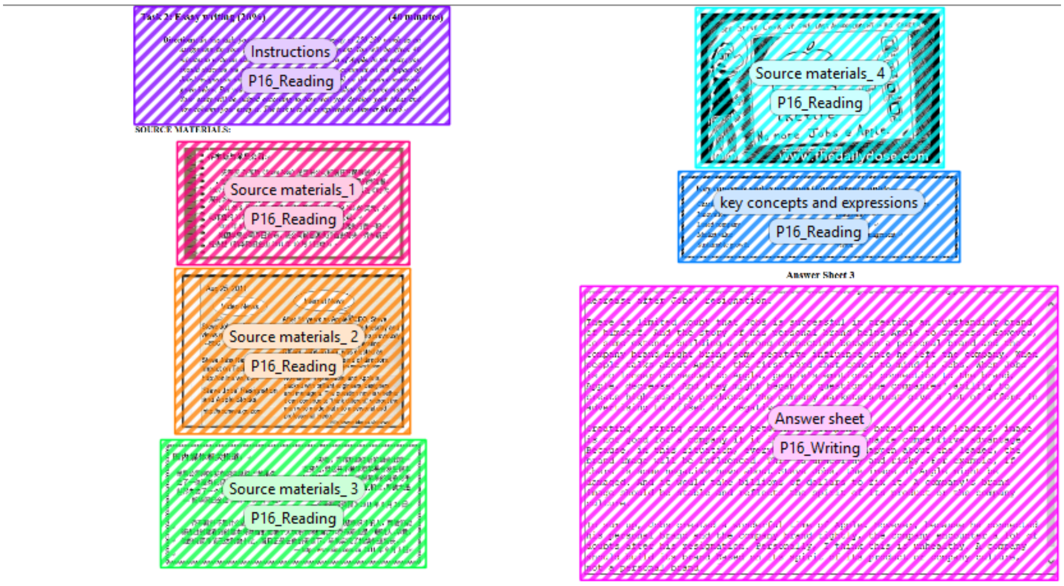

Figure 3.7: Areas of interest on the eye-tracker screen

Unlike the fixation itself, a visit is an interval between the first fixation on an AOI and the end of the last fixation within the same AOI. For example, a visit to the task instructions starts from a participant's first look at this AOI, and ends with this participant looking somewhere else, during which no fixations lie outside the area of instructions. Therefore, when a participant was, say, reading instructions, a visit would contain a number of fixations and last longer, in most cases, than a fixation. In this study, *total visit duration* was examined instead of *total fixation duration* because it is considered as a measure of the overall amount of the participant's processing as not only is the duration of all fixations within an AOI measured, but time spent on saccades (movements between two fixations) is added when calculating the total time of processing on a particular AOI of the subject's cognitive processing during task completion.

All the four measures described above can, to some extent, provide evidence for what writers attended to while completing the TBEM-8 reading-to-write task and thus can inform the answer to RQ2: (1) *time to first fixation* shows how the participant approached the reading-to-write task at the start of task completion; (2) *total visit duration* reports how long the participant spent looking at each part of the task, which can help to show us whether participants spent more time reading or writing; (3) *visit count* reveals the extent to which participants moved frequently between texts and the writing space, which can tell us something about how they used the texts; (4) *visit duration* provides the statistics about the participant's each individual visit such as mean visit duration and max visit duration (the longest visit duration), which can tell us, for example, whether they were engaging in more detailed reading, or whether they were looking quickly to "grab" information.

In addition, in order to gain further insights into participants' source use, Kruskal-Wallis tests were conducted to examine the statistical significance of differences in time spent (*total visit duration*) and number of visits (*visit count*) on each source material and the instructions (RQ2a); and into the relationships between the participants' looking behaviour and their performance levels (RQ2b), correlations were calculated between the eye-tracking results of *total visit duration, mean visit duration, max visit duration* and *visit count* (the independent variables), and the participants' scores on the TBEM-8 reading-to-write task (the dependent variable).

III. Stimulated recall analyses

The participants' verbal reports were first transcribed by the researcher (a native Chinese speaker). As mentioned earlier, the transcriptions were done based on the video recordings of the stimulated recall session, because it had the advantage of being able to link what the participant said to what was seen on the eye-tracker screen and thus facilitating the understanding of the stimulated recalls. The transcriptions were then segmented into a series of units (n=1,142), each of which was related to a single action or idea, such as a plan, or a comment relating to execution of the task, or an evaluation. Below is an example of the segmented units for analysis.

Now I am looking at the first paragraph, to remind myself that I should stick to the point that I have made in the first paragraph, so that the essay could be coherent. (P01-042)*
**P01= Participant 1, 042=the 42th segment in this participant's transcript*

A coding framework was developed based upon the ten categories of cognitive processes proposed in Section 2.4.2, which are presented in Table 3.3 (shown on the next two pages), with examples from the 16 participants' stimulated recalls. It should be noted that the cognitive processes of *text interpretation* and *monitoring* were categorised into several subprocesses so that different types of cognitive processes within these categories can be investigated individually. During the coding process, some segments in the participants' verbal reports did not fall into the nine categories, and therefore two additional codes were arrived at: "*commenting*", when the participant made comments either on the quality of their writings or on their reading or writing processes; "*transcribing*", when the participant reported issues related to their keyboarding skills.

The 1,142 transcript segments were uploaded in Atlas.ti 8 (a qualitative data analysis software) for coding. To ensure the reliability of the coding process, the researcher and a second coder (who were both native Chinese linguists, specialised in language testing and had experience in coding verbal protocols) first applied the coding scheme to one of the 16 transcripts (147 segments), followed by a discussion of the segments' codings on which they disagreed, and then refined the working definitions

Table 3.3: Coding framework for stimulated recalls

Cognitive processes/subprocesses (Codes)		Working definitions of cognitive processes/subprocesses	Examples
Text interpretation (TI)	TI-1	Participants read the instructions	"I was reading the instructions."
	TI-2	Participants read the source texts	"I went back to read the second source material."
	TI-3	Participants read the text-written-so-far	
Task representation (TR)		Participants create an understanding of the task	"I was reading those sentences, the task wants me to describe that event, and then analyse that situation."
Macro-planning (MacP)		Participants plan for writing goals and content	"Then I started to read the source texts and think about how to write this essay at the same time, for example, how should I start the first paragraph, how to write transition sentences, and how to write the ending paragraph."
Organising (O)		Participants organise/prioritise the ideas to be put in the text; use strategies to understand the source texts.	"Later I summarise the main points in media opinions, which I think, can be grouped into two categories, that is, one is about the impacts on Apple itself, and the other is about the impacts on the whole technology industry."
Connecting and generating (CG)		Participants connect ideas in the source texts with their own, generating links between them or new meaning.	"Here I was giving my opinions on Job's resignation, basing on people's comments on it in the texts."

Selecting (S)		Participants select ideas or information from the source materials/memory.	*"I used the 'vision' in the key concepts and expressions box."*
Micro-planning (MicP)		Participants plan for the text that is about to be produced.	*"Now I wanted to start the paragraph with a sentence which describes Job's personalities, or something that happened to him."*
Translating (Trans)		Participants transcribe abstract ideas into linguistic forms.	*"I was writing that sentence."* *"Then I began to write the first sentence."*
Monitoring (M)	M-1	Participants check the mechanical accuracy of the text produced (e.g., spelling, syntax).	*"I went to read the instructions again to make sure I wasn't deviating from the topic."*
	M-2	Participants check higher-level aspects of text quality such as argument and coherence.	*"I sometimes look back to check whether there is any grammar problem or whether the collocation is right."*
Revising (R)		Participants make a revision of their writings.	*"I changed the structure of the sentence in the middle of the second paragraph."* *"I found a mistake of grammar, and then revised it."*

of the cognitive processes after which a final coding was arrived at. The two coders then coded another four participants' transcripts. Among the 390 instances (20% of the total instances) of coding, 333 instances (85% agreement rate) were agreed by both coders, achieving a good inter-coder reliability value of Cohen's Kappa=0.833 with $p<0.001$. The researcher then coded the remaining segments of transcripts. A total of 1,956 instances of cognitive processes were obtained from the 1,142 segments of the participants' stimulated recalls.

In order to investigate what cognitive processes participants employed while completing the TBEM-8 reading-to-write task (RQ1), the number of occurrences for each cognitive process/sub-process was calculated and divided by the total instances in the stimulated recalls and then presented as percentages in the results chapter. Also, each type of cognitive processes was illustrated in detail, with quotes from participants' stimulated recalls. Furthermore, the frequency counts were used as the independent variables to explore relationships (Spearman's rank-order correlations) between the participants' cognitive processing and their scores (the dependent variable) on the TBEM-8 reading-to-write task (RQ1a).

In summary, this section has presented the research design of Study I, which sought to investigate test-takers' cognitive processes while completing the TBEM-8 reading-to-write task by using a combination of the eye-tracking and stimulated recall techniques. It starts with an explanation of the participants, followed by a detailed description of the equipment (Tobii TX300 eye-tracker) and the instrument (the TBEM-8 reading-to-write task), and the data collection procedure. It then describes the procedures set up for data analysis, beginning with the scoring of the participants' written products and moving on to the eye-tracking visualisation (heat-maps) and four eye-tracking metrics to be examined, and the transcribing, coding and analysis of the stimulated recall protocols. In the next section, methods for Study II will be introduced.

3.4 Study II: Reading-to-write process questionnaire – Methods

This section introduces the second source of data for the research. It was decided, based on the discussion of methodology in Section 3.2, to use a reading-to-write process questionnaire (see Section 3.2.2 for details) developed by Chan (2013) to elicit the participants' cognitive processes while completing the TBEM-8 reading-to-write task, so as to offset the drawbacks of eye-tracking and stimulated recall methods. This set of data complements and triangulates the data collected in the eye-tracking and stimulated recall study.

Section 3.4.1 first describes in detail a pilot study conducted before the main study and revisions of the pilot reading-to-write process questionnaire. Section 3.4.2 then introduces the recruitment of participants and their background information. Section 3.4.3 describes the procedures for data collection. Finally, methods for data analysis are presented in Section 3.4.4.

3.4.1 Pilot study

First, in order to test the readability of the preliminary version of the reading-to-write process questionnaire, it was first trialled with two Chinese Master's students at Lancaster University. They were encouraged to read through the instructions and source materials before they went on to look at the questionnaire, during which they were asked to identify items that were unclear to them.

According to their feedback, 12 items were modified (see Appendix F for a list of the modified items). For example, the word "text" in Item 1.2, i.e., "I thought of what I might need to write to make my text relevant and adequate to the task", was replaced by the word “essay”, because both of the students claimed that they were uncertain about the meaning of “text” here. Similarly, other items which had the same issue as Item 1.2 were reworded accordingly. Another way of modifying the items was to add examples into the original sentence to facilitate the understanding of the item, for example, in Item 4.8, the students said that they had not much idea of what the word “coherent” referred to, and so “e.g. appropriate use of topic sentences, connectives, etc.” was added.

After the trial, the modified version of this questionnaire was piloted with 81 Chinese third-year undergraduate students (a representative sample of the target population of the TBEM-8 reading-to-write task), all majoring in business English, from two universities in China. They were invited to complete the TBEM-8 reading-to-write task as an assessment of their reading-to-write proficiency, which was one of the main themes of an academic writing course they attended at the universities. The task was conducted in a classroom setting, in which the students were first given a lecture (lasting about 45 minutes) by their course lecturers (linguists specialised in language testing), on the nature and types of integrated writing tasks. After the lecture, the students then went on to complete the TBEM-8 reading-to-write task, which were delivered through the paper and pencil testing method. When they finished the task, they were asked to respond to the reading-to-write process questionnaire, as a reflection of their cognitive processes employed during task completion.

A total of 81 responded questionnaires were collected. Four of them were considered as invalid ones due to participants’ insufficient responses to the items (less than 30 items were responded), and discarded from further analysis. The remaining 77 questionnaires were submitted to a series of reliability and item analyses through SPSS. The results are presented below.

I. Internal consistency of each category of cognitive process

As described in Section 3.2.2, the questionnaire was designed to measure seven types of cognitive processes, i.e., *task representation, macro-planning, text interpretation, connecting and generating, organising, low-level editing (monitoring and revising)* and *high-level editing*. These types of cognitive processes were grouped into five hypothesised phases of academic writing: *conceptualisation, meaning and discourse construction, organising, low-level monitoring and revising* and *high-level monitoring and revising*. In order to understand the extent to which each group of questionnaire items reliably measured the same type of cognitive processes, a Cronbach's alpha was run to assess the internal consistency of these items. The overall reliability of each of the seven categories of cognitive processes, and each of the five hypothesised writing phases were obtained. Results are presented in Table 3.4. Items whose item-total correlations were lower than 0.30 are highlighted grey and correlations lower than 0.20 are highlighted yellow.

Overall, the results showed that all the five writing phases achieved a Cronbach's alpha of over 0.50 or above, ranging from 0.53 to 0.86, indicating a moderate to high level of internal consistency for each component of the questionnaire. Among the seven categories of cognitive processes, items designed to measure the processes of *low-level editing* (r=0.85) and *high-level editing* (r=0.86) achieved high levels of reliability, while items assigning to *task representation* (r=.321) and *connecting and generating* (r=.48) did not report satisfactory internal reliability of 0.50 or above. Out of the 42 individual items, 11 items did not yield satisfactory item-total correlations of 0.30 or above; four items (Item 1.4, Item 1.1, Item 4.4 and Item 2.7) reported item-total correlations that were lower than 0.20.

Table 3.4: Reliability statistics of the pilot questionnaire (42 items)

Item No.	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted	Cronbach's Alpha
Conceptualisation phase					
Task representation					
1.4	7.70	2.633	.057	.427	
2.5	7.94	1.509	.242	.092	.321
4.3	8.00	1.579	.252	.068	
Macro-planning					
1.2	13.03	6.586	.274	.601	
1.3	13.57	5.982	.367	.558	
1.5	13.67	5.930	.389	.548	.613
2.11	14.43	5.529	.389	.547	
4.5	14.46	5.345	.416	.531	
Overall reliability					.607
Meaning and discourse construction phase					
Text interpretation					
1.1	23.90	9.910	.157	.598	
2.1	24.09	8.373	.440	.505	
2.2	24.48	7.674	.394	.517	
2.3	24.25	8.767	.396	.524	.590
2.4	23.87	9.114	.429	.523	
2.6	24.14	8.440	.232	.593	
4.4	24.05	9.945	.179	.590	
Connecting and generating					
2.8	7.22	2.043	.226	.503	
2.10	7.57	1.538	.366	.261	.483
4.2	7.47	1.779	.320	.352	
Overall reliability					.639
Organising phase					
Organising					
2.7	19.61	5.036	.191	.528	
2.9	19.85	4.731	.316	.473	
3.1	19.50	5.021	.224	.513	.534
3.2	19.85	4.210	.366	.442	
3.3	19.99	4.698	.207	.529	
4.1	19.51	4.418	.394	.433	
Overall reliability					.534
Low-level monitoring and revising phase					
Low-level editing					
4.11	18.49	13.296	.351	.877	
4.13	18.72	10.992	.719	.815	
4.14	18.82	10.573	.693	.820	
5.6	18.57	12.812	.439	.864	
5.8	18.71	10.576	.828	.794	.854
5.9	18.64	10.431	.841	.790	
Overall reliability					.854
High-level monitoring and revising phase					
High-level editing					
4.6	39.92	33.807	.460	.856	
4.7	40.36	30.525	.633	.844	
4.8	40.32	31.126	.658	.843	.862
4.9	40.54	30.732	.593	.847	
4.10	40.00	32.427	.509	.853	

4.12	40.71	32.395	.399	.862	
5.1	40.05	33.011	.484	.855	
5.2	40.33	31.770	.627	.846	
5.3	40.30	30.987	.678	.842	
5.4	40.46	31.772	.551	.850	
5.5	40.09	32.165	.526	.852	
5.7	40.63	32.902	.410	.860	
Overall reliability					.862

Item 1.4 reads as "*I was able to understand the instructions for this writing task very well*" (while reading the task prompt). It only reported an item-total correlation of 0.57, indicating that the participants did not respond to this item similarly as how they responded to the other two items (Items 2.5 and 4.3) in this group. The reason may be that, Item 2.5 and Item 4.3 were designed to elicit participants' answers to questions asking whether they read the instructions at different phases of writing. It was about a participant's behaviour during task completion. While Item 1.4 was actually eliciting responses as to whether the participants understood the instructions. It was about results of a participant's reading behaviour. Therefore, discrepancy between these items may occur, but considering that Item 4.1 was useful in understanding how well the participants understood the instructions, it remained in this group of items.

Item 1.1 reads as "*I read the task prompt (i.e., instructions) carefully to understand each word in it*" (while reading the task prompt). This item was more likely to measure the participants' task representation process, as Items 2.5 and 4.3 did, rather than the process of text interpretation, although the wording of this item contained "read" and "carefully". Therefore, it was regrouped into *task representation*, whose internal consistency then improved from 0.32 to 0.44 (item-total correlation for Item 1.1 increased from 0.16 to 0.30; item-total correlation for Item 1.4 increased from 0.6 to 0.21).

Item 4.4 reads as "*I selectively re-read the source texts*" (while writing the first draft). This item was designed to measure the process of text interpretation. It reported a low item-total correlation of 0.18. The low correlation may be because this item and the other items in this group were measuring different types of text interpretation process at different phases of writing. More specifically, Item 4.4 was meant to measure selective reading skills such as scanning and search reading, while other Items such as Item 2.1 and Item 2.4 were meant to measure careful reading process. According to the eye-tracking and stimulated recall data (see Chapter 4 and 5 for details), the participants adopted different reading approaches at different writing phases (more careful global reading occurred while reading the source materials before writing, and more expeditious local reading occurred during writing), and this may lead to the discrepancy in the participants' responses to Item 4.4 and other items in this group. Item 4.4 remained in this group of items, and after removing Item 1.1 from *text interpretation,* the internal consistency of this type of cognitive process improved to 0.60.

Item 2.7 reads as "*I prioritised important ideas in the source texts in my mind*" (while reading the source texts). It was meant to measure the process of organising, however, it did not yield a satisfactory item-total correlation as other items in *organising* did. This may be due to the fact that two types of organising process were involved in completing the task, as evidenced by the stimulated recall data (see Chapter 5 for details), one of which is using strategies to understand and organise the structure of the source materials, and the other is to think about the structure of the participants' own text. Item 2.7 was designed to measure the former type of organising, while other items such Item 3.1, Item 3.2 and Item 3.3 were meant to measure the latter type of organising. Both types of organising process were considered essential in participants' reading-to-write process, so Item 2.7 remained in this group of items.

To sum up, Item 1.4 and Item 1.1 grouped together into *task representation*. Item 4.4 and Item 2.7 stayed in the old groups. As stated before, there were also another seven items which did not yield satisfactory item-total correlations of 0.30 (but above 0.20), for examples, Item 2.8 in *connecting and generating*, Item 3.1 and Item 3.3 in *Organising*. However, considering the relatively small sample size of the pilot study, and the small number of items in each cognitive process component, plus the possibility that different subcategories within some of these cognitive processes (e.g. *organising*) may exist, those seven items remained in the groups they were originally assigned to at this stage of analysis.

II. Item discrimination

In order to know how well each item distinguishes between participants at different levels of engagement in cognitive processing, the students were first ranked according to their total scores on the questionnaire, and then a t-test was conducted to compare the mean scores of each item between students in the top fourth of the sample with those in the bottom fourth. Overall, the results showed that the mean scores on all items in the questionnaire were statistically significantly different (*p*-values were less than 0.05) between the two groups of students, except the four items presented in Table 3.5 (see Appendix G for the results of all items).

Table 3.5: Independent samples t-test results on Item 2.1, Item 2.11, Item 4.5 and Item 5.7

Levene's Test for Equality of variances				t-test for Equality of Means			
		F	Sig.	t	df	Sig (2-tailed)	Mean Difference
Item 2.1	Equal variances not assumed	6.990	.012	1.964	31.469	.058	.586
Item 2.11	Equal variances not assumed	6.050	.018	1.527	31.194	.137	.481
Item 4.5	Equal variances assumed	.970	.331	1.988	39	.054	.567
Item 5.7	Equal variances assumed	.000	.991	1.851	39	.072	.495

Item 2.1 reads as "*I read through the whole of each source text carefully*" (at the stage of reading source materials). This item was designed to measure the process of text interpretation, i.e., reading different types of materials in the TBEM-8 reading-to-write task. More than four-fifths (83.1 percent) of the participants chose either "strongly agree" or "agree" in this item (see Figure 3.8). This is in accord with the findings in the eye-tracking and stimulated recall study that the majority of participants did spend much time reading the source materials during task completion, which may be the reason that Item 2.1 did not distinguish well between the higher- and lower-scoring participants.

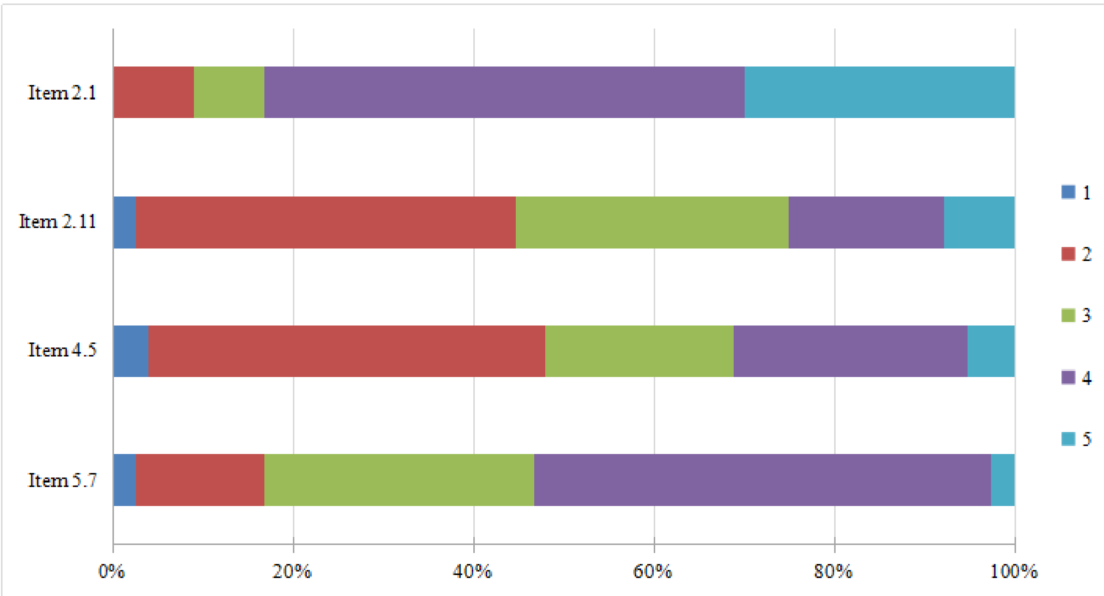

Figure 3.8: Stacked bar chart of the participants' responses to Item 2.1, Item 2.11, Item 4.5 and Item5.7 (5=strongly agree; 4=agree; 3=no view; 2=disagree; 1=strongly disagree)

Item 2.11 reads as "*I changed my writing plan (e.g. structure, content etc.)*" (at the stage of reading source materials). This item was to measure the process of macro-planning, i.e., planning for writing goals and content at the macro-level. More than two-fifths (42.1 percent) of the participants disagreed that they revised their writing plans while reading the source texts, and 30.3 percent of participants claimed that they had no view on this item. It appeared that the participants had a negative tendency towards the response to this item. This is somewhat understandable, as the stimulated recalls showed that the participants had only generated an initial, rough plan at this phase of writing, after which they began to write the essay, and it was during the writing phase that they either revised their original plan or consolidated it as their representation of the task and the source materials became clearer. The participants' responses to **Item 4.5** (see Figure 3.1), "*I changed my writing plan (e.g. structure, content etc.)*" (at the stage of writing the first draft), to some extent, proves this speculation. Although the number of participants who claimed they did not change their plans for writing remained stable, the percentage of participants who reported that they did revise the writing plans increased from 17.1 to 26.0 percent compared to Item 2.11. These two items (Item 2.11 and 4.5) were dropped from the questionnaire in the main study due to some practical constraints, which will be discussed later in this section.

Item 5.7 reads as "*I checked the possible effect of my writing on the intended reader*" (after writing the first draft). More than half of the participants (50.6 percent) claimed that they considered the target readership when monitoring the first draft, and 29.9 percent of participants said they had no view on this item. This is worth noting as in the stimulated recalls (see Chapter 5 for details), almost no participants reported that they engaged in this type of monitoring, and only one participant recalled thinking about the readership throughout task completion. This obvious disagreement between the findings in stimulated recall and questionnaire data is very likely due to the wording of this item. More specifically, the noun phrase "intended reader" may result in different interpretations between the participants and the researcher. According to the comments from the participants' lecturers, they had not specifically paid attention to the "readership" issue in their writing course, instead, in a testing-oriented environment in most Chinese universities, the students were often taught to write essays tailored to the raters' preferences. In other words, the "intended reader" in the students' mind is more likely to be the raters who are going to mark their written products, rather than some imaginary readers who would read their essays (in this study, the professor of course strategic management). Therefore, this item and another two items (Items 1.3 and 4.12) relating to "intended reader" were dropped in the main study questionnaire.

III. Revisions according to the eye-tracking and stimulated recall data

As shown in the above two statistical analyses, data from the eye-tracking and stimulated recall study has provided extra knowledge on the understanding of the results of questionnaire data; furthermore, it fed into some other revisions of the pilot questionnaire.

First, as will be discussed in Chapters 4 and 5, the process of selecting played an essential role in the participants' reading-to-write process, when they were either selecting ideas from the source materials to combine with their prior knowledge, or selecting specific information (e.g., a specific word) in the source materials to support their own writing. However, the selecting process was not well represented in the pilot questionnaire. It was then decided to add several items into the questionnaire to measure the process of selecting.

A total of five items were added for *selecting*. Two of them were added into the "while reading the source materials" stage, and read as "*I used the materials to help me get ideas on the topic*", and "*The materials helped me choose an opinion on the issue*". The other three items were added into "while writing the first draft" stage, and read as

"*I used some of the ideas from the source materials in my essay*", "*I paraphrased part(s) of the source materials in my writing*" and "*I copied phrases and sentences directly from the source materials into my essay*".

Besides the above five items, another item was added to measure the process of organising, and reads as "*I used the materials to help me organise my essay*". This was meant to examine if representation of the source materials helped the participants to think about the structure of their own writing.

IV. Some practical constraints

Based on the information gathered through the above analyses, it was decided (1) to put Item 1.1 and Item 1.4 together into the group of items which examined the process of task representation, (2) to drop Item 1.3, Item 4.12 and Item 5.7 because these items were suspected not being able to assess the process they were designed to measure, and (3) to add six new items to measure the process of selecting and further investigate the process of organising. After these revisions, the number of items in this reading-to-write process questionnaire increased from 42 to 45. The revised pilot questionnaire was ready to be used in the main study. However, after contacting with the course lecturers from the two universities where the researcher was going to collect data (see Section 3.4.3 for details of data collection), some practical constraints showed up and resulted in several further revisions to the pilot questionnaire.

First, it was decided to deliver the TBEM-8 reading-to-write task and the reading-to-write process questionnaire via the computer in one university (most of the participants were recruited from this university). Item 2.6, i.e., "*I took notes on or underlined the important ideas in the source texts*", thus seemed inappropriate in light of the task delivery method. It was dropped in the main study questionnaire .

Second, students participating in the main study had different English language proficiency levels. In order to ensure that all of them had the same understanding of each questionnaire item, the questionnaire was translated to Chinese and the translation was checked by one of the course lecturers (Chinese native speaker who had extensive experience in teaching English as a foreign language). So, a Chinese version of the questionnaire was administered in the main study.

Third, it was decided to further remove another four items to keep the number of questions within 40. This was because (1) the TBEM-8 reading-to-write task was integrated into an end-of-term English test as its writing component in one University; the participants were required to complete this English test in 120 minutes, after which they were expected to respond to the questionnaire; considering that the participants would have had spent two hours doing high cognitive work, they might be less motivated and willing to fill in a questionnaire if it was unpleasantly long; the lecturer from that university insisted that the number of items should not be exceeding 35, but finally the number was compromised to 40, and (2) it was estimated, at the beginning of data collection, that only about 100 participants could be recruited in the main study, so the number of items (variables) in the questionnaire should not be too large, so that some potential statistical analyses could be run to assess quality of the questionnaire.

The last four items removed from the questionnaire were Item 1.5, Item 2.2, Item 2.11 and Item 4.5. Item 2.2, "*I read the whole of each source text more than once*", was dropped, because there were already two other similar items (Item 2.1 and Item 2.4) which examined the use of careful reading approach while reading the source texts, and according to the eye-tracking data, there is ample evidence that most of the participants carefully read through the source materials more than once before they started to compose the essay. Items 2.11 and 4.5, as discussed in the previous item discrimination analysis, did not distinguish well between different levels of participants, and quite a number of participants chose “no view” as their responses to these two questions (designed to measure macro-planning process). Item 1.5, i.e., “*After reading the prompt, I thought about the purpose of the task*”, was also designed to measure the process of macro-planning, but the phrase “the purpose of the task” seemed to be ambiguous and may cause different interpretations of the meaning it referred to among different participants, so it was removed as well. After this series of revisions, there was only one item (Item 1.2) left to measure the macro-planning process, which may be one limitation of the main study that this process was not fully represented in the questionnaire. Considering the strong relationship between the task representation and macro-planning process, which was exemplified in the stimulated recall data (see Chapter 5 for details), these two processes were combined as one and is the main process participants engaged in during the conceptualisation phase of writing.

Through the above four stages of revisions to the pilot questionnaire, the version of questionnaire to be used in the main study was finalised (See Appendix H for the main study questionnaire and Appendix I for an English version of it). The new structure of the reading-to-write process questionnaire is presented in Table 3.6; new items added into the questionnaire were highlighted grey.

Table 3.6: Structure of the main study questionnaire (40 items)

Phases of academic writing	Cognitive Processes	Stages					No. of items
		Reading task prompt	Reading source materials	Before writing	Writing the 1st draft	After writing the 1st draft	
Conceptualisation	Task representation and macro-planning	1.1 1.2 1.3	2.6		4.4		5
Meaning and discourse construction	Text interpretation		2.1 2.2 2.3		4.5		12
	Selecting		2.4 2.5		4.2 4.6 4.7		
	Connecting and generating		2.8 2.10		4.3		
Organisation	Organising		2.7 2.9 2.11	3.1 3.2 3.3	4.1		7
Low-level monitoring and revising	Low-level editing				4.13 4.14 4.15	5.6 5.7 5.8	6
High-level monitoring and revising	High-level editing				4.8 4.9 4.10 4.11 4.12	5.1 5.2 5.3 5.4 5.5	10

3.4.2 Participants

A total of 172 undergraduate students participated in the main study. The participants were all native Chinese learners of English and enrolled in a Business English program at two public universities in China. Of the participants, 120 were from University A and, at the time of data collection, were in their second year of study; 52 students were from University B and were in the third year of their study during data collection. With regards to gender, 8.6% were male and 91.4% were female (there has been a high percentage of female students majoring in business English in Chinese universities). Their ages ranged between 20 and 21 years old. Their English proficiency levels were estimated to be between CEFR B2 and C1 (based on their scores on the TEM-4 test, i.e., Test for English Majors-Band four, a national test for English majors in China; for those who had not sat the TEM-4 test, their scores on the end-of-term English test were referenced). It was believed that these participants formed a representative sample of the target population of the TBEM-8 reading-to-write task.

3.4.3 Data collection procedures

In University A, The TBEM-reading-to-write task was administered to the 120 students as a mid-term classroom assessment for an English academic writing course. The task was delivered via the computer in a multi-media classroom on campus. Immediately after the participants had completed the task, the reading-to-write process questionnaire was used to prompt them to report the extent to which they employed different types of cognitive processes throughout task completion. A total of 120 questionnaires were collected from these participants.

In University B, The TBEM-8 reading-to-write task was integrated into an end-of-term English test for a course of integrated English, as its writing component to assess the participants' reading-to-write abilities. This test was administered through the traditional paper and pencil tests method. It lasted about 120 minutes. After the participants finished the test, they were asked to complete the reading-to-write process questionnaire. 52 questionnaires were collected from these participants.

3.4.4 Analyses

A total of 172 questionnaires were collected from the participants. Two of the questionnaires were discarded because of insufficient completion (more than 10 items were left unresponded to), and the remaining 170 valid questionnaires were submitted to SPSS for further statistical analyses. The scoring of these participants' written products is first described below.

I. Scores on the TBEM-8 reading-to-write task

The researcher scored the 170 essays using the TBEM-8 reading-to-write task rating scale (see Section 3.3.4 for details). The lecturer from University B (had a PhD degree in language testing) randomly scored 50 (about 30 percent) essays. The inter-rater reliability between the two raters was calculated. As Shapiro-Wilk tests of normality indicated that both sets of scores' distributions were not statistically significantly different from normality (p>.05), Pearson product-moment correlations were run, achieving a good correlation coefficient of r=.85 (p=.00**).

II. Descriptive and inferential analyses on questionnaire data

First, a Cronbach's alpha was run to understand whether each group of items in the main study questionnaire (particularly the newly added group of *selecting*) reliably measured the same category of cognitive processes.

Second, in order to understand the extent to which participants employed the specified cognitive processes while completing the TBEM-8 reading-to-write task (RQ1), a frequency analysis was performed to know the percentage of participants choosing each number (1 to 5) for each question, and the agreement rate for each question was calculated by adding up the percentage of those who agreed and strongly agreed. Each category of cognitive processes will be examined individually in the order as Table 3.6 presents. Finally, a set of Mann-Whitney U tests was run to investigate if higher- and lower-scoring participants responded to each item in the questionnaire differently (RQ1a).

In summary, this section has presented the research design of Study II, in which a reading-to-write process questionnaire was used to complement and triangulate the data from the eye-tracking and stimulated recall study (Study I). A pilot study of the preliminary version of the questionnaire was first described in detail, with revisions of it based on several statistical analyses afterwards. It then introduces the background information of participants and the procedures for data collection, and finally, the methods for data analysis.

3.5 Summary

This chapter has mainly introduced the methodology for this research study. First, in Section 3.2, some methodological underpinning to each chosen research method was discussed to explain (1) why the combination of eye-tracking and stimulated recall methods was useful in Study I; and (2) why a reading-to-write process questionnaire was utilised (Study II) to complement the data collected in Study I. Sections 3.3 and 3.4 then present the details of data collection and analysis in each separate study (including the recruitment and background information of participants, the equipment and instrument, pilot studies, etc.).

The next three chapters, that is, Chapters Four, Five and Six will present the results from the analysis of the three sources of data: eye-tracking, stimulated recall and the reading-to-write process questionnaire.

CHAPTER 4 RESULTS I: EYE-TRACKING (STUDY I)

4.1 Introduction

This chapter looks at the findings from the analysis of eye-tracking data collected in the eye-tracking and stimulated recall study (Study I) described in Chapter Three. It begins with the results of a correlation analysis in Section 4.2, which demonstrates the relationship between the participants' performance on the TBEM-8 reading-to-write task and the IELTS test (particularly of the reading and writing components respectively). Section 4.3 then presents a heat map to illustrate the overall distribution of the participants' attention throughout task completion. Section 4.4 reports the results on the four eye-tracking measures defined and explained in Chapter Three (see Section 3.3.4 for details), including *time to first fixation* (Section 4.4.1), *total visit duration* (Section 4.4.2), *visit count* (Section 4.4.3) and *visit duration* (Section 4.4.4). Section 4.5 investigates the relationships between the results of eye-tracking measures (*total visit duration*, *mean visit duration*, *max visit duration* and *visit count*) and participants' performance on TBEM-8 reading-to-write task. Finally, a summary of this chapter is provided in Section 4.6.

4.2 Relationships between test-takers' performance on TBEM-8 reading-to-write task and IELTS test

In order to gain an initial understanding of the associations between participants' reading-to-write ability and their reading and writing abilities, Spearman's rank-order correlations (*p*-values for Shapiro-Wilk tests of normality for some of the measures were less than 0.05) between the participants' scores on TBEM-8 reading-to-write task and their IELTS test scores were calculated. The correlation coefficients among these variables are shown in Table 4.1 below.

Table 4.1: Spearman's rank-order correlations between participants' scores on the TBEM-8 reading-to-write task and their IELTS test scores

Scores	1	2	3	4
1. TBEM-8 reading-to-write task	--	.78**	.68**	.59*
2. IELTS overall		--	.69**	.56*
3. IELTS reading			--	.26
4. IELTS writing				--

**. Correlation is significant at the 0.01 level.
*. Correlation is significant at the 0.05 level.

Overall, the results show that correlations among these four sets of scores were all statistically significant, showing a positive relationship, with coefficients higher than 0.50, at either the 0.01 or the 0.05 level except the one between IELTS reading and writing scores (r_s=0.26). A strong, significant positive correlation (r_s=0.78**) was found between participants' scores on the reading-to-write task and IELTS overall scores, while the correlations between the reading-to-write task scores and IELTS reading/writing scores (r_s=0.68**, r_s=0.59* respectively) are less strong though still relatively robust, suggesting that the reading-to-write task was more related to participants' overall language proficiency, than to stand-alone reading or writing abilities considered individually. Also, it should be noted that the correlation between participants' reading-to-write task scores and their IELTS reading scores (r_s=0.68**) is significantly positive, however, as shown in Table 4.1, no significant correlation was found between their IELTS reading and writing scores. This may be partly due to the in of IELTS writing component which is, essentially, an independent writing test rather than an integrated one which requires integration of different skills (especially reading-writing skills), and this also implies that the participants' L2 reading ability may be, to some extent, a factor that had impact on their performance on an integrated writing task such as the one examined in this study.

These findings concur with Esmaeili's (2002) study of the role of reading in a reading-to-write task, in which he investigated participants' writing strategies through a questionnaire and interview taken after the task and found that reading played a critical role in participants' writing process concluding that "Examining participants' writing strategies, overall, reveals how writing involves reading. In fact, one can hardly view reading and writing as stand-alone skills" (p. 615). Another study of the importance of reading skills in integrated writing tasks was conducted by Plakans (2009a), in which she used think-aloud protocols and interviews to look into the reading strategies of 12 participants who completed an integrated academic writing task. Results indicated that reading was actively involved and had an effect on participants' writing performance, and there were differences in choice of strategies between lower- and higher-scoring participants, for example, higher-scoring writers used more mining (selecting words from source texts for use in writing) and global strategies, whereas lower-scoring writers used more word-level but fewer global strategies such as skimming and scanning, which corresponded to the results found in some earlier studies (Carrell,1989; Koda, 2005).

Esmaeili's (2002) and Plakans's (2009a) process studies concluded that reading plays an important role in completing an integrated writing task, however, other researchers have done some correlation studies which showed somewhat contradictory results. Watanabe (2001) investigated L2 writers' performance on a reading-to-write task by correlating the scores with those on an independent writing task and a reading test, and found that participants' performance on the independent writing task was a stronger predictor of their scores on the reading-to-write task, while the reading scores had a low correlation and the predictive power of which was presumably due to participants' general language proficiency rather than their reading abilities. Ascensión Delaney (2008) also looked into the relationships between participants' scores on a reading test and two kinds of integrated writing tasks, a short summary and a response essay. Results showed that "reading-to-write scores were weakly related to reading ability" (p. 147). These findings suggest that reading was not a major factor accounting for the scores in a reading-to-write task, which is, to some extent, opposed to the results of the present study. One possible reason for this contradiction is: in Watanabe's (2001) and Ascensión Delaney's (2008) studies, results may be impacted by the use of holistic scoring to establish the relation between scores and the skills required to finish the task, in other words, reading skills may not be properly measured in scoring but they are necessary to complete the task, as demonstrated in Esmaeili's (2002) and Plakans's (2009a) studies.

4.3 Heat map output

To gain insights into test-takers' looking behaviour while completing the TBEM-8 reading-to-write task, the 16 participants' eye movements were recorded by the eye-tracker as they responded to the task. An absolute duration heat map (see Figure 4.1)

Figure 4.1: Heat map output for the TBEM-8 reading-to-write task

was first generated based on these eye traces to provide an initial response to the second research question (RQ2), i.e., to what extent do test-takers engage with the source materials (including the task instructions). It illustrates the amount of time participants spent on each individual AOI. Red indicates the longest time and green the least, with varying levels in between.

As can be seen in this figure, participants' attention covered all the seven parts of the task and few fixations were found outside these areas. Therefore, the total visit duration on these seven AOIs can be roughly counted as the total time participants spent completing the task. Overall, the majority of participants' attention was on the answer sheet (where they wrote the essay), which is understandable as this is ultimately a writing task that requires test-takers to produce a written product. Most of the area in Source 4, the only graphic input, remain transparent, indicating that very limited attention was paid to this source.

To look at each AOI separately, first, in the area of task instructions, attention within the texts seems to be evenly spread, with a light intensity of focus in the middle of this area where content requirements for the task are stated, i.e., "*In the essay, you should describe the event, analyze the situation and comment on the impact of Job's resignation*". This suggests a careful and more global (i.e., to handle the majority information in this section) reading approach to the task instructions by these participants. In the area of Source 1, the visualisation shows a similar picture to that in the instructions. Participants' attention was almost equally scattered over every sentence, though they seemed to spend relatively more time reading this material than the instructions. In Source 2, the right half of the texts appears to receive more attention compared with the other half. This may because only news headlines were provided in the left half, with less useful information that participants may take in their writing. Again, the visualisation on the right half shows a similar visual pattern which indicates more careful global readings may be involved. Less attention was given to Source 3, as compared to the first two materials, and participants seemed to adopt an expeditious approach of reading as their attention was unevenly distributed within the texts in this material.

As mentioned earlier, little attention was directed towards Source 4. These participants appeared to only attend to the limited texts in the picture. Interestingly, the face of the cartoon Steve Jobs also received quite a high amount of attention. Participant 4 gave an explanation in his stimulated recall, "*When I am thinking, I prefer to look at places where there are no texts, because texts may interrupt my thought processes…I just unconsciously went to look at the picture, particularly the face of the little man, although I did not use any information in the picture*". Source 5 (key concepts and expressions) attracted considerable attention from participants, who reported that the words and phrases in this material either provided lexical support when they were looking for a particular word during writing or helped them to generate new ideas for the text to be produced in their writing.

Last, the answer sheet received the most attention among the seven AOIs. It can be seen in the heat map that relatively more time was spent in the upper part of this area, especially where the starting sentence of the first paragraph was composed. This is probably due to the macro-planning process that may be involved at this point during which participants may need time to think of a general plan for composing their essay. The intensity of focus in this area gradually decreased as participants went on composing, and the last several sentences of the ending paragraph received much less attention from participants as compared to the other parts of the essay (the majority of participants fully used the space provided to write the essays). This is understandable because, in the last paragraph, participants may only need to summarise the main points that had been described in previous paragraphs rather than generate new ideas.

4.4 Eye-tracking metrics

The results on the four eye-tracking metrics are presented in this section to gain further insights into how much the 16 participants engage with the source materials (including the instructions).

4.4.1 Time to first fixation

Time to first fixation measures the amount of time from when an AOI was shown on the screen until the start of the first fixation within it. Table 4.2 shows the results on this metric by the 16 participants. For reasons of interpretability, the time-related metrics are all presented in seconds, though original output was expressed in milliseconds due to the sampling rate of the eye-tracker.

Table 4.2: *Time to first fixation* on each individual AOI by participant

Areas of interest	Participant																	
	1 (F)	2 (F)	3 (M)	4 (M)	5 (F)	6 (M)	7 (F)	8 (F)	9 (F)	10 (M)	11 (F)	12 (F)	13 (F)	14 (F)	15 (M)	16 (F)	Min	Max
	Sec	Sec	Sec	Sec	Sec	Sec	Sec	Sec	Sec	Sec	Sec	Sec	Sec	Sec	Sec	Sec	Sec	Sec
Instructions	0.9	97.2	0.6	1.1	1.3	4.0	0.6	1.5	5.3	5.1	3.8	5.6	1.1	1.8	0.8	3.9	0.6	97.2
Source 1	0.5	1.1	0.4	0.9	17.5	1.3	2.8	0.7	0.9	1.2	0.6	47.3	0.9	2.6	0.6	1.3	0.4	47.3
Source 2	3.5	0.7	128.4	0.6	15.5	2.0	135.5	45.1	31.6	11.6	2.4	3.5	76.4	1.1	60.8	2.4	0.6	135.5
Source 3	60.4	156.5	3.7	0.4	12.6	12.5	5.0	41.4	106.6	12.0	150.2	67.7	163.4	3.8	61.1	2.9	0.4	163.4
Source 4 (picture)	181.3	490.8	1.5	147.1	10.1	3.5	209.3	37.7	136.1	1005.6	1.9	1.5	356.8	99.0	120.2	94.7	1.5	1005.6
Source 5	4.6	183.4	1.7	1.8	11.0	7.3	6.0	1.2	4.4	68.5	37.5	1.8	1.8	1.2	1.2	5.5	1.2	183.4
Answer sheet	5.0	50.4	5.6	146.2	12.4	5.9	7.3	1.0	2.1	1.8	198.9	4.5	2.8	1.4	1.5	8.3	1.0	198.9

It can be seen in the table that although participants approached the task quite differently in terms of the time when they looked at each AOI for the first time, a major pattern that seems to emerge from these measures (and by investigating participants' eye-movement recordings) is that participants started responding to the task by having a quick and short browse of all the seven parts of the task, and then went back to read the task instructions and the source materials one after another in a slow and careful manner. Figure 4.2 shows the heat map output for the first 30 seconds of recording of

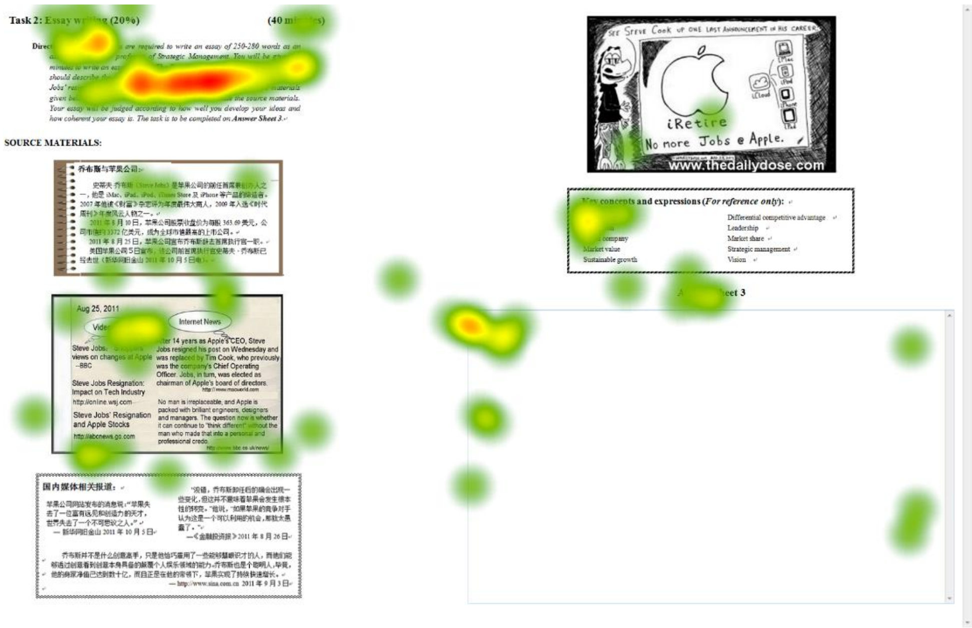

Figure 4.2: Heat map output for the first 30 seconds of recording of Participant 6

Participant 6, which can be used as an example to illustrate this common pattern. During the first half minute, this participant's attention was scattered loosely over different areas of the task, with a relatively strong focus on the task instructions. This is natural that at the beginning of the task completion test-takers may spend some time having a quick browse at each part of the task in order to get a general idea of what different parts are about, especially when they are not familiar with the task type. Participants then typically returned to read the instructions in detail to gain further understanding of the task.

There is only one participant, Participant 2, who did not follow this major pattern. Figure 4.3 shows how she approached the task in the first 90 seconds of the recording. Instead of reading the task instructions first, she started by moving straight to the first and second source texts, and the reading approach she adopted seemed to be more expeditious and local as her attention was unevenly spread within these materials. It was after these 90 seconds when she read the instructions for the first time and then went on reading through the other source materials. Also, it is interesting to note that it took Participant 10 nearly 17 minutes (maximum figure for Source 4) before he had the first fixation on Source 4 (the picture) and it seems that this material needs, on average, more time to attract participants' attention as compared to other source materials.

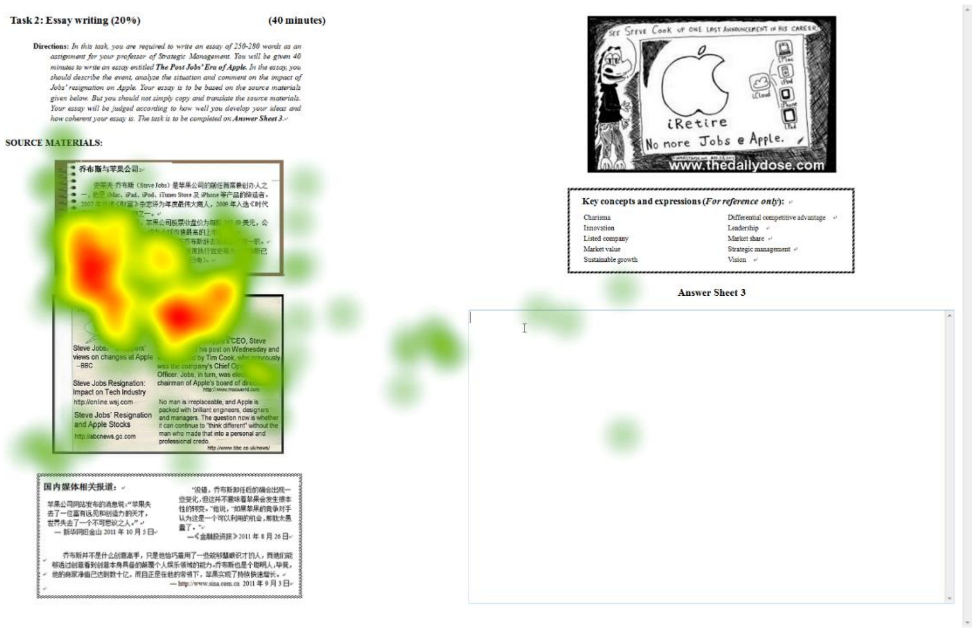

Figure 4.3: Heat map output for the first 90 seconds of recording of Participant 2

4.4.2 Total visit duration

Total visit duration measures the duration of all visits within an AOI or AOI group. As discussed in Chapter Three (see Section 3.3.4), this metric is considered as a measure of the overall amount of participants' attention on a specified AOI during task completion because not only are the duration of all fixations measured, but time spent on saccades (movements between fixations) is added into the calculation of total processing time.

Table 4.3 presents the total visit duration on the task instructions and the five source materials by each participant. Overall, the participants spent, on average, 580.8 seconds (*SD*=117.8) reading these parts of the task. Source 2 seems to be the material at which participants had the longest stay, with a mean of 157.7 seconds (*SD*=76.1), which accounts for 27.2 percent of the total time spent on reading. Source 4, the picture, received the least attention from participants, with a mean of 18.8 seconds (*SD*=20.3). Participant 3 spent only 1.8 seconds looking at the picture throughout task completion, while Participant 12 spent the longest time (81.0 seconds), but she recalled in her protocols, "*I don't know why I always went to look at the face of that little dinosaur, it's quite attractive to me, maybe because that's an animal image...I looked for information in the 'key concepts and expressions' (Source 5) rather than the picture...*". It seems that her attention on the picture was more of an unconscious behaviour rather than a careful act trying to dig out useful information.

Table 4.3: *Total visit duration* on each individual AOI in the reading group by participant

Areas of interest	Participant 1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	Mean	*SD*	Min	Max
	Sec *%*	Sec *%*	Sec *%*	Sec *%*	Sec *%*	Sec *%*	Sec *%*	Sec *%*	Sec *%*	Sec *%*	Sec *%*	Sec *%*	Sec *%*	Sec *%*	Sec *%*	Sec *%*	Sec *%*		Sec	Sec
Instructions	131.5 *26.8*	67.0 *11.5*	142.7 *22.6*	113.4 *19.3*	153.9 *22.4*	107.6 *16.1*	73.8 *18.5*	120.5 *25.1*	73.8 *11.2*	94.8 *22.4*	128.7 *20.5*	188.0 *32.1*	174.8 *25.2*	108.1 *17.6*	93.6 *11.9*	59.7 *16.1*	114.5 *19.7*	*37.8*	59.7	188.0
Source 1	37.0 *7.6*	124.3 *21.4*	186.2 *29.5*	174.9 *29.8*	129.7 *18.8*	93.3 *13.9*	68.8 *17.2*	59.4 *12.4*	234.2 *35.5*	93.5 *22.0*	189.6 *30.2*	79.9 *13.6*	166.8 *24.0*	100.6 *16.4*	135.9 *17.2*	60.3 *16.3*	120.9 *20.8*	*56.6*	37.0	234.2
Source 2	122.0 *24.9*	292.8 *50.4*	119.1 *18.9*	109.5 *18.7*	151.6 *22.0*	292.5 *43.7*	143.1 *35.8*	113.0 *23.6*	178.3 *27.0*	96.0 *22.6*	138.9 *22.1*	85.2 *14.5*	203.4 *29.3*	130.1 *21.2*	296.1 *37.6*	51.3 *13.9*	157.7 *27.2*	*76.1*	51.3	296.1
Source 3	83.0 *16.9*	82.5 *14.2*	65.9 *10.4*	99.9 *17.0*	134.9 *19.6*	108.0 *16.1*	53.4 *13.4*	73.5 *15.3*	84.3 *12.8*	60.6 *14.3*	101.0 *16.1*	75.1 *12.8*	66.1 *9.5*	78.4 *12.8*	114.7 *14.5*	62.7 *16.9*	84.0 *14.5*	*22.3*	53.4	134.9
Source 4 (picture)	6.1 *1.2*	8.0 *1.4*	1.8 *0.3*	42.7 *7.3*	35.3 *5.1*	3.7 *0.6*	18.4 *4.6*	11.4 *2.4*	6.4 *1.0*	10.0 *2.4*	13.5 *2.1*	81.0 *13.8*	5.2 *0.8*	9.4 *1.5*	20.6 *2.6*	27.9 *7.6*	18.8 *3.2*	*20.3*	1.8	81.0
Source 5	110.5 *22.5*	6.6 *1.1*	115.5 *18.3*	46.0 *7.8*	83.1 *12.1*	63.8 *9.5*	42.0 *10.5*	101.6 *21.2*	83.1 *12.6*	69.2 *16.3*	56.9 *9.0*	77.1 *13.1*	78.6 *11.3*	188.3 *30.6*	127.7 *16.2*	108.0 *29.2*	84.9 *14.6*	*41.6*	6.6	188.3
Total	490.2 *100*	581.1 *100*	631.2 *100*	586.5 *100*	688.6 *100*	668.8 *100*	399.5 *100*	479.3 *100*	660.1 *100*	424.1 *100*	628.6 *100*	586.3 *100*	695.0 *100*	614.9 *100*	788.6 *100*	369.9 *100*	580.8 *100*	*117.8*	369.9	788.6

An average of 120.9 seconds (*SD*=56.6) were spent on Source 1, which provides a brief description of Steve Jobs and Apple Company, accounting for 20.8 percent of the total reading time. Task instructions received roughly the same amount of participants' attention (Mean=114.5; *SD*=37.8) as Source 1 in terms of the mean, but the distribution of total visit durations had a lower standard deviation, which indicates that each participant's time spent on the instructions tends to be somewhat more aligned than for Source 1. Source 3 was given less attention (Mean=84.0; *SD*=22.3) compared with the first two source materials. As regards Source 5 (key concepts and expressions), it should be noted that although the number of words in it is much less than that in Source 3, it received as much attention (Mean=84.9; *SD=41.6*) as Source 3.

To test the statistical significance of differences in time spent on each source material and the instructions (RQ2a), the total visit duration data were submitted to the Kruskal-Wallis test (a non-parametric procedure was used as the assumptions of normality and equal variances were violated). The number of words (one Chinese character counted as one word) in each AOI were controlled by dividing the total visit duration by the total number of words in each part of the task. Results are shown in Table 4.4 (Source 4, the picture, was not included in this test). Participants spent the most time on Source 5, followed by Source 2, Instructions, Source 1, and finally Source 3 (χ^2=56.68, df=4, p<0.001). The Mann-Whitney tests (see Table 4.5) were also conducted as post-hoc tests to compare the time differences between AOIs: there were significant differences among each AOI, with medium to large effect size, except between Instructions and Source 2 (p =.121).

Table 4.4: Results of Kruskal-Wallis test of *total visit duration* on different AOIs with number of words controlled

AOIs	Total visit duration (mean)	Number of words	Total visit duration (mean) with number of words controlled	Mean rank
Instructions	114.49	117	114.49/117 – **.98**	44.22
Source 1	120.91	213	120.91/213 – **.57**	26.72
Source 2	157.68	120	157.68/120 – **1.31**	50.78
Source 3	84.00	275	84.00/275 – **.31**	12.16
Source 5	84.87	24	84.87/24 – **3.54**	68.63

Table 4.5: Significant differences in *total visit duration* between different AOIs

Comparisons	Mann-Whitney U	Z	*p*	Effect size
Instructions – Source 1	43.000	-3.224	.001	.32
Instructions – Source 2	87.000	-1.551	.121	.08
Instructions – Source 3	.500	-4.838	.000	.73
Instructions – Source 5	16.000	-4.225	.000	.56
Source 1 – Source 2	26.000	-3.857	.000	.46
Source 1 – Source 3	47.000	-3.116	.002	.30
Source 1 – Source 5	13.500	-4.324	.000	.58
Source 2 – Source 3	3.000	-4.746	.000	.70
Source 2 – Source 5	24.500	-3.904	.000	.48
Source 3 – Source 5	8.000	-4.559	.000	.65

Table 4.6 shows the total visit duration on the AOI reading group (which contains the AOI instructions and the five source materials; time spent on this group was roughly counted as the total reading time) and writing group (which contains only the answer sheet, time spent on which was counted as the total writing time). It can be seen in this table that participants spent, on average, over a quarter (26.4 percent; 580.8 seconds) of their time in reading, and 73.6 percent (1623.1 seconds) in writing. Among the 16 participants, Participant 11 spent the largest proportion of time (41.3 percent) reading the instructions and source materials. This percentage seems to be high and may imply

Table 4.6: *Total visit duration* on the AOI reading group and writing group by participant

AoI Group	Participant 1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	Mean	*SD*	Min	Max
	Sec *%*	Sec *%*	Sec *%*	Sec *%*	Sec *%*	Sec *%*	Sec *%*	Sec *%*	Sec *%*	Sec *%*	Sec *%*	Sec *%*	Sec *%*	Sec *%*	Sec *%*	Sec *%*	Sec *%*		Sec	Sec
Reading	490.2 *22.3*	581.1 *27.4*	631.2 *32.4*	586.5 *27.2*	688.6 *27.8*	668.8 *31.4*	399.5 *17.9*	479.3 *20.6*	660.1 *31.0*	424.1 *17.9*	628.6 *41.3*	586.3 *25.0*	695.0 *28.7*	614.9 *31.7*	788.6 *27.7*	369.9 *17.4*	580.8 *26.4*	*117.9*	369.9	788.6
Writing	1706.8 *77.7*	1541.1 *72.6*	1314.8 *67.6*	1569.5 *72.8*	1792.7 *72.2*	1458.7 *68.6*	1827.4 *82.1*	1842.2 *79.4*	1467.5 *69.0*	1938.7 *82.1*	893.6 *58.7*	1759.3 *75.0*	1723.7 *71.3*	1326.4 *68.3*	2055.1 *72.3*	1752.1 *82.6*	1623.1 *73.6*	*286.7*	893.6	2055.1
Total	2197.0 *100*	2122.2 *100*	1946.0 *100*	2156.0 *100*	2481.3 *100*	2127.5 *100*	2226.9 *100*	2321.5 *100*	2127.6 *100*	2362.8 *100*	1522.2 *100*	2345.6 *100*	2418.7 *100*	1941.3 *100*	2843.7 *100*	2122.0 *100*	2203.8 *100*	*285.2*	1522.2	2843.7

that reading played an important role in this participants' reading-to-write process. By looking at this participant's stimulated recalls, it was found that she reported many instances when she was summarising the source materials and categorising them in order to use them in different parts of her own essay, for example, she said "*...so I re-read the first three materials and categorised them to decide in which paragraph of the essay their information can be put into…then I found that the content in the first material can be used in the first part of my essay…*". In contrast, Participant 16 spent the least proportion of time (369.9 seconds; 17.4 percent) on AOIs in the reading group: she gave much less attention to the first two source materials (see Table 4.3) compared to other participants, but spent almost the same amount of time (108.0 seconds) on Source 5, accounting for 29.2 percent of the total reading time. Again, this demonstrates that Source 5 may provide some particularly important information that participants deemed helpful while they were composing the essays. The reason that this participant spent relatively less time on reading the materials may be that she based her essay more often on her own knowledge rather than the information provided in the source texts, and when she went back to the materials to search for information, she, most of the time, was looking for mechanical support, such as the spelling of a particular word, for example, she recalled "*I was looking for the word 'resignation'*".

The above is an overall depiction of how much participants engage with different source materials in terms of the total time spent on each AOI during task completion. To investigate the data in more detail and identify possible patterns of participants' looking behaviour, the whole task completion was divided into three phases: before writing, during writing and after writing. 'Before writing' refers to the period during which participants get themselves ready for the writing process, which in this study involves, but is not limited to, familiarising themselves with the task environment, reading the instructions, reading the source materials and planning for writing goals. In practice, this period was considered to start from the time at which participants' first fixation appeared on the eye-tracker screen and to end at the moment they typed the first word on the answer sheet. 'During writing' is the major phase in task completion when participants compose a first draft of the essay, during which they are expected to integrate information from the source materials into their writing and translate thoughts into words. Finally, the 'after writing' phase is when participants finish the first draft and make revisions to their writing.

Table 4.7 shows participants' total visit duration on each part of the task (not including the answer sheet) during the **before writing** phase. Overall, the participants spent, on average, 208.7 seconds (*SD*=87.5) reading the task instructions and the five source materials before they began composing their written response. The instructions received the most attention from participants, with a mean of 73.1 seconds (*SD*=29.1), accounting for over one third (35.0 percent) of the total time spent reading in the before writing phase. Three participants (Participant 1, 13 and 15) spent over half their time on the instructions. This is not surprising as it is important to develop an understanding of the task at the start of task completion, especially in an integrated writing task which involves source materials for extra reading.

Time spent on other parts of the task before writing is proportionally similar to that for the whole task completion (see Table 4.3). Source 2 is the text at which participants had the longest stay, with a mean of 52.4 seconds (*SD*=26.8), which takes up 25.1 percent of the total reading time before writing. Source 1 was less attended to (Mean=37.2; *SD*=19.1) than Source 2, while Source 3 received the least attention (Mean=22.1; *SD*=10.6) among the first three source materials. It is interesting to find that although the number of words in Source 1 and 3 together doubles that in Source 2,

Table 4.7: *Total visit duration* on each individua AOI in the reading group *before writing* by participant

Areas of interest	Participant																			
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	Mean	*SD*	Min	Max
	Sec *%*	Sec *%*	Sec *%*	Sec *%*	Sec *%*	Sec *%*	Sec *%*	Sec *%*	Sec *%*	Sec *%*	Sec *%*	Sec *%*	Sec *%*	Sec *%*	Sec *%*	Sec *%*	Sec *%*		Sec	Sec
Instructions	44.6 *51.0*	51.7 *21.9*	100.1 *38.0*	80.2 *28.1*	142.7 *31.3*	72.0 *41.7*	54.2 *31.9*	54.5 *27.1*	51.9 *27.4*	68.6 *35.6*	100.3 *38.3*	107.4 *43.1*	98.7 *54.1*	46.7 *39.0*	57.0 *55.2*	39.4 *23.2*	73.1 *35.0*	*29.1*	39.4	142.7
Source 1	16.0 *18.3*	50.1 *21.2*	52.0 *19.7*	72.1 *25.3*	79.5 *17.4*	26.9 15.6	24.0 *14.1*	17.4 *8.7*	36.2 *19.1*	37.6 *19.5*	47.7 *18.2*	19.9 *8.0*	28.5 *15.6*	18.6 *15.6*	27.4 *26.6*	40.7 *24.0*	37.2 *17.8*	*19.1*	16.0	79.5
Source 2	7.4 *8.5*	115.3 *48.7*	49.8 *18.9*	56.3 *19.7*	88.4 *19.4*	48.5 *28.1*	45.0 *26.5*	61.3 *30.5*	60.0 *31.7*	47.5 *24.6*	63.6 *24.3*	70.4 *28.3*	52.5 *28.8*	30.5 *25.5*	8.6 *8.3*	32.8 *19.4*	52.4 *25.1*	*26.8*	7.4	115.3
Source 3	2.4 *2.7*	16.4 *6.9*	27.2 *10.3*	37.0 *13.0*	72.6 *15.9*	8.9 *5.2*	15.0 *8.8*	27.7 *13.8*	17.2 *9.1*	32.2 *16.7*	28.5 *10.9*	25.9 *10.4*	2.3 *1.3*	9.2 *7.7*	6.2 *6.0*	24.4 *14.4*	22.1 *10.6*	*17.3*	2.3	72.6
Source 4 (picture)	- -	- -	1.2 *0.5*	15.8 *5.5*	30.0 *6.6*	0.4 *0.2*	11.2 *6.6*	8.1 *4.0*	6.1 *3.2*	- -	2.0 *0.8*	4.6 *1.8*	- -	0.2 *0.2*	0.3 *0.3*	14.9 *8.8*	5.9 *2.8*	*8.4*	0	30.0
Source 5	17.0 *19.5*	3.1 *1.3*	33.3 *12.6*	23.8 *8.3*	42.9 *9.4*	15.9 *9.2*	20.3 *12.0*	31.8 *15.8*	17.9 *9.5*	6.8 *3.5*	19.8 *7.6*	20.8 *8.4*	0.5 *0.3*	14.4 *12.0*	3.7 *3.6*	17.3 *10.2*	18.1 *8.7*	*11.5*	0.5	42.9
Total	87.4 *100*	236.6 *100*	263.6 *100*	285.2 *100*	456.1 *100*	172.6 *100*	169.7 *100*	200.8 *100*	189.3 *100*	192.7 *100*	261.9 *100*	249.0 *100*	182.5 *100*	119.6 *100*	103.2 *100*	169.5 *100*	208.7 *100*	*87.5*	87.4	456.1

participants spent only a slightly lower proportion of time on Source 2 (25.1 percent) alone than on Source 1 and 3 (28.4 percent) together. One possible reason for this is that Source 1 and 3 are Chinese materials, whereas Source 2 is in English. Several participants talked about this in their stimulated recalls, for example, Participant 3 said, "*The third material is a Chinese one, and listed some comments from media, so I had a quick read through it*", and Participant 4 reported that "*the first paragraph, because it is in Chinese, so I read it very fast...because the second paragraph is in English, so I read it relatively slowly*".

An average of 18.1 seconds (*SD*=11.5) were spent on Source 5 before writing, and Source 4, the picture, received the least attention during the before writing phase (Mean=5.9; *SD*=8.4). Four participants did not even look at the picture before they started to compose, and another three participants (Participant 6, 14 and 15) spent less than one second looking at it, which can be considered as merely a glance. This is noteworthy because it seems that these participants rapidly disregarded this source material after glimpsing that it was not text-based, or they made a very fast evaluation that the picture provided very limited information which would help with beginning their writing.

Table 4.8 shows participants' total visit duration on each part of the task **during writing**. The mean of duration of all visits to these AOIs rises from 208.7 seconds before writing to 357.4 seconds (*SD*=132.5). Time spent on Source 1 and 2 accounts for over half (52.2 percent) the total time for reading, with a mean of 83.0 seconds (*SD*=49.9) and 103.6 seconds (*SD*=78.0) respectively. Source 3, again, received less attention (Mean=59.9; *SD*=23.0) compared with the first two source materials, taking up 16.8 percent of participants' reading time during writing.

Source 5 was slightly more attended to (Mean=60.6; *SD*=33.1) than Source 3 in

Table 4.8: *Total visit duration* on each individual AOI in the reading group *during writing* by participant

Areas of interest	Participant 1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	Mean	*SD*	Min	Max
	Sec *%*	Sec *%*	Sec *%*	Sec *%*	Sec *%*	Sec *%*	Sec *%*	Sec *%*	Sec *%*	Sec *%*	Sec *%*	Sec *%*	Sec *%*	Sec *%*	Sec *%*	Sec *%*	Sec *%*		Sec	Sec
Instructions	86.5 *21.7*	14.4 *4.2*	42.2 *11.5*	24.5 *10.0*	10.2 *4.4*	35.5 *7.2*	18.5 *8.7*	65.7 *24.5*	21.9 *4.7*	26.1 *12.2*	28.2 *7.7*	79.8 *23.8*	49.9 *10.4*	61.6 *14.9*	36.0 *5.3*	20.4 *10.4*	38.8 *10.9*	*23.5*	10.2	86.5
Source 1	21.0 *5.3*	74.2 *21.6*	134.0 *36.6*	102.0 *41.6*	50.6 *21.9*	64.0 *13.0*	42.1 *19.8*	41.5 *15.5*	198.0 *42.1*	55.6 *26.0*	141.5 *38.7*	60.0 *17.9*	137.2 *28.6*	78.3 *18.9*	108.5 *15.8*	19.3 *9.8*	83.0 *23.2*	*49.9*	19.3	198.0
Source 2	114.6 *28.7*	177.5 *51.7*	69.1 *18.9*	31.2 *12.7*	63.2 *27.4*	243.9 *49.4*	96.6 *45.4*	51.4 *19.2*	117.8 *25.1*	48.3 *22.6*	75.1 *20.5*	14.6 *4.4*	149.2 *31.1*	98.4 *23.8*	287.6 *42.0*	18.5 *9.4*	103.6 *29.0*	*78.0*	14.6	287.6
Source 3	78.1 *19.5*	66.1 *19.2*	38.3 *10.5*	47.4 *19.3*	61.7 *26.8*	99.1 *20.1*	38.4 *18.1*	45.7 *17.1*	67.0 *14.2*	20.4 *9.5*	72.6 *19.8*	48.7 *14.5*	63.8 *13.3*	63.6 *15.4*	108.5 *15.8*	38.3 *19.5*	59.9 *16.8*	*23.0*	20.4	108.5
Source 4 (picture)	5.8 *1.5*	8.0 *2.3*	0.2 *0.1*	18.7 *7.6*	5.2 *2.3*	3.3 *0.7*	3.3 *1.6*	3.2 *1.2*	0.3 *0.1*	1.3 *0.6*	11.5 *3.1*	76.1 *22.7*	5.2 *1.1*	8.6 *2.1*	20.3 *3.0*	13.1 *6.7*	11.5 *3.2*	*18.3*	0.2	76.1
Source 5	93.5 *23.4*	3.3 *1.0*	81.9 *22.4*	21.6 *8.8*	39.7 *17.2*	47.5 *9.6*	13.7 *6.4*	60.5 *22.6*	65.2 *13.9*	62.3 *29.1*	36.9 *10.1*	56.0 *16.7*	73.7 *15.4*	103.6 *25.0*	124.0 *18.1*	86.4 *44.1*	60.6 *17.0*	*33.1*	3.3	124.0
Total	399.5 *100*	343.5 *100*	365.7 *100*	245.4 *100*	230.6 *100*	493.3 *100*	212.6 *100*	268.0 *100*	470.2 *100*	214.0 *100*	365.8 *100*	335.2 *100*	479.0 *100*	414.1 *100*	684.9 *100*	196.0 *100*	357.4 *100*	*132.5*	196.0	684.9

terms of the mean, but as discussed earlier, considering the average time spent on each word, this amount of attention still outnumbered that on other source materials. It should also be noted that participants did spend some time (Mean=38.8; *SD*=23.5) on Instructions at this phase of writing, although about half less than they did before writing. Participants' reading strategies when looking back to the Instructions during writing will be examined with more evidence from stimulated recalls data in the next chapter. Still, little attention was given to Source 4 during writing. Participant 12 had the longest duration (76.1 seconds) at the picture, but she recalled in her protocols that this seemed to be unconscious and she did not get any useful formation out of it.

Data for participants' total visit duration on each AOI **after finishing** the first draft is presented in Table 4.9. Overall, most participants gave very limited attention to these AOIs once they had finished composing their response. In fact, there were only three participants (Participant 4, 13 and 14) who spent more than 30 seconds looking at the instructions and source materials. For Participant 4, 67.3 percent of the total time (55.9 seconds) was given to Source 2 and 3, this may be because he decided to integrate some information from these materials into the essay when revising the draft, as he recalled, "*...I found that I should add the concluding sentence in the third source material. It serves a transition purpose, so I added this sentence*". Participant 13 spent 78.2 percent of the total time on Instructions. She explained this in her protocols, "*Here I went to reread the instructions, because I had written some critical comments in the last paragraph, I wanted to check if they were what the task requires to write*". For Participant 14, 70.1 seconds were given to Source 5. She reported that "*I was reading through the text I had written...and I was looking at the 'key concepts and expressions' to see if there were any other words I could use in the essay*".

Table 4.10 presents participants' total visit duration on the answer sheet at different

Table 4.9: *Total visit duration on* each individual AOI in the reading group *after writing* by participant

Areas of interest	Participant																			
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	Mean	*SD*	Min	Max
	Sec *%*	Sec *%*	Sec *%*	Sec *%*	Sec *%*	Sec *%*	Sec *%*	Sec *%*	Sec *%*	Sec *%*	Sec *%*	Sec *%*	Sec *%*	Sec *%*	Sec *%*	Sec *%*	Sec *%*		Sec	Sec
Instructions	0.5 *14.7*	- -	0.4 *22.2*	8.7 *15.6*	0.7 *35.0*	0.2 *20.0*	0.8 *4.7*	0.9 *8.3*	- -	0.2 *1.1*	0.3 *25.0*	0.9 *40.9*	26.2 *78.2*	0.1 *0.1*	- -	- -	2.5 *17.0*	*6.7*	0	26.2
Source 1	- -	- -	0.1 *5.6*	0.8 *1.4*	- -	0.3 *30.0*	3.0 *17.4*	0.5 *4.6*	- -	0.3 *1.7*	0.5 *41.7*	- -	1.1 *3.3*	3.7 *4.6*	- -	0.3 *6.4*	0.7 *4.8*	*1.1*	0	3.7
Source 2	- -	- -	0.2 *11.1*	22.0 *39.4*	- -	0.1 *10.0*	1.5 *8.7*	0.2 *1.9*	0.3 *75.0*	0.3 *1.7*	0.3 *25.0*	0.2 *9.1*	1.7 *5.1*	1.2 *1.5*	- -	- -	1.8 *12.2*	*5.4*	0	22.0
Source 3	2.5 *73.5*	- -	0.4 *22.2*	15.6 *27.9*	0.7 *35.0*	- -	- -	0.1 *0.9*	0.1 *25.0*	8.0 *45.5*	- -	0.5 *22.7*	- -	5.6 *6.9*	- -	- -	2.1 *14.3*	*4.3*	0	15.6
Source 4 (picture)	0.4 *11.8*	- -	0.4 *22.2*	8.2 *14.7*	0.1 *5.0*	- -	3.9 *22.7*	0.1 *0.9*	- -	8.7 *49.4*	- -	0.4 *18.2*	- -	0.6 *0.7*	- -	- -	1.4 *9.5*	*2.9*	0	8.7
Source 5	- -	0.2 *100*	0.3 *16.7*	0.6 *1.1*	0.5 *25.0*	0.4 *40.0*	8.0 *46.5*	9.0 *83.3*	- -	0.1 *0.6*	0.1 *8.3*	0.2 *9.1*	4.5 *13.4*	70.1 *86.2*	- -	4.4 *93.6*	6.2 *42.2*	*17.3*	0	70.1
Total	3.4 *100*	0.2 *100*	1.8 *100*	55.9 *100*	2.0 *100*	1.0 *100*	17.2 *100*	10.8 *100*	0.4 *100*	17.6 *100*	1.2 *100*	2.2 *100*	33.5 *100*	81.3 *100*	- -	4.7 *100*	14.6 *100*	*23.4*	0	81.3

Table 4.10: *Total visit duration* on the answer sheet at different writing phase by participant

Writing phase	Participant 1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	Mean	*SD*	Min	Max
	Sec %	Sec %	Sec %	Sec %	Sec %	Sec %	Sec %	Sec %	Sec %	Sec %	Sec %	Sec %	Sec %	Sec %	Sec %	Sec %	Sec %		Sec	Sec
Before writing	2.2 *0.1*	2.8 *0.2*	4.3 *0.3*	7.6 *0.5*	110.1 *6.1*	6.9 *0.5*	46.5 *2.5*	23.5 *1.3*	16.8 *1.1*	16.3 *0.8*	21.5 *2.4*	10.4 *0.6*	1.0 *0.1*	1.2 *0.1*	14.2 *0.7*	40.1 *2.3*	20.3 *1.2*	*27.4*	1.0	110.1
During writing	1695.3 *99.3*	1354.3 *87.9*	1301.1 99.0.	1052.7 *67.3*	1567.4 *87.5*	1283.7 *88.0*	1500.2 *82.1*	1643.2 *89.2*	1400.1 *95.6*	1865.2 *96.3*	867.2 *97.0*	1746.6 *99.3*	1576.7 *91.5*	831.2 *62.7*	2032.5 *98.9*	1652.5 *95.2*	1460.6 *90.1*	*337.7*	831.2	2032.5
After writing	9.4 *0.6*	184.1 *11.9*	9.4 *0.7*	505.0 *32.3*	112.9 *6.3*	168.1 *11.5*	280.2 *15.3*	175.4 *9.5*	48.1 *3.3*	54.9 *2.8*	4.9 *0.5*	1.3 *0.1*	146.0 *8.5*	493.4 *37.2*	7.6 *0.4*	42.4 *2.4*	140.2 *8.6*	*162.7*	1.3	505.0
Total	1706.9 *100*	1541.2 *100*	1314.8 *100*	1565.3 *100*	1790.4 *100*	1458.7 *100*	1826.9 *100*	1842.1 *100*	1465.0 *100*	1936.4 *100*	893.6 *100*	1758.3 *100*	1723.7 *100*	1325.8 *100*	2054.3 *100*	1735.0 *100*	1621.2 *100*	*286.0*	893.6	2054.3

phases of writing. It is not surprising to find that participants looked at the answer sheet the longest in the **during writing** phase, because this task ultimately requires test-takers to produce a written product. Also, it can be seen that most participants spent some amount of time on this AOI after writing. This is likely due to the monitoring and revising processes that participants may employ after they completed the first draft. More evidence on this will be presented in the next chapter.

4.4.3 Visit count

In order to answer the second research question (RQ2), participants' eye movements have been investigated in terms of the time to first fixation and total time spent on each part of the task. This section seeks to examine participants' looking behaviour from the perspective of *visit count*, i.e., how many times they visited each AOI throughout task completion and at different phases of writing.

Table 4.11 presents the number of visits in each part of the task (not including the answer sheet). As shown in the table, Source 2 and Source 1 were looked at most frequently, with a mean of 83.3 (*SD*=61.6) and 73.3 (*SD*=36.1) respectively, followed by Source 5 (Mean=67.1; *SD*=41.8), Source 3 (Mean=43.8; *SD*=18.3), Instructions (Mean=41.4; *SD*=13.8), and finally Source 4 (Mean=22.3; *SD*=23.9). It should be noted that, according to the total visit duration data (see Table 4.3), although participants spent less or roughly the same amount of time on Source 5 than on Instructions and Source 3, Source 5 was visited more often than those two AOIs. This may be because of the nature of Source 5 that it contains separate words and phrases with independent meanings, so that participants may spend less time on it per visit than they did when reading through other materials. Participant 14 had significantly more visits (197) in Source 5 than other participants. Participants visited Source 4 (picture) the least often,

Table 4.11: *Visit count* on each individual AOI in the reading group by participant

Areas of interest	Participant																			
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	Mean	*SD*	Min	Max
	Ct *%*	Ct *%*	Ct *%*	Ct *%*	Ct *%*	Ct *%*	Ct *%*	Ct *%*	Ct *%*	Ct *%*	Ct *%*	Ct *%*	Ct *%*	Ct *%*	Ct *%*	Ct *%*	Ct *%*		Ct	Ct
Instructions	39 *14.8*	64 *10.5*	54 *14.7*	42 *11.7*	46 *16.5*	33 *9.6*	37 *18.3*	58 *17.6*	25 *8.7*	28 *16.9*	30 *10.2*	36 *11.3*	27 *10.4*	40 *8.4*	71 *13.6*	32 *14.5*	41.4 *12.5*	*13.8*	25	71
Source 1	37 *14.1*	164 *26.9*	112 *30.5*	99 *27.5*	55 *19.8*	45 *13.2*	38 *18.8*	51 *15.5*	99 *34.4*	37 *22.3*	102 *34.6*	38 *11.9*	79 *30.5*	87 *18.3*	81 *15.5*	48 *21.7*	73.3 *22.1*	*36.1*	37	164
Source 2	52 *19.8*	255 *41.9*	72 *19.6*	74 *20.6*	58 *20.9*	151 *44.2*	56 *27.7*	58 *17.6*	85 *29.5*	24 *14.5*	57 *19.3*	28 *8.8*	73 *28.2*	85 *17.9*	178 *34.1*	26 *11.8*	83.3 *25.2*	*61.6*	24	255
Source 3	48 *18.3*	70 *11.5*	38 *10.4*	44 *12.2*	49 *17.6*	77 *22.5*	29 *14.4*	57 *17.3*	37 *12.8*	14 *8.4*	35 *11.9*	26 *8.1*	29 *11.2*	54 *11.3*	70 *13.4*	23 *10.4*	43.8 *13.2*	*18.3*	14	77
Source 4 (picture)	7 *2.7*	27 *4.4*	8 *2.2*	53 *14.7*	18 *6.5*	3 *0.9*	14 *6.9*	19 *5.8*	2 *0.7*	9 *5.4*	32 *10.8*	98 *30.6*	8 *3.1*	13 *2.7*	26 *5.0*	20 *9.0*	22.3 *6.7*	*23.9*	2	98
Source 5	80 *30.4*	29 *4.8*	83 *22.6*	48 *13.3*	52 *18.7*	33 *9.6*	28 *13.9*	86 *26.1*	40 *13.9*	54 *32.5*	39 *13.2*	94 *29.4*	43 *16.6*	197 *41.4*	96 *18.4*	72 *32.6*	67.1 *20.3*	*41.8*	28	197
Total	263 *100*	609 *100*	367 *100*	360 *100*	278 *100*	342 *100*	202 *100*	329 *100*	288 *100*	166 *100*	295 *100*	320 *100*	259 *100*	476 *100*	522 *100*	221 *100*	331.1 *100*	*118.0*	166	609

which concurs with the findings from the total visit duration data. In addition, it is the only material on which the number of visits (Mean=22.3) outnumbers the duration of all visits (Mean=18.8), which means that these participants, on average, spent less than a second per visit on the picture.

The visit count data were submitted to the Kruskal-Wallis test (RQ2a), and results are shown in Table 4.12. According to the visit count mean, Source 2 was visited most frequently, followed by Source 1, Source 5, Source 3, and finally Instructions (χ^2=15.671, df=4, p=0.003). The Mann-Whitney tests (see Table 4.13) then confirmed that there were no significant differences in number of visits between Source 1, Source 2 and Source 5, nor between Source 3 and Instructions, but the latter two materials had significantly fewer visits. Participants read and reread the instructions the least often, which may because that the participants had a longer stay at the AOI Instructions per visit.

Table 4.12: Results of Kruskal-Wallis test of *visit count* on different AOIs

AOIs	Visit count (mean)	Std. deviation	Mean rank
Instructions	41.38	13.77	27.06
Source 1	73.25	36.10	51.00
Source 2	83.25	61.60	50.00
Source 3	43.75	18.31	29.28
Source 5	67.13	41.83	45.16

Table 4.13: Significant differences in *visit count* between different AOIs

Comparisons	Mann-Whitney U	Z	*p*	Effect size
Instructions – Source 1	49.000	-2.979	.003	.28
Instructions – Source 2	59.500	-2.583	.010	.21
Instructions – Source 3	121.000	-.264	.792	.00
Instructions – Source 5	67.500	-2.281	.023	.16
Source 1 – Source 2	126.000	-.075	.940	.00
Source 1 – Source 3	58.500	-2.622	.009	.21
Source 1 – Source 5	106.500	-.811	.418	.02
Source 2 – Source 3	62.000	-2.489	.013	.19
Source 2 – Source 5	112.500	-.584	.559	.01
Source 3 – Source 5	77.000	-1.923	.054	.12

Table 4.14 shows the visit count in the AOI reading group and writing group. Unlike the findings from the total visit duration data (see Table 7.4), which show that participants spent a lot more time writing (Mean=1623.1 seconds) than reading (Mean=580.8 seconds), the five source materials, together with the instructions, were looked at more frequently (Mean=331.1; *SD*=118.0) than the answer sheet (Mean=229.6; *SD*=50.4). The higher number of visit counts for the reading group (versus the answer sheet) may indicate that the participants were not moving from the writing to one text and back to the writing. Instead, they may be moving between the texts, and then going back to the writing, which suggests that the participants were synthesising the information they were selecting from the texts, and not just filling in the writing with discrete pieces of information.

Table 4.14: *Visit count* on AoI reading group and writing group by participant

AoI group	Participant 1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	Mean	*SD*	Min	Max
	Ct *%*	Ct *%*	Ct *%*	Ct *%*	Ct *%*	Ct *%*	Ct *%*	Ct *%*	Ct *%*	Ct *%*	Ct *%*	Ct *%*	Ct *%*	Ct *%*	Ct *%*	Ct *%*	Ct *%*		Ct	Ct
Reading	263 *49.7*	609 *67.4*	367 *62.6*	360 *61.0*	278 *56.0*	342 *63.0*	202 *53.0*	329 *53.0*	288 *60.3*	166 *42.1*	295 *57.5*	320 *55.9*	259 *53.4*	476 *58.6*	522 *72.7*	221 *63.3*	331.1 *59.1*	*118.0*	166	609
Writing	266 *50.3*	295 *32.6*	219 *37.4*	230 *39.0*	218 *44.0*	201 *37.0*	179 *47.0*	292 *47.0*	190 *39.7*	228 *57.9*	218 *42.5*	252 *44.1*	226 *46.6*	336 *41.4*	196 *27.3*	128 *36.7*	229.6 *40.9*	*50.4*	128	336
Total	529 *100*	904 *100*	586 *100*	590 *100*	496 *100*	543 *100*	381 *100*	621 *100*	478 *100*	394 *100*	513 *100*	572 *100*	485 *100*	812 *100*	718 *100*	349 *100*	560.7 *100*	*149.7*	349	904

Again, the whole task completion was divided into three phases: before writing, during writing and after writing. Participants' visit count in each part of the task (not including the answer sheet) at each of the three phases are displayed in Tables 4.15, 4.16 and 4.17 respectively. First, before participants started to write (see Table 4.15), they had, on average, 70.6 visits (*SD*=48.7) in these AOIs. The Instruction AOI was looked at most frequently (Mean=20.1; *SD*=10.4), accounting for nearly 30 percent of the total number of visits during this phase, which agrees with the results from total visit duration data (see Table 4.7) that participants attended most to the instructions before writing. Source 1 and 2 were the two mostly visited materials before writing, with a mean of 19.1 (*SD*=20.3) and 13.6 (*SD*=17.8) times respectively. It should be noted that although Source 2 contains fewer words and was less frequently visited compared with Source 1, according to the total visit duration data, participants spent more time reading it than Source 1. This indicates that participants tended to stay in Source 2 longer per visit, which may be due to the nature of this material: that it is an English text which takes more time for participants to process compared with a Chinese text. Source 5 was also looked at relatively often (Mean=9.3; *SD*=5.7) during this phase. Participant 3 visited this material 17 times and he recalled in his protocols when he found that his eye fixations jumped constantly between Source 5 and other source materials at some interval, "*I looked back at the source materials when I finished reading the key words, I was thinking about in which part of the essay these words might be used, so I looked back at the source materials again, that's why there was a constant jump of fixations between these parts at this point*".

Table 4.15: *Visit count* on each individual AOI in the reading group *before writing* by participant

Areas of interest	Participant 1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	Mean	*SD*	Min	Max
	Ct %	Ct %	Ct %	Ct %	Ct %	Ct %	Ct %	Ct %	Ct %	Ct %	Ct %	Ct %	Ct %	Ct %	Ct %	Ct %	Ct %		Sec	Sec
Instructions	9 *33.3*	47 *21.1*	20 *27.8*	21 *22.8*	37 *28.2*	17 29.3	25 *43.1*	24 *34.8*	13 *21.7*	15 *32.6*	20 *40.0*	11 *35.5*	8 *27.6*	9 *22.5*	22 *41.5*	23 *25.6*	20.1 *28.5*	*10.4*	8	47
Source 1	11 *40.7*	89 *39.9*	19 *26.4*	23 *25.0*	31 *23.7*	13 *22.4*	12 *20.7*	10 *14.5*	13 *21.7*	11 *23.9*	11 *22.0*	2 *6.5*	12 *41.4*	10 *25.0*	7 *13.2*	32 *35.6*	19.1 *27.1*	*20.3*	2	89
Source 2	3 *11.1*	77 *34.5*	6 *8.3*	12 *13.0*	25 *19.1*	15 25.9	5 *8.6*	8 *11.6*	17 *28.3*	4 *8.7*	6 *12.0*	6 *19.4*	6 *20.7*	10 *25.0*	8 *15.1*	10 *11.1*	13.6 *19.3*	*17.8*	3	77
Source 3	1 *3.7*	6 *2.7*	5 *6.9*	4 *4.3*	14 *10.7*	4 *6.9*	2 *3.4*	4 *5.8*	8 *13.3*	7 *15.2*	3 *6.0*	5 *16.1*	1 *3.4*	5 *12.5*	5 *9.4*	8 *8.9*	5.1 *7.2*	*3.2*	1	14
Source 4 (picture)	- -	- -	5 *6.9*	11 *12.0*	8 *6.1*	2 *3.4*	6 *10.3*	5 *7.2*	1 *1.7*	- -	3 *6.0*	2 *6.5*	- -	1 *2.5*	3 *5.7*	6 *6.7*	3.3 *4.7*	*3.3*	0	11
Source 5	3 *11.1*	4 *1.8*	17 *23.6*	21 *22.8*	16 *12.2*	7 *12.1*	8 *13.8*	18 *26.1*	8 *13.3*	9 *19.6*	7 *14.0*	5 *16.1*	2 *6.9*	5 *12.5*	8 *15.1*	11 *12.2*	9.3 *13.2*	*5.7*	2	21
Total	27 *100*	223 *100*	72 *100*	92 *100*	131 *100*	58 *100*	58 *100*	69 *100*	60 *100*	46 *100*	50 *100*	31 *100*	29 *100*	40 *100*	53 *100*	90 *100*	70.6 *100*	*48.7*	27	223

Table 4.16 shows the number of visits in each AOI **during writing**. Overall, participants made an average of 245.4 visits (*SD*=97.6) to the instructions and source materials. Source 2 was looked at the most often, with a mean of 67.9 (*SD*=49.9). Participant 2 visited this material 178 times during writing, which accounts for nearly half of her total number of visits at this phase. Source 1 was the second most visited (Mean=52.9; *SD*=27.4) source material. Source 5 was looked at as often as Source 1, with a mean of 50.5 (*SD*=29.3). Surprisingly, Participant 14 made 113 visits in Source 5. Also, it is worth noting that participants did look back at Instructions (Mean=19.5; *SD*=12.0) during writing, which may indicate that task representation is an ongoing process that participants engaged in throughout the whole task completion, rather than a process that only occurs at a particular phase of writing. Although participants visited Source 4 (picture) 17.4 times, according to the total visit duration data, they spent only an average of 5.9 seconds on the picture, which indicates again that they paid very little attention to this source material.

Data for the number of visits in each AOI after finishing the first draft is presented in Table 4.17. Overall, most participants made very limited visits to the task instructions and source materials during this phase. There are two participants (Participant 4 and 14) who looked markedly more often at the reading AOIs than other participants. Participant 4 had a total of 52 visits; Source 2 was the most visited (15 visits). As discussed in total visit duration data, this is likely because that participant wanted to integrate some information from that material to the draft that had been written; she said in the recalls that "*I was wondering if this paragraph can be added into the essay, but then I found it was not necessary to do that*". Participant 14 had 95 visits to reading texts at this phase, and the majority (79 visits) were to Source 5. To illustrate, Figure 4.4 shows the eye movements of Participant 14 in a two-minute interval starting from

Table 4.16: *Visit count* on each individual AOI in the reading group *during writing* by participant

Areas of interest	Participant 1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	Mean	*SD*	Min	Max
	Ct *%*	Ct *%*	Ct *%*	Ct *%*	Ct *%*	Ct *%*	Ct *%*	Ct *%*	Ct *%*	Ct *%*	Ct *%*	Ct *%*	Ct *%*	Ct *%*	Ct *%*	Ct *%*	Ct *%*		Sec	Sec
Instructions	28 *12.1*	17 *4.4*	33 *11.5*	14 *6.7*	6 *4.4*	15 *5.4*	8 *6.2*	32 *13.0*	12 *5.3*	12 *10.8*	9 *3.7*	24 *8.5*	13 *6.0*	29 *8.5*	50 *10.6*	10 *7.8*	19.5 *7.9*	*12.0*	6	50
Source 1	26 *11.2*	75 *19.5*	92 *32.1*	73 *34.8*	24 *17.5*	31 *11.1*	24 *18.6*	39 *15.9*	86 *37.9*	25 *22.5*	90 *37.3*	37 *13.0*	64 *29.4*	71 *20.8*	74 *15.7*	16 *12.4*	52.9 *21.6*	*27.4*	16	92
Source 2	49 *21.1*	178 *46.2*	65 *22.6*	47 *22.4*	33 *24.1*	135 *48.2*	49 *38.0*	49 *19.9*	68 *30.0*	19 *17.1*	50 *20.7*	21 *7.4*	66 *30.3*	71 *20.8*	170 *36.2*	16 *12.4*	67.9 *27.7*	*49.9*	16	178
Source 3	46 *19.8*	64 *16.6*	31 *10.8*	31 *14.8*	32 *23.4*	73 *26.1*	27 *20.9*	52 *21.1*	28 *12.3*	6 *5.4*	32 *13.3*	19 *6.7*	28 *12.8*	47 *13.8*	65 *13.8*	15 *11.6*	37.2 *15.2*	*18.9*	6	73
Source 4 (picture)	6 *2.6*	27 *7.0*	1 *0.3*	31 *14.8*	9 *6.6*	1 *0.4*	5 *3.9*	13 *5.3*	1 *0.4*	5 *4.5*	29 *12.0*	95 *33.5*	8 *3.7*	10 *2.9*	23 *4.9*	14 *10.9*	17.4 *7.1*	*23.0*	1	95
Source 5	77 *33.2*	24 *6.2*	65 *22.6*	14 *6.7*	33 *24.1*	25 *8.9*	16 *12.4*	61 *24.8*	32 *14.1*	44 *39.6*	31 *12.9*	88 *31.0*	39 *17.9*	113 *33.1*	88 *18.7*	58 *45.0*	50.5 *20.6*	*29.3*	14	113
Total	232 *100*	385 *100*	287 *100*	210 *100*	137 *100*	280 *100*	129 *100*	246 *100*	227 *100*	111 *100*	241 *100*	284 *100*	218 *100*	341 *100*	470 *100*	129 *100*	245.4 100	*97.6*	111	470

Table 4.17: *Visit count* on each individual AOI in the reading group *after writing* by participant

Areas of interest	Participant 1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	Mean	*SD*	Min	Max
	Ct *%*	Ct *%*	Ct *%*	Ct *%*	Ct *%*	Ct *%*	Ct *%*	Ct *%*	Ct *%*	Ct *%*	Ct *%*	Ct *%*	Ct *%*	Ct *%*	Ct *%*	Ct *%*	Ct *%*		Sec	Sec
Instructions	2 *50.0*	- -	2 *22.2*	7 *13.5*	2 *22.2*	1 *20.0*	3 *21.4*	2 *14.3*	- -	1 *11.1*	1 *25.0*	1 *16.7*	6 *50.0*	1 *1.1*	- -	- -	1.8 *12.0*	*2.0*	0	7
Source 1	- -	- -	1 *11.1*	3 *5.8*	- -	2 *40.0*	2 *14.3*	2 *14.3*	- -	1 *11.1*	1 *25.0*	- -	3 *25.0*	7 *7.4*	- -	1 *25.0*	1.4 *9.3*	*1.8*	0	7
Source 2	- -	- -	1 *11.1*	15 *28.8*	- -	1 *20.0*	2 *14.3*	1 *7.1*	1 *50.0*	1 *11.1*	1 *25.0*	1 *16.7*	1 *8.3*	4 *4.2*	- -	- -	1.8 *12.0*	*3.7*	0	15
Source 3	1 *25.0*	- -	2 *22.2*	9 *17.3*	3 *33.3*	- -	- -	1 *7.1*	1 *50.0*	1 *11.1*	- -	2 *33.3*	- -	2 *2.1*	- -	- -	1.4 *9.3*	*2.2*	0	9
Source 4 (picture)	1 *25.0*	- -	2 *22.2*	11 *21.2*	1 *11.1*	- -	3 *21.4*	1 *7.1*	- -	4 *44.4*	- -	1 *16.7*	- -	2 *2.1*	- -	- -	1.6 *10.7*	*2.8*	0	11
Source 5	- -	1 *100*	1 *11.1*	7 *13.5*	3 *33.3*	1 *20.0*	4 *28.6*	7 *50.0*	- -	1 *11.1*	1 *25.0*	1 *16.7*	2 *16.7*	79 *83.2*	- -	3 *75.0*	6.9 *46.0*	*19.3*	0	79
Total	4 *100*	1 *100*	9 *100*	52 *100*	9 *100*	5 *100*	14 *100*	14 *100*	2 *100*	9 *100*	4 *100*	6 *100*	12 *100*	95 *100*	- -	4 *100*	15.0 *100*	*24.5*	0	95

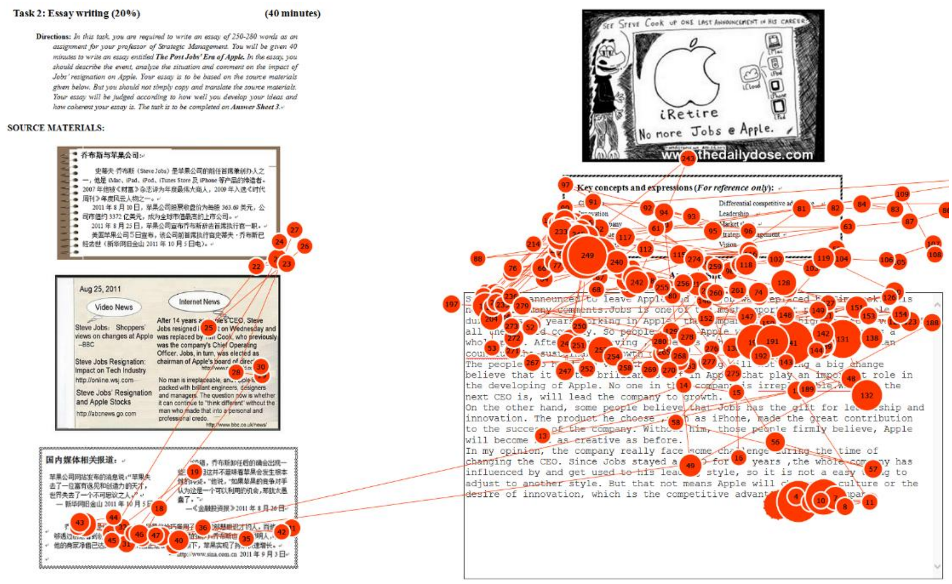

Figure 4.4: Gaze plot for the interval from 31:00 to 33:00 of recording of Participant 14

31:00 to 33:00 of the recording, during which she was reading through the text that had been written and making revisions. It can be seen in this figure that most of the fixations were on Source 5 and the first paragraph of the essay, as Participant 14 recalled in the protocols that she "*was reading through the essay from the start, and went back to look at the key words to decide if there were any more words that can be used in the essay*". This may provide an explanation as to why this participant visited Source 5 so frequently while she was revising the draft.

Table 4.18 presents the number of visits within the answer sheet by participants at different phases of writing. It can be seen that the majority of visits were done during writing, which is not surprising because this is ultimately a task that requires participants to produce a written product for assessment. Participants were also found looking at the answer sheet before starting to compose. Participant 5 had significantly more visits to the answer sheet at this phase than other participants, and by looking into her eye-movement recordings and stimulated recalls, she spent much time on macro-

Table 4.18: *Visit count* on the answer sheet at different writing phase by participant

Writing phase	Participant																			
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	Mean	*SD*	Min	Max
	Ct *%*	Ct *%*	Ct *%*	Ct *%*	Ct *%*	Ct *%*	Ct *%*	Ct *%*	Ct *%*	Ct *%*	Ct *%*	Ct *%*	Ct *%*	Ct *%*	Ct *%*	Ct *%*	Ct *%*		Ct	Ct
Before writing	3 *1.1*	9 *3.0*	8 *3.6*	20 *8.6*	72 *32.9*	14 *6.9*	25 *13.9*	24 *8.2*	11 *5.8*	22 *9.6*	13 *5.9*	9 *3.6*	3 *1.3*	7 *2.1*	18 *9.1*	21 *16.3*	17.4 *7.5*	*16.2*	3	72
During writing	262 *98.1*	257 *86.8*	207 *94.1*	170 *73.3*	144 *65.8*	173 *85.6*	142 *78.9*	244 *83.0*	178 *93.2*	199 *86.5*	204 *92.7*	242 *95.7*	217 *95.6*	226 *67.1*	178 *90.4*	104 *80.6*	196.7 *85.2*	*44.7*	104	262
After Writing	2 *0.7*	30 *10.1*	5 *2.3*	42 *18.1*	3 *1.4*	15 *7.4*	13 *7.2*	26 *8.8*	2 *1.0*	9 *3.9*	3 *1.4*	2 *0.8*	7 *3.1*	104 *30.9*	1 *0.5*	4 *3.1*	16.8 *7.2*	*26.2*	1	104
Total	267 *100*	296 *100*	220 *100*	232 *100*	219 *100*	202 *100*	180 *100*	294 *100*	191 *100*	230 *100*	220 *100*	253 *100*	227 *100*	337 *100*	197 *100*	129 *1000*	230.9 *100*	*50.5*	129	337

planning, i.e., making a general plan for writing goals and content. For example, she reported in her protocols, "*I divided the requirements in the instructions into several parts, and decided the content to be written in each paragraph corresponding to each part of the requirements*", and on micro-planning before she started to compose the first paragraph, "*I was thinking, because the first paragraph was going to briefly describe this incident, so I needed to find some information about the facts of Apple Company, for example, what kind of company it is, and what is the relation between Steve Jobs and this company*". Participants also visited the answer sheet after they finished writing the first draft, as discussed when looking at the total visit duration data (see Table 4.10), this may because participants were doing some revision work on the draft.

4.4.4 Visit duration

Unlike *total visit duration*, which is calculated by adding up the duration of all visits within an AOI or AOI group, *visit duration* measures the duration of each individual visit in an AOI, and it can provide some descriptive statistics such as mean visit duration, i.e., how long on average each visit lasts, and the maximum visit duration, i.e., how long the longest visit was.

Table 4.19 displays the participants' visit duration data. As shown in the table, the minimum visit duration within each AOI by these participants was rather short, most of which were around 0.10 seconds. These short visits were likely to be participants' unconscious eye visits within an AOI, which might be composed of a single short fixation along the path of a long eye movement and did not hold any meaningful looking behaviour. The mean visit duration within each individual AOI was less than three seconds, with the exception of that on the answer sheet, which was 7.5 seconds. This indicates that participants constantly switched between these AOIs, which may happen

Table 4.19: Descriptive statistics for *visit duration* on each individual AoI by participant

Areas of interest	Participant	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	Mean
		Sec	Sec	Sec	Sec	Sec	Sec	Sec	Sec	Sec	Sec	Sec	Sec	Sec	Sec	Sec	Sec	
Instructions	Mean	3.4	1.1	2.6	2.7	3.4	3.3	2.0	2.1	3.0	3.4	4.3	5.2	6.5	2.7	1.3	1.9	3.0
	Median	1.6	0.7	0.8	1.0	1.1	1.3	0.7	0.7	1.4	1.2	1.1	1.4	1.8	0.6	0.5	1.1	
	SD	*4.3*	*1.1*	*8.1*	*6.8*	*9.1*	*5.9*	*3.6*	*3.3*	*5.0*	*7.3*	*6.4*	*11.8*	*11.0*	*4.7*	*2.9*	*2.7*	
	Min	0.11	0.06	0.15	0.09	0.08	0.14	0.19	0.09	0.21	0.18	0.15	0.06	0.34	0.07	0.07	0.09	
	Max	22.6	4.8	59.4	44.3	61.1	29.8	19.2	16.2	24.2	37.2	27.9	60.1	52.7	25.1	20.9	12.9	
Source text_1	Mean	1.0	0.8	1.7	1.8	2.6	2.1	2.6	2.0	2.1	4.0	2.4	3.0	2.8	1.5	1.7	2.0	2.1
	Median	0.5	0.3	0.9	1.2	0.8	0.9	1.2	0.4	1.1	0.7	0.6	0.5	1.6	0.5	0.9	0.5	
	SD	*1.0*	*1.0*	*4.0*	*2.4*	*5.3*	*3.7*	*4.2*	*7.2*	*5.0*	*7.5*	*7.3*	*9.8*	*4.5*	*3.1*	*2.6*	*4.2*	
	Min	0.08	0.06	0.06	0.09	0.09	0.10	0.08	0.09	0.11	0.11	0.10	0.10	0.07	0.02	0.06	0.09	
	Max	3.2	5.9	40.9	13.3	33.0	17.6	23.8	55.2	44.2	23.9	54.0	51.7	26.4	17.1	19.9	19.1	
Source text_2	Mean	2.4	1.2	1.7	1.5	2.8	1.9	1.8	1.3	2.3	4.3	2.9	2.9	2.3	1.5	1.6	2.7	2.2
	Median	1.3	0.4	0.4	0.3	1.7	0.7	1.2	0.5	1.9	2.6	1.4	1.2	1.6	0.7	0.9	0.7	
	SD	*3.0*	*3.1*	*4.9*	*5.1*	*4.1*	*3.3*	*2.6*	*2.6*	*1.8*	*4.4*	*4.0*	*4.0*	*2.5*	*1.9*	*1.9*	*4.4*	
	Min	0.11	0.06	0.08	0.07	0.16	0.10	0.09	0.07	0.12	0.23	0.16	0.13	0.13	0.08	0.08	0.16	
	Max	15.0	36.2	39.7	43.7	24.2	22.9	14.1	15.4	7.6	14.0	18.0	18.9	10.3	8.7	11.3	16.4	
Source text_3	Mean	1.7	1.2	1.7	2.3	2.0	1.4	1.3	0.6	3.2	1.1	0.4	0.8	0.7	0.7	0.8	1.4	1.3
	Median	0.8	0.5	0.5	1.2	0.4	1.0	0.7	0.2	3.2	0.3	0.2	0.5	0.5	0.3	0.2	1.0	
	SD	*2.7*	*1.7*	*3.3*	*4.8*	*5.3*	*1.5*	*1.8*	*1.2*	*4.1*	*2.1*	*0.4*	*0.9*	*0.8*	*1.1*	*1.1*	*1.6*	
	Min	0.17	0.06	0.10	0.12	0.09	0.07	0.10	0.07	0.32	0.14	0.06	0.07	0.10	0.10	0.06	0.10	
	Max	17.6	9.4	17.7	31.6	22.9	7.9	6.6	4.9	6.1	6.7	1.5	5.9	2.4	4.1	4.0	5.9	

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	Mean
Source text_4 (picture)	Mean	0.9	0.3	0.2	0.8	2.4	1.2	1.8	1.2	2.4	2.5	1.9	2.1	2.1	1.2	1.7	1.3	1.5
	Median	0.5	0.2	0.2	0.6	0.7	0.3	1.2	0.5	1.4	1.2	1.0	1.1	1.0	0.7	0.8	0.6	
	SD	*1.0*	*0.3*	*0.1*	*1.1*	*5.0*	*1.8*	*2.6*	*1.7*	*2.7*	*3.4*	*2.5*	*3.3*	*3.4*	*1.4*	*2.3*	*1.3*	
	Min	0.27	0.06	0.12	0.09	0.10	0.06	0.06	0.10	0.08	0.14	0.08	0.08	0.06	0.06	0.07	0.13	
	Max	3.2	1.7	0.4	7.4	29.7	3.3	14.5	9.5	14.9	16.8	17.1	18.5	21.4	8.2	11.8	5.1	
Source text_5 (key concepts and expressions)	Mean	1.4	0.2	1.4	1.0	1.6	1.9	1.5	1.2	2.1	1.3	1.5	0.8	1.8	1.0	1.3	1.5	1.3
	Median	0.6	0.1	0.7	0.4	0.6	1.3	0.8	0.6	0.8	0.6	0.5	0.3	0.6	0.5	0.7	0.9	
	SD	*2.1*	*0.4*	*2.4*	*1.6*	*3.2*	*2.6*	*2.6*	*1.6*	*3.0*	*2.0*	*2.3*	*1.7*	*2.5*	*1.1*	*1.6*	*1.7*	
	Min	0.14	0.06	0.10	0.07	0.07	0.09	0.11	0.08	0.10	0.07	0.07	0.08	0.10	0.06	0.06	0.13	
	Max	16.3	1.9	19.4	8.1	20.8	13.2	13.4	12.3	15.9	11.0	13.2	12.6	12.0	7.0	8.8	9.0	
Answer sheet	Mean	6.4	5.2	6.0	6.8	8.2	7.3	10.2	6.3	7.7	8.5	4.1	7.0	7.6	4.0	10.5	13.7	7.5
	Median	2.0	1.5	2.6	3.2	2.4	3.5	3.0	2.0	3.1	2.0	1.8	2.8	2.2	1.1	6.1	3.8	
	SD	*10.7*	*9.1*	*9.1*	*9.8*	*14.7*	*11.0*	*16.7*	*12.3*	*12.4*	*21.2*	*5.6*	*10.9*	*14.0*	*8.3*	*14.9*	*24.7*	
	Min	0.10	0.06	0.10	0.09	0.06	0.11	0.07	0.07	0.07	0.06	0.17	0.09	0.07	0.06	0.08	0.09	
	Max	62.4	67.4	67.6	63.1	98.6	88.1	132.7	142.0	95.3	211.5	29.3	79.1	126.4	91.7	120.5	185.7	

between different source materials or from answer sheet to source materials and vice versa. Moreover, most of the participants' median visit durations within each AOI in the reading group were around a second, which means that half of these visits were around a second. This may imply that participants adopted more often an expeditious style of reading, for example, searching for information that they thought would be useful in their writing. Interestingly, the median visit duration within the answer sheet for each participant is much less than the corresponding mean. Ten participants' median visit duration on this AOI in less than three seconds, meaning that half of the visits lasted less than three seconds. This again provided evidence for participants' looking behaviour, i.e., they constantly and frequently went to look at the instructions and source materials while writing.

4.5 Relationships between eye-tracking measures and test-takers' performance on TBEM-8 reading-to-write task

In order to gain further insights into the eye-tracking data, Spearman's rank-order correlations were run to look at if there were any relationships between the results of eye-tracking measures presented in Section 4.4 and the participants' performance on the TBEM-8 reading-to-write task (RQ2b).

First, to explore how the attention paid to reading and writing correlate with the participants' performance, correlations were calculated between two of the eye-tracking measures, i.e., *total visit duration* and *visit count*, on the AOI reading group (the task instructions and five source materials) and writing group (the answer sheet), and the 16 participants' scores on the TBEM-8 reading-to-write task (IELTS test scores were also added as dependent variables to identify potential trends). Table 4.20 shows the results of this correlation analysis.

Table 4.20: Spearman's rank-order correlations between eye-tracking measures on different AOI groups and the participants' performance on TBEM-8 reading-to-write task and IELTS test

AOI groups	Scores	Total visit duration	Visit count
AOI reading group	TBEM-8 reading-to-write task	.137	-.194
	IELTS overall	-.095	-.322
	IELTS reading	-.166	-.073
	IELTS writing	.012	-.110
AOI writing group (Answer sheet)	TBEM-8 reading-to-write task	-.367	-.026
	IELTS overall	-.370	-.042
	IELTS reading	-.405	.135
	IELTS writing	-.100	.385

As can be seen in Table 4.20, there were no statistically significant correlations (at the 0.05 level) between these variables. The only correlation approaching the 0.1 level of significance was found between the participants' L2 reading proficiency (scores on the IELTS reading component) and their *total visit duration* on the answer sheet (ρ=-.405, p=.120); higher reading ability participants spent less time on writing, indicating that participants' L2 reading proficiency may be a factor that affects their writing process. No significant correlation was found between participants' total visit duration on the answer sheet and their L2 writing proficiency (scores on the IELTS writing component), and the correlations between the time spent on reading and participants' performance on the TBEM-8 reading-to-write task as well as other IELTS test scores are weak and not significant (p>0.5). Similarly, results on *visit count* did not report any statistically significant correlation between these measures. All correlations were in a weak and negative direction except the ones between the number of times that participants visited the answer sheet and their scores on the IELTs reading and writing components.

To further explore the relationships between the results of eye-tracking measures and the participants' performance on TBEM-8 reading-to-write task, Spearman's correlations were calculated between four eye-tracking measures (*total visit duration*, *mean visit duration*, *max visit duration* and *visit count*) on each individual AOI in the reading group and the participants' scores on the reading-to-write task (IELTS test scores were also included as dependent variables). The results are displayed in Table 4.21. The correlations that are statistically significant at the 0.05 level have been highlighted yellow in the table. Considering the relatively small sample size in this study, which may "have masked some smaller yet extant effect sizes" (Brunfaut and McCray, 2015, p. 34), the correlations that are statistically significant at the 0.1 level have also been highlighted in grey for consideration, following the method taken by Brunfaut and McCray (2015).

As shown in Table 4.21, no significant correlations were found between the four eye-tracking measures on the instructions and the first two source materials, and participants' performance on the TBEM-8 reading-to-write task. There was a light trend that participants who scored higher in the IELTS writing task had a longer *max visit duration* on the third source material (ρ=-.442, p=.086), but those who visited Source 3 less frequently achieved higher IELTS overall scores (ρ=-.433, p=.094).

For the results on the fourth source material (picture), it is clear that all the statistically significant correlations (at the 0.1 level) are in a negative direction, suggesting that the time spent on the picture, as well as the number of visits in it, decreased as the ability of the participants improved. This might relate to the L2 writing proficiency of the participants, since most participants reported in their protocols their

Table 4.21: Spearman's rank-order correlations between eye-tracking measures on AOIs in the reading group and the participants' performance on TBEM-8 reading-to-write task and IELTS test

AOIs	Scores	Total visit duration	Mean visit duration	Max visit duration	Visit count
Instructions	TBEM-8 reading-to-write task	.227	.174	.277	-.150
	IELTS overall	.242	.276	.153	-.332
	IELTS reading	.297	.088	.084	.033
	IELTS writing	.402	.294	.338	-.072
Source material 1	TBEM-8 reading-to-write task	.361	.009	-.029	.237
	IELTS overall	.349	.051	.129	.221
	IELTS reading	.117	-.158	-.067	.276
	IELTS writing	.092	-.053	-.058	-.026
Source material 2	TBEM-8 reading-to-write task	.206	-.046	.032	.061
	IELTS overall	-.155	.103	-.019	-.219
	IELTS reading	-.076	.002	.280	-.194
	IELTS writing	-.181	-.141	.047	-.028
Source material 3	TBEM-8 reading-to-write task	-.183	.417	.388	-.153
	IELTS overall	-.253	.236	.122	-.433
	IELTS reading	-.336	.001	.140	-.271
	IELTS writing	-.239	.338	.442	-.082

Source material 4 (picture)	TBEM-8 reading-to-write task	-.448	-.057	-.024	-.458
	IELTS overall	-.193	-.042	.047	-.171
	IELTS reading	-.210	-.335	-.163	.065
	IELTS writing	-.449	-.202	-.165	-.448
Source material 5 (key concepts and expressions)	TBEM-8 reading-to-write task	-.168	.329	.499*	-.429
	IELTS overall	-.179	.210	.375	-.319
	IELTS reading	-.158	-.075	.220	-.234
	IELTS writing	.036	-.008	.226	.021

*. Correlation is significant at the 0.05 level.

attention on the picture was unconscious during which they were actually either making micro-plans for the sentence to be written or trying to translate abstract ideas into linguistic forms; or because the higher proficiency participants were more likely to quickly disregard the picture as less relevant overall.

The only statistically significant correlation at the 0.05 level (ρ=.499, p=.049) was found between the participants' *max visit duration* on the fifth source material (key concepts and expressions) and their performance on the TBEM-8 reading-to-write task; the *max visit duration* of the higher-scoring participants lasted longer. Interestingly, however, the correlation between the participants' number of visits in this material and their performance shows a negative direction though less strong towards significance (ρ=-.429, p=.098). This may suggest that the higher-performing participants are more efficient in using the words or expressions provided to conceptualise the ideas in their writing; they attended to the key words for a longer amount of time, and needed to visit this AOI less often as a result.

4.6 Summary

This chapter has mainly reported results from the analysis of the eye-tracking data. First, a correlation analysis showed that the participants' scores on the TBEM-8 reading-to-write tasks were moderately related to their reading abilities. Second, a heat-map output presented that the participants' attention covered all the main parts of the task, with various focus within each part. Third, the four eye-tracking measures, i.e., *time to first fixation*, *total visit duration*, *visit count* and *visit duration* illustrated in much detail how these participants engaged with the source materials throughout task completion. Finally, the correlation analysis between the eye-tracking measures and participants' performance on the TBEM-8 reading-to-write task yielded no significant results (at the 0.05 level) except that a moderate positive correlation was found between the participants' max visit duration on the 'key concepts and expressions' and the reading-to-write scores.

The next chapter presents the results from the analysis of the stimulated recall data collected in the eye-tracking and stimulated recall study (Study I).

CHAPTER 5 RESULTS II: STIMULATED RECALL (STUDY I)

5.1 Introduction

To gain insights into test-takers' cognitive processing while completing the TBEM-8 reading-to-write task, stimulated recall protocols were gathered immediately after the participants finished the task. These protocols were first transcribed and divided into short segments which were then coded according to a coding scheme based on Shaw and Weir's (2007) model of writing processes and Spivey's (1990, 1997, 2001) discourse synthesis model (see Section 3.3.4 for details of the coding scheme), as well as codes which were generated from the data itself. Section 5.2 summarises the results of coding and examines each type of cognitive processes in detail with quotes from participants' stimulated recalls. Section 5.3 then looks at the relationships between the use of these cognitive processes and participants' performance on the TBEM-8 reading-to-write task. A summary of this chapter is provided in Section 5.4.

5.2 Cognitive processes employed during task completion

Tables 5.1 and 5.2 shows the results of coding. Together they inform the answer to the first research question (RQ1), that is, what cognitive processes do test-takers employ while completing the TBEM-8 reading-to-write task. In Table 5.1, the coding categories, which represent cognitive processes and subprocesses, run down the left side of the table, while the frequency and percentage of each process are displayed in columns on the right-hand side of the table (see Table 3.3 for the definitions of each code). In Table 5.2, the number of each type of cognitive process employed by individual participants are displayed.

Table 5.1: Stimulated recall results on cognitive processes employed to complete the TBEM-8 reading-to-write task (n=1,956)

Cognitive processes/subprocesses		Frequency (n=1,956; *100%*)	
		f	%
Text interpretation (TI)	TI-1	131	*6.7*
	TI-2	375	*19.2*
	TI-3	162	*8.3*
Task representation (TR)		81	*4.1*
Macro-planning (MacP)		27	*1.4*
Organising (O)		60	*3.1*
Connecting and generating (CG)		91	*4.6*
Selecting (S)		307	*15.7*
Micro-planning (MicP)		164	*8.4*
Translating (T)		33	*1.7*
Monitoring (M)	M-1	155	*7.9*
	M-2	115	*5.9*
Revising (R)		111	*5.7*
Other processes	Commenting (C)	127	*6.5*
	Transcribing (TB)	17	*0.9*
Total		1,956	*99.9*

As shown in Table 5.1, a wide range of cognitive processes (1,956 instances; these processes represent the total number of processes across all participants) were reported by the participants as they completed the TBEM-8 reading-to-write task. This table allows us to consider the general patterns in the data, which will be reported here, following which each coding category will be discussed in more detail. Table 5.2 shows that the total number of these cognitive processes split evenly across participants, with a mean of 122.3 (*SD*=29.3), ranging from 89 to 209.

Overall, the most reported process (375 instances, 19.2 percent of all instances of cognitive processing) was *text interpretation-2* (TI-2), i.e., participants reading the source materials. This is in line with the results from the eye-tracking analyses that the participants spent considerable time looking at the source materials. The second most frequently reported cognitive process (307 instances, 15.7 percent) was *selecting* (S),

Table 5.2: Stimulated recall result on cognitive processes employed by each individual participant

P	Cognitive processes															Totals
	TI-1	TI-2	TI-3	TR	MacP	O	CG	S	MicP	Trans	M-1	M-2	R	C	TB	
1	17	34	22	11	2	10	11	28	27	8	10	16	6	7	0	209
2	8	26	7	5	1	7	5	14	11	0	1	10	8	15	1	119
3	10	33	8	5	0	5	4	19	24	3	3	11	10	10	1	146
4	8	25	10	7	5	3	1	11	11	1	3	20	9	11	3	128
5	3	19	12	3	4	3	3	7	11	3	3	8	5	13	2	99
6	8	18	9	5	2	7	4	10	13	1	3	6	5	0	0	91
7	6	28	4	4	1	0	7	21	5	2	2	19	12	2	0	113
8	13	26	19	6	0	2	5	25	7	0	6	24	12	3	1	149
9	6	30	10	2	0	2	1	24	4	1	4	11	6	5	0	106
10	5	22	19	8	5	4	8	15	11	0	5	18	10	5	2	137
11	6	14	1	4	2	6	9	20	8	1	1	2	0	11	4	89
12	11	22	12	6	1	1	10	26	5	1	5	11	7	9	0	127
13	10	17	11	2	1	3	10	32	4	3	2	8	3	3	0	109
14	10	22	2	4	0	3	3	20	3	0	11	12	4	6	1	101
15	4	24	10	1	1	0	6	21	11	7	0	18	10	12	0	125
16	6	15	6	8	2	4	4	14	9	2	7	10	4	15	2	108
Total	131	375	162	81	27	60	91	307	164	33	155	115	111	127	17	1956

i.e., participants selecting ideas or information from the source materials. It is important to note that a total of 34.9 percent of all reported instances of cognitive processing were dedicated to reading the source materials (TI-2) and gathering information from them (S). This percentage is similar to that of *total visit duration* (26.4 percent) on the reading texts, suggesting that participants paid close attention to the source materials. This in a way also implies that, at a very broad level, the participants seemed to understand the task requirements in the instructions that they should write an essay based on the given materials.

Micro-planning (MicP) was the third most used cognitive process (164 instances, 8.4 percent); the participants reported, in many cases, that they were planning, at either the sentence level or the paragraph level, for the text about to be produced. Their eye traces also indicated that they seemed to pause often before starting to write a new sentence, or when they were in the middle of writing a sentence, during which they usually went back to read the instructions, source texts or sentences they had written.

Another common process reported by these participants is *text interpretation-3* (TI-3, 162 instances, 8.3 percent). For this process, the participants reported looking back at the text they had produced. It was explained in some L1 writing studies (Chenoweth and Hayes, 2001; Hayes and Berninger, 2009) that when people write, they typically generate text in short "language bursts" of six to twelve words separated by pauses that may involve either planning of the next segment to be produced or evaluation of the text written so far (Das and Misra, 2015). This process may not differ as much as other processes across different types of writing tasks. *Monitoring-1* (M-1) accounted for 7.9 percent of all instances of cognitive processing, which is slightly less than the percentage of TI-3. Monitoring refers to a type of process when the participants reread the text that had been written to check the mechanical accuracy of the text (low-level monitoring), for example, spelling, grammar and syntax.

131 instances (6.7 percent) of *text interpretation-1* (TI-1) process were found in the protocols. The participants seemed keen to refer to the task instructions for different purposes at different stages of writing. This behaviour has been confirmed by looking at participants' eye-tracking recordings that they did reread the instructions many times throughout the whole task completion. The most revisited part of the instructions, according to the participants' verbal reports, is where the requirements for the expected content in the written product are stated: "*In the essay, you should describe the event, analyze the situation and comment on the impact of Jobs' resignation*"; this can also be supported by some eye-tracking evidence, for example, the heat map output shown in Figure 7.1. *Monitoring-2* (M-2) comes after TI-1 as the seventh most often used process (115 instances, 5.9 percent), in which the participants were monitoring at a more advanced level (high-level monitoring), i.e., examining the text produced to determine the extent to which it accords with the writers' macro- or micro-plans, its relevance to and adequacy for the task set and the development of arguments.

The above seven cognitive processes/subprocesses comprise over 70 percent (72.1 percent) of all instances of cognitive processing in the participants' stimulated recalls. The remaining cognitive processes were used more rarely, though they were also essential to the participants' reading-to-write process. *Revising* (R, 111 instances, 5.7 percent) is a process which is involved, basically, in any type of writing tasks; writers revise what they have written after they have produced a certain amount of text, in order to either improve the quality of the text or make it more suitable for a particular purpose. 91 instances (4.6 percent of all instances) of the *connecting and generating* (CG) process were found in the protocols. The participants reported that they, not infrequently, generated new thoughts by connecting the ideas in the source materials with knowledge they retrieved from memory, or created links, at least, between these ideas. This process is essential to our understanding of how test-takers write from source materials, since it reflects, to a large extent, the interaction between reading and writing, which will be further investigated later in this section.

The participants reported that they were trying to understand or clarify the task demands at any stage during task completion (81 instances, 4.1 percent). This process (*task representation,* TR) has been examined with L1 writers (Flower et al., 1990) as well as L2 writers (Ruiz-Funes, 1999, 2001; Allen, 2004; Wolfersberger, 2007; Plakans, 2010), and proved to be an essential step for producing a written product in any task setting. 60 instances (3.1 percent) of the participants' *organising* (O) process were found; only one participant (Participant 15) did not report any instance of organising. This process is a bit different in use between independent and integrated writing tasks that in the latter case test-takers may not only engage in ordering the ideas they would like to put into the text (often at the macro-level), they also use strategies to organise the relationships between ideas in the source materials for the text to be produced (Spivey, 1991).

The three least reported processes (*commenting* will not be discussed here as it is not relevant to the focus of this study) were *translating* (T, 33 instances, 1.7 percent), *macro-planning* (MacP, 27 instances, 1.4 percent) and *transcribing* (17 instances, 0.9 percent). It should be noted that stimulated recalls presented only part of the participants' cognitive processing, and thus the low frequency of processes do not necessarily mean that they were seldom used in completing the task, for example, the process of *translating*, although it is important in writing (Field, 2005; Shaw and Weir, 2007), may not be adequately reported as participants tend to be less aware of the use of this process due to its automatized nature. It is also interesting to find that a few participants reported that they were not familiar with the keyboard provided during writing, which caused a bit of disturbance to them that they had to look down to the keyboard time and time again while typing. This indicates that *transcribing* may, to some extent, affect the participants' writing performance.

In summary, Table 5.1 illustrates a general picture of the cognitive processes these participants employed to complete TBEM-8 reading-to-write task. Findings indicate that the whole range of targeted cognitive processes were used, with varying frequencies, throughout the whole task completion. Thus, it may safely be concluded that the TBEM-8 reading-to-write task elicited all types of cognitive processes proposed in the literature review, from the majority of participants. Each of these processes will be further investigated to provide a qualitative account of their nature. The sequence will follow the order as presented in the left column of Table 5.1, with quotes from participants' stimulated recalls, in Sections 5.2.1 to 5.2.9 below.

5.2.1 Text interpretation

Text interpretation is a process that creates "internal representations from linguistic and graphic inputs" (Hayes, 1996, p. 13). In the context of an independent writing task, the text to be interpreted normally includes the text in task instructions (TI-1) and the text writers have written (TI-3), while in a typical reading-to-write task, the text in source materials (TI-2) is also added into the whole text and thus resulting in differences in writers' cognitive processing during task completion.

The whole process of completing the TBEM-8 reading-to-write task was, for ease of analysis, divided into three phases: before writing, during writing and after writing (after the completion of first draft). The participants' protocols were then parsed to differentiate the use of text interpretation process between these phases. Table 5.3

Table 5.3: *Text interpretation-1* at different phase of writing by participant

Writing phase	Participant																
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	Total
Before writing	6	6	9	3	2	6	3	6	3	5	4	3	5	4	3	6	74
During writing	9	2	1	3	1	2	3	6	2	3	1	8	3	6	0	2	52
After writing	1	0	0	2	0	0	0	0	1	0	0	0	1	0	0	0	5
Total	16	8	10	8	3	8	6	12	6	8	5	11	9	10	3	8	131

shows the number of occurrences of the participants reading the task instructions (TI-1) at different phase of writing. Before writing, they started by reading through the entire instructions carefully to create an initial understanding of the task, for example, Participant 8 reported that "*I read through the instructions very slowly and paid close attention to what I should write and how many aspects I should cover...*". Most of the participants reread the instructions several times before moving on to the source materials; this may be due to the complexity of instructions in reading-to-write tasks that test-takers may spend more time creating a task representation than they do in an independent writing task. This can also be supported by participants' eye-traces that they read back and forth between the instructions and source materials in the first five minutes of task completion. Participant 13 explained this kind of looking behaviour in her recalls:

> *I was reading the second source material, and then I went back to read the instructions again, I wanted to make sure what this material was for, and what was the connection between it and the instructions, then I could decide in which part (of the essay) I could use the information in this material.*

Some participants also claimed that reading the source materials imposed an extra cognitive load on their minds that they forgot what the instructions were about and thus reread the instructions either during or after reading the source materials.

During writing, the participants most often used an expeditious form of reading: scanning, to locate specific information in the task instructions, as evidenced in their eye-tracking recordings and the protocols, for example, Participant 12 went back to read the instructions while composing, "*I was talking about resignation, so I had a look back to the instructions and found that it wanted me to discuss the impacts of Jobs' resignation, this was to make sure that I was on the right track*". The purpose of reading the instructions at this phase is, to a large extent, monitoring the progress of writing, more specifically, to check if the text written so far is not deviating from the topic specified in the task instructions, and to determine whether writing plans need to be modified for the text to be produced. After completing the first draft, few participants reported that they revisited the task instructions when they re-checked if the essays fulfilled the requirements of the task. This may be partly because most of the participants tended to monitor their writing frequently as they wrote, but less often after finishing the draft. This will be looked at in more detail in Section 5.2.8 which is on monitoring process.

Reading the source materials (TI-2) was the most reported process of text interpretation (375 instances, 19.2 percent) in the participants' stimulated recalls. Table 5.4 shows that they reported the most instances of TI-2 (271 instances, 72.3 percent) while they were writing; 98 instances (26.1 percent) were devoted to comprehending the source materials before writing and only two participants mentioned that they did read the materials after finishing the first draft. Before writing, the form of reading adopted by these participants was, as evidenced in eye-tracking recordings, mostly careful reading. They read through the source materials in a slow and careful manner, particularly when they were reading the English source material (Source 2). The organising process (using strategies to understand the structure of readings) was also engaged in reading activities during this phase, for example, Participant 1 summarised the main points in different source materials when she was reading, "*...I found that there were some similarities as well as some differences in these source texts, so I thought I might need to think critically on this issue, I re-evaluated the requirements of the task*".

Table 5.4: *Text interpretation-2* at different phase of writing by participant

Writing phase	Participant																
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	Total
Before writing	4	7	10	7	8	7	5	7	8	5	8	6	3	5	5	3	98
During writing	30	19	23	17	11	12	23	19	22	14	6	16	14	14	19	12	271
After writing	0	0	0	0	0	0	0	0	0	3	0	0	0	3	0	0	6
Total	34	26	33	24	19	19	28	26	30	22	14	22	17	22	24	15	375

During writing, the participants used scanning most often to locate the specific information in the source materials they considered useful in their writing, which is similar to the process of reading the task instructions during writing; for instance, Participant 1 told the researcher why she frequently went back and forth between Source 2 and the answer sheet when writing the first paragraph of her essay, "*…there were some key words I could add into my introduction, they helped to describe what kind of person Steve Jobs was, and what the impacts of his resignation were…they helped me to elaborate my points* ". Similarly, Participant 4 looked back to the same source material at some point during writing, but he stated that he was checking whether the word "resignation" was spelt correctly in his writing.

As expected, rare instances of reading the source materials were found after writing. This may be because of the time limit under testing situations that test-takers may focus on examining the textual quality of their written products such as accuracy of spelling, word use and sentence structure rather than the appropriateness of the content which probably needs spending time to refer back to the source materials.

Reading the text that had been written (TI-3) is the last subprocess of text interpretation found in the participants' stimulated recalls. Unlike the other two processes discussed above, this process may be expected not to differ much between independent and integrated writing tasks, as it is by nature more associated with test-takers' writing abilities rather than the integration of reading and writing skills. As shown in Table 5.5, most of the TI-3 processes (133 instances) were reported during writing; the participants said they were trying to plan for the text to be produced by reading the text that had just been written. For example, Participant 5 stated that "*I didn't know what to write in the concluding paragraph, so I went back to have a look at what I had written*". Also, they reported that they were checking the qualities of the

Table 5.5: *Text interpretation-3* at different phase of writing by participant

Writing phase	Participant																
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	Total
Before writing	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
During writing	21	3	8	8	12	7	3	17	8	13	1	11	7	1	9	4	133
After writing	1	4	0	2	0	2	1	3	2	6	0	0	4	1	1	2	29
Total	22	7	8	10	12	9	4	20	10	19	1	11	11	2	10	6	162

text produced. This process of text interpretation will be discussed in more detail with other process (micro-planning, monitoring etc.) in later sections.

5.2.2 Task representation

Task representation is an interpretive process during which test-takers create an understanding of the task demands. It is important because test-takers' performance is dependent on how they understand and approach the task. The 16 participants' protocols were coded as they were done in the above section, to differentiate the use of task representation processes between different writing phases (see Table 5.6). The results are in line with the findings in Table 5.3 that the most instances of task representation occurred before participants started to write, fewer instances were found during writing and only four participants reported that they revisited the instructions after completing the first draft.

Before writing, with only one exception (see Figure 4.3), the participants started by reading through the instructions a first time to construct an initial understanding of the task demands, for example, Participant 6 read the instructions carefully, word by word, instantly after he had a quick browse of each part of the task at the beginning, "*...I was reading the instructions, because they are very important, the title of the essay and task requirements were provided, I read closely this part...*". Following this first reading, participants then moved on to read the source materials, during which they were found, through their eye-traces, going back to read the instructions a second or even more time. Participant 11 explained her circular process of reading and rereading the instructions:

Table 5.6: *Task representation* at different phase of writing by participant

Writing phase	Participant																
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	Total
Before writing	3	4	4	3	2	3	2	3	1	5	4	2	1	1	1	6	45
During writing	7	1	1	3	1	2	2	3	0	3	0	4	0	3	0	2	32
After writing	1	0	0	1	0	0	0	0	1	0	0	0	1	0	0	0	4
Total	11	5	5	7	3	5	4	6	2	8	4	6	2	4	1	8	81

> *I went on to read the following materials and went back to reread the instructions and then I understood what this task wanted us to write, especially about how many specific parts we should cover in the essay, for example, analysis of the situation and comments on Job's resignation.*

During writing, most of the participants devoted less time to *task representation* than they did before starting to write. Two major patterns emerged when examining the protocols at this phase. First, participants revisited the task instructions for support on text they were about to produce. For example, Participant 1 recalled, "*...I was wondering what to write in the first sentence, then I went to read the first source material and the instructions, the instructions said that you should describe the event, so I thought I needed to describe it...*". Second, as discussed previously, the participants went back to the instructions to check if they went off the topic, for example, Participant 12 explained, "*...when writing essays in Chinese, I always remind myself to stay on the topic. I worried about deviating from the topic, I thought it would be a serious problem, so I looked back to the instructions time and time again*". Four participants reported that they engaged in the task representation process after they finished the first draft. Participant 4 was one of them: "*now I had done checking the grammar, tense, vocabulary etc., then I went back to look at the instructions again to check if I missed any points.*"

To sum up, the participants used task representation processes during all phases of writing, especially before writing, when they read through the instructions for an initial understanding of the task demands, and during writing, when they revisited the instructions for either guidance on the text to be produced or monitoring the progress of their writing. When they looked back to the instructions, they seemed to spare relatively little time for the contextual constraints of the task such as input length and time limit, while they paid most of their attention to the content demands, i.e., what content is expected in the written product, either before writing, during writing or after writing.

Their protocols also indicated that the use of source materials in integrated writing tasks may complicate test-takers' task representation process, which agrees with findings in other research (Plakans, 2010; Wolfersberger, 2007). Participants in this study provided several reasons for their effort to understand the task. First, they reported that they had little experience with writing from source materials in a test setting, for example, Participant 2 commented on her recursion in the task representation process, "*…then I read the materials and looked back to the instructions sometimes, because I have not done this kind of task before, so I needed to go back to reconsider its requirements*". Second, due to the extra cognitive load of reading the source materials, their working memory seemed not capable to hold the content of the task instructions, so they tended to forget the information in the instructions. For instance, Participant 14 reported "*…After reading (the source materials), I forgot what I was required to do, and then I went back there (the instructions) and checked it out again*". Lastly, although some participants claimed that the source materials were helpful for understanding the task, for example, as described earlier, Participant 11 seemed not able to construct an accurate task representation in her first reading of the task instructions, but reading the source texts facilitated her understanding of it. However, there are other participants who thought the inclusion of source materials made their conceptualization of the topic even more complicated, for example, Participant 12 stated, "*…After reading the source materials I became confused about what the task wanted me to write*", then she returned to the instructions and reread the source materials several times for clarification, which makes her task representation process rather complex. This may be due to the participants' reading proficiency in that if they had trouble building a representation of the source materials or the instructions, they were less likely to comprehend the relationship between the materials and task, and thus hindered their task representation and other relevant processes, for example, macro-planning.

In summary, evidence from eye-tracking and stimulated recall reveals that, to complete the TBEM-8 reading-to-write task, test-takers engaged in task representation processes during all phases of writing, proving that this process is not a single, simple act, but an extended, repetitive interpretive process throughout the task completion. The inclusion of source materials may complicate test-takers' task representation process by introducing more reading into the process of writing, calling for more interaction between these two skills.

5.2.3 Macro-planning

As discussed in the above section, task representation is an important process of creating an understanding of the task before starting to write, another process related to this phase is the process of macro-planning, in which writers plan for writing goals and content, and identify major constraints of a task such as genre and the level of formality required, the target readership, etc., on the basis of their representation of task (Field, 2004; Shaw and Weir, 2007).

Twelve participants reported that they used macro-planning processes (27 instances) while completing the TBEM-8 reading-to-write task. Almost all of these instances occurred before the participants started to compose and they were mainly concerned with goal setting and consideration of content, for example, Participant 5 said that she "*divided the content requirements in the instructions into several parts, and planned for the paragraphs to be written according to the focus in each part…*", and Participant 6 reported that "*when I was reading (the source materials), I already* *began to think about what content I should wrote, because I thought that I did not have to write everything in the materials, for example, I decided not to write about Jobs' death*". No instances of consideration of the genre and target readership were found in the protocols, which indicates that the participants may not have been aware of the importance of these two aspects in successful task completion.

It should be noted that the macro-planning process employed when completing the TBEM-8 reading-to-write task requires more of participants' reading abilities because of the inclusion of source materials, making this process more complex, as compared to macro-planning in completing an independent writing task. Many participants were found building connections between source materials and macro-plans for their writing, for example, Participant 11 reported that:

> *I found that the most important thing was to integrate information in these source materials to my writing, and it was not necessary to include many of my own opinions, so I reread the first three materials and categorised them to decide in which paragraph of the essay their information can be put into.*

Also, participants were found referring to the task instructions when making macro-plans before writing, for instance, Participant 10 said that he started to make plans for writing after he "*understood the content and structure of the source materials*" and "*referred constantly to the content requirements in the instructions*", proving that the process of macro-planning can be influenced by the constraints of the task (Grabe and Kaplan, 1996; Shaw and Weir, 2007).

In short, macro-planning processes were employed by the participants while completing the TBEM-8 reading-to-write task; they seemed to focus on the planning of the content of the text, while little effort was made to consider the target readership, or the genre and style of the piece. Participants' macro-planning process not only involved gathering of ideas, but also building connections between source materials and plans for writing to determine what and how the text was to be written to successfully complete the task.

5.2.4 Organising

During the organising process, writers were (1) organising the relationships between ideas in the source materials and/or (2) structuring their own writing. It was hypothesised that writers may have difficulty in generating coherent texts if they could not first assemble the texts they read into a unified coherent whole; in other words, they would need to build a representation of source materials in order to build a representation of their own text (Stein, 1990).

The 16 participants' protocols were parsed to differentiate the use of the two types of organising process between different writing phases. The number of occurrences for each type of organising process are displayed in Table 5.7 and 5.8. As shown in the tables, a total of 60 instances (3.1 percent) of using organising processes were found in the participants' stimulated recalls. Both before and during writing, the participants spent time organising (42 instances) to support their reading. Before writing (25 instances), they not only tried to comprehend the ideas in the source materials, but also to discover the relationships between these ideas for the text they were about to produce. For example, Participant 2 recalled her organising process before starting to write:

> *I was reading through the source materials, there were some relationships between them, some of them were talking about the same issue…there were two points of view in these materials, one was that Apple Company would not change after Jobs' resignation, the other one was that his resignation would have impact on the company, I categorised these materials into the two sides.*

Table 5.7: *Organising*-thinking about the structure of the source materials at different phase of writing by participant

Writing phase	Participant																
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	Total
Before writing	3	4	4	0	2	5	0	1	0	1	3	0	1	0	0	1	25
During writing	4	3	0	2	1	0	0	1	1	2	0	0	2	1	0	0	17
After writing	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Total	7	7	4	2	3	5	0	2	1	3	3	0	3	1	0	1	42

Table 5.8: *Organising*-thinking about the structure of their writing at different phase of writing by participant

Writing phase	Participant																
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	Total
Before writing	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	2
During writing	3	0	1	0	0	1	0	0	1	1	3	1	0	2	0	3	16
After writing	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Total	3	0	1	1	0	2	0	0	1	1	3	1	0	2	0	3	18

Over half of the participants reported that they also spent time trying to understand the source materials during writing (17 instances), using different strategies such as summarising the main ideas for each reading and identifying rhetorical structures (Plakans, 2009b). For instance, Participant 2 said:

> *I was looking for descriptions of Jobs, what kind of person he was, how talented he was, I was looking for materials about Jobs, the first source material described the Apple Company, so I did not look at it, the other two source texts talked about Jobs. I went to the third material to look for descriptions of Jobs and comments on him.*

And Participant 10 stated, "*I was reading these two paragraphs (in the third source material), because they seemed to talk about the same issues, so I kept reading them back and forth and tried to summarise main ideas from them*". No instances of the organising process were found after the participants finished writing the first draft.

18 instances were devoted to organising the ideas to be put in the essays (see Table 5.8). It is worth noting that only two participants (Participants 4 and 6) reported in their protocols that they were thinking about the overall structure of their writing before they started to produce any text. For example, Participant 6 said, "*I was reading (the source materials) and thinking about how to write the introduction part, the transition paragraphs, and the ending paragraph*". Most of the instances of structuring the essays were found during writing, when participants either referred back to the source materials to build a clearer representation of the input texts and then continued to think about the structure of their own writing in order to incorporate different ideas from the materials, or they just went about structuring the essays based on their own evaluations, for example, Participant 16 stated, "*I thought these two paragraphs were both descriptive ones, it's not necessary to separate them, so I put them together, and then I*

decided to start a new paragraph and to analyse the situation".

In summary, the participants did appear to spend time thinking about the structure of the source materials while completing the TBEM-8 reading-to-write task, especially before they started to produce any text, however few of them had worked out a rough outline of key points to include in their writing, instead, most participants seemed more prone to using strategies to understand the source materials *during* writing, while concurrently structuring the text to be produced. This is likely due to the complexity and difficulty of the organising process in completing an integrated writing task which includes source materials; test-takers may not be able to create a complete representation of the input texts instantly, which impedes the progress of building their own text on the source materials.

5.2.5 Connecting and generating

Connecting and generating is a process in which writers bring what they already know into the reading and create meaning-enhancing additions (Levin, 1988). In other words, writers combine their prior knowledge with the ideas they select from source materials to generate new meaning (Kucer, 1985; Spivey, 1987, 1997). As they connect and generate during the reading phase, they are "creating a pool of ideas from which to draw during the writing process" (Stein, 1990, p. 147). To some extent, then, the selecting, connecting and organising of ideas that occurs during reading may become the basis of plans for their writing.

As shown in Table 5.1, the 16 participants reported 91 instances of connecting and generating through task completion. The majority of these instances occurred during writing; few participants reported that they connected ideas in the source materials with their own knowledge while reading the materials before writing. This may be because of the limitation of the stimulated recall methodology that participants may not be able to recollect adequately their thought processes at the beginning of task completion due to memory decay. On the other hand, it may also be the case that the participants did actually not engage in the connecting and generating process before writing, instead, they were prone to selecting, connecting and organising ideas from both the source materials and/or their memory during writing.

Analysis of the 16 participants' stimulated recalls indicates that the use of connecting and generating processes during writing mainly served two different purposes: to develop ideas already found in the source materials and to generate new ideas. First, almost all of the participants reported that they used connecting and generating as a means of elaborating ideas found in the source texts (73 instances), when they selected relevant ideas from the input texts, combined them with their prior knowledge, and generated further development of these ideas. For example, Participant 1, when writing the second paragraph of her essay, said:

> *I went to read some sentences that I had written in this paragraph, and was probably thinking about what to write for the next sentence. I found the 'market value' (in the word list) above, I thought I could write on this, and also I could write something about the 'advantage' above…*

This elaboration of ideas from the source texts became what she wrote in her final essay, that is:

> *the world will not be surprised to see this listed company suffer from a decreased market value without much differential competitive advantage over its counterparts any more.*

The other function of connecting and generating is to produce different kinds of additional materials, i.e., information not found in the source materials, much of which is at a very detailed level, for example, new words that participants include in their essays. In this analysis, however, the generation of new materials refers specifically to those instances when the participants were attempting to re-evaluate the ideas provided in the source materials, by connecting their own knowledge with them, and generated ideas from a new perspective. Thus, this function of connecting requires writers to elaborate more critically on the ideas found in the source texts. Six of the 16 participants recalled in their protocols that they engaged in this type of connecting and generating process (18 instances). The protocol of Participant 11 provides some good examples of it. For example, when reading the third source material before starting to write, she said:

> *when I was reading this material, I was thinking (about the topic), it says that the Jobs' resignation on Apple had no great impact on the development of the company, but in my point of view, considering the current status of Apple, there is great impact…so when I was reading this material, I was also thinking about if there were any counter examples that could prove that his resignation did impact on the development of Apple*.

This generation of new ideas during reading also became part of this participant's writing plan and content later when she was writing the third paragraph of her essay, where she wrote:

> *Jobs' leave can be a severe loss of Apple" and stated the reason: "since Jobs' personal charisma is a very significant identity icon of Apple and maybe even the power gathering all the talents together.*

In summary, there is evidence of the usefulness of the process of connecting and generating in the 16 participants' s stimulated recalls. It may lead to elaboration of the ideas found in the source materials, by applying the participants' prior knowledge (including world knowledge, experiences, preferences etc.) to what they were reading, and may also promote critical thinking, creating ideas from a new perspective, when participants used their prior knowledge as a basis for comparison and evaluated the validity of the propositions in the source materials. Eventually, the process of connecting and generating creates an individualised pool of ideas; as the participants planned their writing, and made decisions about what to write, they would select information from that pool. The protocols also show that most of the participants used connecting and generating processes during writing, when they constantly referred back to the source materials (this can also be demonstrated by the eye-tracking data) when producing text for their writing. This may be likely due to the difficulty of building a complete representation of these source materials at the start of task completion, and the fact that the connecting and generating is an ongoing process of meaning building, rather than a one-off act. Participant 10 provided an explanation on this:

> *…at the beginning, I first read through the materials, and generated an overall impression, which might not be completely clear, but I knew roughly what it was in different part of the essay I should write, then during writing, when I felt that some information might be missed, I would go back to the materials to check if there were any content worth adding into the writing.*

As the participants went on reading and writing, their representations of the source materials and the essay both became clearer and more complete, and thus a coherent written product was more likely to be produced.

5.2.6 Selecting

The above section has discussed the process of connecting, in which test-takers combined their prior knowledge with the ideas they 'selected' from the source materials to generate new meaning. The process of selecting will be investigated with more details in this section to gain insights into how the 16 participants used this process during task completion and what different purposes this process served

As shown in Table 5.1, a total of 307 instances (15.7 percent) of the selecting process were found in the protocols. They mainly served three purposes: to select ideas from both memory and source materials for connecting, to select information (often at the level of detail, for example, a specific word) from source materials to support writing, and to select sentences from source materials for paraphrasing or translating. First, as discussed in the above section, when the participants composed from sources, they selected ideas from both their prior knowledge and source materials (91 instances), connecting them to generate either links between ideas or new meaning. By using this type of selecting, together with the process of connecting and generating, participants created a pool of ideas, from which they would like to draw during writing.

Another type of selecting, which was more frequently reported by the participants in their stimulated recalls, was to select specific information from the source materials to support the writing process (182 instances). This information was, most of the time, a certain word that the participants decided to bring into their own writing. For example, Participant 1, while writing a sentence in the second paragraph of her essay, said, "*I was thinking about what noun I can use to describe him, I found the word 'leadership' above, so I decided to use leader*". Similarly, Participant 3 explained why he went to look at the words in the 'key concepts and expressions', "*Then I went on writing, I wanted to talk about how the Apple company kept developing, I was wondering if there was any word in the 'key words' list that can be used in my writing, and then I found the phrase 'differential competitive advantage', so I added it into my essay*". These two examples indicate that the words in the source materials provided some lexical support for the participants' composing process, especially when they were trying to transcribe their abstract ideas into concrete linguistic forms. The participants also reported that they sometimes went back to the materials during writing to look for words that they could not recall correctly after their reading of the sources, for example, Participant 5 stated that "*I went back to the source material to find the word 'resign', I wanted to make sure whether there is an 's' or two 's's in it*", and "*I wanted to use the word 'charisma', so I went back to find it*". Last, searching for specific information and selecting it from the source materials was another common activity performed by these participants, when they found that certain information they read before was needed in their writing, and they would go back to the materials to look for it. For instance, Participant 8, when writing a sentence introducing the Apple products designed by Steve Jobs, said, "*I was going to write some examples, what products had been designed, so I went back to read the first paragraph*".

The third function of the selecting process discovered in the 16 participants' protocols was to select, paraphrase and/or translate original sentences in the source materials and integrated them into the participants' own writing. Fourteen participants reported that they engaged in this type of selecting during writing. It seemed that the participants were aware of the restraints in the instructions that they "should not simply copy and translate the source materials", so they adopted paraphrasing as a means of incorporating the sentences in the materials. For example, Participant 9 said, "*This sentence was to introduce what I was going to write in my essay. Basically, I paraphrased the task requirements in the instructions, so I would look at the instructions*" and:

> *I was looking for the information about his post, although the instructions said that "you should not copy", I was basically paraphrasing that sentence…so in this paragraph, I was composing sentences and looking at that source material*

at the same time. That's why my eye fixations frequently switched between these two areas.

Also, as the input materials included Chinese texts, the participants were found simply translating sentences in these materials and put them into their own writing, for instance, Participant 5 stated that "*I used the information in the third source material. I was translating those Chinese sentences*". Participant 13 even claimed that she copied one sentence in the materials, "*I was writing about its management team, and I just copied the sentence talking about that*".

In summary, when the participants composed from sources, there was ample evidence to support the notion that they used the selecting process to choose ideas from both memory and source materials, then connected them to generate new meaning that may have value for the content of their writing. Also, these participants were found, more frequently, to seek both lexical and syntactical support from the source materials during the process of translating (transcribing abstract ideas into linguistic forms), which concurs with the findings in the eye-tracking study, that they constantly switched their attention between different AOIs during writing. In the next section, the process of organising will be examined.

5.2.7 Micro-planning and translating

We have investigated, in the previous sections, the processes of macro-planning and organising, in which test-takers plan for writing goals and content, identify major constraints of the task (target readership, genre, style), and think about the structure of the readings and their own writing at the macro-level. But it is believed that planning and organising may also take place at the micro-level, i.e., at the sentence and paragraph level (Field, 2004; Shaw and Weir, 2007), during which writers plan for the goal, content and structure of a particular paragraph or an upcoming sentence, possibly with constant reference back to the macro-plans established earlier as well as the text produced so far.

164 instances (8.4 percent) of micro-planning process were found in the 16 participants' protocols, the majority of which occurred during writing. At the paragraph level (62 instances), all of the participants reported that they engaged in planning for either the content or the overall structure of an upcoming paragraph. For example, Participant 4, when monitoring the progress of task completion, said "*At this moment, based on what I had just read, I thought I should add another paragraph here talking about the current status, analysing the impact of Jobs' resignation on Apple company*", and Participant 1 stated that "*I was going to include two aspects of information in this paragraph, at this moment I was wondering which aspect I should write first, later I decided to first write about the impact on the company within itself*". At the sentence level (102 instances), participants were also found to plan for the content and structure of an upcoming sentence. For example, Participant 8 recalled that "*I was writing the topic sentence here, and I was thinking about using which sentence to state the topic…I went to the readings to find possible material, but later I decided to write it myself*", and Participant 6 reported on how he went about structuring a sentence, "*I was thinking about the structure of this sentence, should I write a simple sentence, or should I write a complex sentence with a relative clause…*".

It is worth noting that the participants were found often going back to read the instructions, source materials or the text that had been produced when they were micro-planning at both the sentence and paragraph level. First, for example, Participant 8 stated that she went back to reread the instructions when she finished writing the first paragraph of her essay, "*I was looking at the instructions and thinking about what to* *write in the next paragraph*"; the reason for this may be that the participant did not plan well at the macro-level, and so she may have needed to refer back to the instructions to create a clearer representation of the task and redo macro-planning before she continued to plan what content she was going to write in the next paragraph.

Second, when the participants reread the source materials while micro-planning, they were either selecting ideas from the materials to connect with their own knowledge to generate new meaning, or just choosing information they may need for their writing, and these processes of selecting and connecting may ultimately generate a micro-plan for the next paragraph or sentence. For instance, Participant 3 recalled that "*I had finished writing my first point of view in this part, and then I reread the source materials and the text I had written, and began to think about how to write the second point of view*", and Participant 1, when writing a sentence in the third paragraph of her essay, said:

> *I was thinking about how to write his traits, what words I could use to describe him, and then I went to source material to look for any possible words. I read through almost all the key words that might be useful in the source materials, and now I finally came up with some words in my mind, so I began to write this sentence.*

Last, the participants were also found frequently going back to reread the text that they had written, either before starting to write a new sentence or in the middle of producing part of an existing one. For example, Participant 10 recalled that "*I was going to write some comments, but it seemed to me that I had not fully described this incident in that sentence, so I went back to read what I had written and the source materials, and decided to add some information into the sentence*". It is natural that, as the participants went on writing, the evolving textual output became part of the context that they had to consider in order to drive further planning for the content to be produced.

The output of micro-planning was stored in the participants' mind in the form of goals at the sentence and paragraph level, which then became the bases of their translating process, during which the abstract ideas were transcribed into concrete linguistic forms. It is at this critical point that "the writer moves from an internal 'private' representation, which is abstract and only understood by him or her, to its expression in the 'public' shared code of language" (Shaw and Weir, 2007, p. 39). The participants reported 33 instances (see Table 5.1) of the process of translating, whose number was very likely to be underestimated, because, as discussed earlier in this chapter, translating may not be adequately reported as the participants tended to be less aware of the use of it due to its automatized nature, although it is an important process when producing the actual text during writing (Field, 2004, 2005). Also, the limitation of stimulated recall methodology may be another factor that hinders these participants' reflection on the translating process.

By looking at the limited instances of the translating process, together with the participants' eye-traces in the eye-movement recordings, we could see that this process was rather complicated and highly demanding in terms of cognitive processing in the context of L2 writing. This may because it is at this point that L2 writers face critical problems regarding the translation of abstract ideas for which they may not possess the necessary language resources. In other words, their knowledge of, for example, vocabulary and grammar of L2 may not be adequate to represent the ideas stored in the micro-plans. For example, Participant 12 stated in her protocol that, "*I was thinking about how to write this sentence, I already had an idea of what I was going to write, but I kept thinking about the language issues*". Problems related to language resources appeared to exert additional cognitive demands on the participants' translation process, during which they were very likely turning to the source materials provided in the task to seek both lexical and syntactical support, which was evidenced in their stimulated recalls and eye-tracking traces. For example, participants frequently, especially in the middle of sentence production, referred to source materials for extra support, which was often at a very detailed level, for example, looking for specific words or sentence structures, or even either copying, translating or paraphrasing the information in the input texts. The high cognitive demands of the translating process may, in turn, have hindered the execution of other processes such as connecting and organising, and thus have impacted on the quality of the final written product.

In summary, there is evidence of the 16 participants using micro-planning processes when they conducted planning and organising at the sentence and paragraph level, during which the processes of text interpretation, selecting and connecting were also frequently employed to facilitate the process of micro-planning. The output of the micro-planning process may be, to a certain degree, stored in the participants' mind in the form of abstract ideas, which were then likely to be translated into linguistic forms.

5.2.8 Monitoring and revising

So far, we have looked into how the 16 participants created representation of the TBEM-8 reading-to-write task, how they planned for the writing goals and considered the overall structure of the essay, to what extent and in what way they engaged with the source materials to aid their writing, and how they made micro-plans and translated their ideas into actual text. This section looks at the last two categories of cognitive processes that were found in the participants' stimulated recalls, i.e., the processes of monitoring and revising, which may not differ as much as other cognitive processes, between independent and integrated writing tasks.

The participants reported 270 instances (13.8 percent) of the monitoring process during and after writing, which related to two different levels of analysis. First, at a basic level (low-level monitoring), monitoring involves checking the mechanical accuracy of the text produced, such as spelling, punctuation, word use and syntax (Field, 2004, 2005; Shaw and Weir, 2007). 155 instances of monitoring processes, in the participants' protocols, were devoted to this level of analysis. The participants reported that they engaged in low-level monitoring at different stages of the text production process, either during or after writing a word, a sentence, or a paragraph, or even the first draft. The protocol of Participant 8 provides some good examples of this level of analysis. For example, when thinking about what to write in the next sentence, she said:

> *Most of the time, when I finished writing a sentence, I would go back to read the previous sentences to check if there were any grammar mistakes or any more content to be added into them", and "here I went to reread the first part of this sentence, and I thought the sentence structure was not appropriate, so I decided to delete it and changed to another way of expression.*

One thing to note is that it was not uncommon for the participants to refer back to the source materials when monitoring (this may be a difference of this process between independent and integrated writing tasks), during which they sought clarification of certain words or information. For example, Participant 4 reported that he went back to the source materials to check if he spelt the word 'resignation' right. Also, the participants were found to monitor the word count of their essays during writing. For example, Participant 10 recalled that, "*I read through the whole essay, and counted the number of words in it, because the instructions say that it is required to write 250-280 words, so I checked the word count*".

Second, at a more advanced level, the monitoring process may involve examining the text produced to determine the extent to which it accords with the writers' macro- or micro-plans, whether it is relevant to and adequate for the task set and whether it fits the development of the discourse structure of the text (Shaw and Weir, 2007). 115 instances of monitoring processes found in the participants' protocols were conducted at this level. For example, Participant 1, when writing the third paragraph of her essay, went back to reread what she had written in the previous paragraph, and claimed that "*I wanted to make sure what I wrote in this paragraph was different from that in the previous one*", and Participant 8 stated that "*I was wondering if I can add any more points into the writing, because I thought the opinions were not adequately described, so I wanted to add more analysis to…*". It should be noted that, in over half (66 instances) of the 115 instances, the participants reported that they referred back to the instructions to check if what they had written was relevant to the task set. This indicates that these participants paid considerable attention to this aspect of task fulfilment, and it may be because they were trained to do so, as Participant 12 said in her protocols, "*Do not go off the topic, this is what I learned from writing Chinese essays. If we go off the topic, there might be a problem, so I kept looking back to the instructions during writing to make sure I was on the right track*". Another function of high-level monitoring that can be found in the participants' protocols was to examine the cohesion and coherence of the text produced, for example, Participant 1 stated that:

> *I was reading through the paragraphs I had written, I was not checking the grammar points, but thinking about the connection between what I was going to write and the previous paragraphs, how can I write it in a logic way and make these two parts coherent…*

As discussed previously, the high cognitive demands on the translation process may hinder the execution of other processes, very likely including the process of monitoring. Field (2004) therefore argued that during the actual process of text production, writers' attention might be given to the lower-level features such as the accuracy of spelling, word use and syntax, but the higher-level features such as relevance to and adequacy for the task set are more likely to be monitored at a post-production stage. This is in line with the findings in this study with regard to where the instances of monitoring processes occurred, that the participants were found conducting the low-level monitoring literally at any stage of text construction, either during or after writing a word, a sentence, a paragraph, or a first draft, however, the high-level monitoring mostly happened when the participants had just finished writing a sentence or a paragraph (9 instances of high-level monitoring were found after the completion of a first draft). The 270 instances were also further analysed to differentiate the use of the monitoring process between different writing phases. The results show that participants engaged in the two levels of the monitoring process (low-level monitoring: 32 instances; high-level monitoring: 9 instances) after writing much less frequently than they did during writing (229 instances). This agrees with the results of the eye-tracking analysis that the participants spent, on average, limited time going back to look at what they had written after they finished the first draft of their essays.

As a result of monitoring activities, the process of revising may be conducted by writers at any level during or after writing, in which they return to aspects of the text identified as unsatisfactory and make corrections or adjustment (Shaw and Weir, 2007). Although all those aspects identified may not necessarily be revised, "it is very unlikely that revising occurs without monitoring" (Chan, 2013, p. 71). There are two levels of revising process, each corresponding to one of the two levels of the monitoring process: at the basic level, writers make revisions of issues relating to textual features such as accuracy or range of vocabulary, grammar and sentence structure; at the advanced level, writers deal with issues such as the development of arguments and coherence and cohesion.

111 instances (5.7 percent) of the revising process were found in the stimulated recalls, most of which were conducted at the basic level (92 instances). For example, Participant 8 deleted the word she wrote and changed to another one to avoid repetition, "*I found that I had already used 'however' in the previous sentence, so I deleted 'however' and changed it to another word*", and Participant 3 stated that he "*deleted this sentence, and put the adverb clause of time in the front, and then added a declarative sentence to make a simple description*". Nineteen instances of revising at the advanced level were found in the protocols, whose number was much less than that of the corresponding monitoring process (115 instances). Ten participants reported that they engaged in this level of analysis, for instance, Participant 10, when writing the first paragraph, stated that, "*at this time I was looking at the materials, and I found some important information that I had not described about, so I decided to add it into the first paragraph*".

In summary, there is ample evidence of the use of monitoring and revising processes in the 16 participants' stimulated recalls, most of which occurred during writing rather than after writing the first draft. The participants conducted monitoring and revising activities at both a basic level, when they dealt mainly with textual features such as spelling, word use and sentence structure, and a more advanced level, when they monitored issues such as relevance to the task, development of arguments and coherence and cohesion. But it appeared that they made many more revisions at the basic level (92 instances) than they did at the advanced level (19 instances).

5.2.9 Additional codes

The above eight sections presented the 10 main categories of cognitive processes the participants employed while completing the TBEM-8 reading-to-write task, which account for over 92 percent of all the instances of activities found in the participants' stimulated recalls. 144 instances did not fall into the above 10 categories, and two additional codes, *commenting and transcribing*, were arrived at to describe these activities.

First, *commenting*: the participants made a noticeable number of comments on their performance and behaviour in the task (127 instances, 6.5 percent), which are deemed not manifestations of their actual cognitive processing, for example, Participant 2 said that "*I read very slowly*" and "*I almost finished writing the second paragraph*". This is due to the nature of stimulated recall methodology, that it is an offline technique to look into cognitive processes: the participants sometimes commented on their behaviour during the experiment, rather than recalling their thought processes. Second, a few participants reported that the keyboard caused some disturbance to them during writing (17 instances). They claimed that they were not familiar with the keyboard provided, and they had to look down to it time and time again while typing, which may, to some extent, have affected their writing performance (see Chapter Eight for further discussion of limitations of the study).

5.3 Relationships between the use of cognitive processes and test-takers' performance on the TBEM-8 reading-to-write task

In Section 5.2, the analysis of the participants' protocols has shown the intricate interplay of the cognitive systems of reading and writing when completing the TBEM-8 reading-to-write task. It was found that the participants used a wide variety of cognitive processes, which served essential, and various purposes at different points in the reading-to-write process. For example, when a participant was planning the content to be written for an upcoming paragraph, they might go to the source materials to select information which would help them reach the writing goals, to decide which ideas or information to include based on their prior knowledge, and to make sure what they planned to write was relevant to the task set. It seems that it is these contexts, such as the task set, the participants' writing goals, and their prior knowledge, that affect the use of cognitive processes.

The qualitative analysis of the participants' stimulated recalls allowed us to explore the interaction between different cognitive processes in the reading-to-write process, but it has one important limitation, that is, it could only provide us single instances of cognitive processing, which could not be used to generalise about relationships between different categories of cognitive processes as a whole, and between a participant's use of these cognitive processes and their performance on the task. Thus, a quantitative analysis was performed to look at these relationships. Variables included the number of times a participant used a specific process and the quality of their written products (see Section 3.3.4 for details of rating). Spearman's rank-order correlations were calculated as the *p*-values for Shapiro-Wilk tests of normality for some of the variables were less than 0.05. Table 5.9 shows the correlation matrix between these variables. The correlations that are statistically significant at the 0.05 and 0.01 levels have been highlighted in light yellow and dark yellow; due to the small sample size, the statistically significant correlations at the 0.1 level have also been highlighted in grey for consideration.

Table 5.9: Spearman's rank-order correlations between the use of cognitive processes and the participants' performance on TBEM-8 reading-to-write task

Variables	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1. Task performance	--	.02	.43	-.06	-.23	-.16	.13	-.05	.13	.00	.30	.06	-.29	.01
2. Text interpretation-1		--	.30	.34	.58*	-.31	.23	.29	.45	.02	-.15	.14	.58*	.09
3. Text interpretation-2			--	.24	.12	-.43	-.09	-.06	.30	.25	.11	.72**	.05	.68**
4. Text interpretation-3				--	.28	.20	-.08	.26	.31	.26	.17	.23	.45	.28
5. Task representation					--	.40	.45	.17	-.13	.43	-.22	.12	.60*	.19
6. Mac-planning						--	.36	.10	-.51*	.47	.12	-.27	.15	-.19
7. Organising							--	.10	-.35	.60*	-.03	-.44	.08	-.39
8. Connecting								--	.62*	.04	.24	-.03	-.06	.00
9. Selecting									--	-.44	.20	.17	.13	.07
10.Micro-planning										--	.35	.12	-.01	.24
11. Translating											--	-.07	-.16	-.10
12. Monitoring-1												--	-.01	.93**
13. Monitoring-2													--	-.06
14. Revising														--

**. Correlation is significant at the 0.01 level.
*. Correlation is significant at the 0.05 level.

As shown in Table 5.9, the number of times the participants used a specific cognitive process does not correlate significantly (at the 0.05 or 0.01 levels) with their scores on the reading-to-write task, except for a moderate positive correlation (ρ=.43, p=.098) between the number of text interpretation-2 processes found and the quality of written products; that is, the more these participants reported that they read the source materials, the better they performed in the task. Such results may lead one to conclude that there are not any relationships between the use of processes and specific outcomes, but such a conclusion should be taken with caution, for the results may be attributable to the relatively small sample size that only 16 subjects were involved in this study.

The correlations among the processes of text interpretation-1, task representation and monitoring-2 are all significantly positive at the 0.05 level. This may be because these processes are very likely to co-occur during task completion; we have seen in the participants' protocols that they often went back to read the instructions during writing, re-clarifying the task demands, and monitoring the progress of their writing in case any adjustments were needed. Similarly, some slightly stronger positive correlations (significant at the 0.01 level) were found among the processes of text-interpretation-2, monitoring-1 and revising; again, this may because these participants combined these processes to reach a specific goal, which, in this case, was to search for information in the source materials to support their monitoring process, and/or make corresponding revisions (see Section 5.2.8). It is interesting to note that there is a negative correlation between the processes of macro-planning and selecting (ρ=-.51, p=.044), suggesting that the more these participants engaged in the macro-planning process before writing, the less instances of selecting process they reported. This may be due to the extent to which the participants created representations of the task and the source materials; if they understood well the task type and demands, and had created a clear representation of the source materials before starting to write, it might be easier for them to generate more complete macro-plans to be carried out and rely less on the source materials during writing.

5.4 Summary

This chapter has mainly presented results from the analysis of stimulated recall data collected in Study I. The findings showed that the participants employed a wide range of cognitive processes specified in Shaw and Weir's (2007) model of writing and Spivey's (1997) discourse synthesis model, with varying frequencies. Text interpretation-2 (reading the source materials), selecting and micro-planning are the three most frequently used processes according to these participants, and macro-planning and translating are the two least reported processes (but it might be due to the limitation of the stimulated recall methodology). The correlation analysis established no clear relationship between the number of times a participant used a specific process and the quality of the text he or she produced. It seems that the use of cognitive processes is dependent on the various contexts in task completion, such as different task representations, writing goals and the participants' prior knowledge.

The next chapter presents the results from analysis of the reading-to-write process questionnaire data collected in Study II.

CHAPTER 6 RESULTS III: QUESTIONNAIRE (STUDY II)

6.1 Introduction

This chapter reports the results from the analysis of data collected in the reading-to-write process questionnaire study described in Chapter Three. It starts with the results of an internal consistency analysis in Section 6.2, which demonstrates the extent to which each group of items in the main study questionnaire reliably measured the same type of cognitive processes. Section 6.3 then presents the participants' agreement rate for each item, and the results of a set of Mann-Whitney U tests which were run to examine if higher- and lower-scoring participants responded differently to each item in the questionnaire. Finally, Section 6.4 provides a summary of this chapter.

6.2 Internal consistency of each category of cognitive process (main study questionnaire)

As some revisions were made to the pilot questionnaire (see Section 3.4.1 for details), a Cronbach's alpha was run to understand the internal consistency of each group of items designed to measure the same type of cognitive process in the main study questionnaire. The overall reliability of each of the seven categories of cognitive processes, and each of the five hypothesised writing phases were obtained. Results are presented in Table 6.1. Items whose item-total correlations were lower than 0.30 are highlighted in grey.

Table 6.1: Reliability statistics of the main study questionnaire (40 items)

Item No.*	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted	Cronbach's Alpha
Conceptualisation phase					
Task representation and macro-planning					
1.1	15.95	4.919	.292	.525	.562
1.2	16.02	4.813	.377	.474	
1.3	16.17	4.887	.410	.459	
2.6	15.74	5.289	.304	.516	
4.4	16.03	4.945	.249	.555	
Overall reliability					.562
Meaning and discourse construction phase					
Text interpretation					
2.1	12.62	2.286	.244	.548	.533
2.2	12.47	2.552	.416	.404	
2.3	12.31	2.587	.361	.438	
4.5	12.54	2.225	.319	.467	
Selecting					
2.4	16.36	4.195	.298	.455	.518
2.5	16.54	4.226	.321	.441	
4.2	16.00	4.671	.348	.444	
4.6	16.30	4.308	.288	.461	
4.7	16.54	3.926	.229	.517	
Connecting and generating					
2.8	7.36	2.545	.522	.627	.710
2.10	7.51	2.480	.573	.564	
4.3	7.42	2.486	.491	.667	
Overall reliability					.726
Organising phase					
Organising					
2.7	21.87	9.954	.456	.594	.657
2.9	21.10	10.779	.381	.619	
2.11	21.08	10.660	.323	.635	
3.1	21.04	10.471	.347	.628	
3.2	21.16	9.394	.565	.558	
3.3	21.55	10.540	.270	.656	
4.1	20.87	11.704	.247	.652	
Overall reliability					.657
Low-level monitoring and revising phase					
Low-level editing					
4.13	19.13	13.488	.628	.870	.881
4.14	19.41	12.725	.705	.858	
4.15	19.33	12.455	.713	.857	
5.6	19.10	13.459	.664	.865	
5.7	19.29	12.851	.732	.854	
5.8	19.23	12.448	.703	.859	

Overall reliability					.881
High-level monitoring and revising phase					
High-level editing					
4.8	35.47	33.608	.540	.913	.913
4.9	35.61	30.524	.743	.901	
4.10	35.64	30.776	.714	.903	
4.11	35.38	33.768	.608	.909	
4.12	35.37	33.357	.596	.910	
5.1	35.46	31.657	.764	.900	
5.2	35.53	31.152	.795	.898	
5.3	35.63	31.395	.725	.902	
5.4	35.36	32.405	.698	.904	
5.5	35.26	33.600	.656	.907	
Overall reliability					.913

*. The digit in front of the decimal point of an item number indicates which of the five stages this item is in (1=*while reading the task prompt*; 2=*while reading the source materials*; 3=*before writing*; 4=*while writing the first draft*; 5=*after writing the first draft*).

Overall, as shown in Table 6.1, each sub-scale corresponding to the five writing phases achieved a Cronbach's alpha of 0.50 or above, ranging from 0.56 to 0.91. These figures indicate a moderate to high level of internal consistency for each set of items within a given sub-scale. Among the seven categories of cognitive processes, items assigned to *low*- and *high-level editing* achieved high levels of reliability (*r*=0.88 and *r*=0.91 respectively), while items designed to measure the process of *text interpretation* (*r*=0.53) and *selecting* (*r*=0.52) reported the lowest internal reliabilities, but were still considered satisfactory with results above 0.50 (Chan, 2013). Out of the 40 individual items, eight items did not yield satisfactory item-total correlations of 0.30 or above, but half of these items' item-total correlations were above 2.70, and Items 1.1 and 2.4 reported item-total correlations that were very close to 0.30 (0.292 and 0.298 respectively). Four items (Item 4.4, Item 2.1, Item 4.7 and Item 4.1) reported item-total correlations below 2.70. The low correlations may be because the participants did not respond to these items similarly to how they responded to the other items in the same group. These four items will be investigated in detail in the next section, by looking at participants' responses to relevant items.

6.3 Results of descriptive and inferential statistics

In order to understand the extent to which participants employed each category of the cognitive processes specified in the questionnaire while completing the TBEM-8 reading-to-write task (RQ1), a frequency analysis was performed and the results are displayed in the following seven subsections (Sections 6.3.1 to 6.3.7), each of which looks at one type of cognitive processes. Section 6.3.8 presents the results of Mann-Whitney U tests to examine if there were any differences in the responses to each questionnaire item between participants who scored differently on the TBEM-8 reading-to-write task (RQ1a). It should be noted that an exploratory factor analysis was also performed but yielded no meaningful results; this may be because (1) most of the questionnaire items were highly endorsed for the "agree" and "strongly agree" responses, resulting in low correlations (lower than 0.3) between items; and (2) the number of items (N=40) was relatively large considering that of the participants (N=172), which might impact on the effect of this analysis.

6.3.1 Task representation and macro-planning

Five items were designed to measure the process of task representation and macro-planning, during which test-takers are expected to create an understanding of the task demands (task representation) and plan for writing goals and content, identifying major constraints of a task such as genre of the expected response and the level of formality required (macro-planning). Table 6.2 presents these five items and the agreement rate (adding up the percentage of participants who chose "agree" and "strongly agree") for each item. Figure 6.1 shows the percentage of participants choosing each number (1 to 5: strongly disagree, disagree, no view, agree, strongly agree) for these items.

Table 6.2: Agreement with items measuring the process of *task representation and macro-planning*

Items*	agree or strongly agree (n=170)
1.1 I read the directions carefully to understand each word in it.	80.3%
1.2 I thought of what I might need to write to make my essay relevant and adequate to the task.	77.7%
1.3 I was able to understand the directions for this writing task very well.	69.5%
2.6 I read the directions again.	91.1%
4.4 I re-read the directions.	79.3%

*. The digit in front of the decimal point of an item number indicates which of the five stages this item is in (1=*while reading the task prompt*; 2=*while reading the source materials*; 3=*before writing*; 4=*while writing the first draft*; 5=*after writing the first draft*).

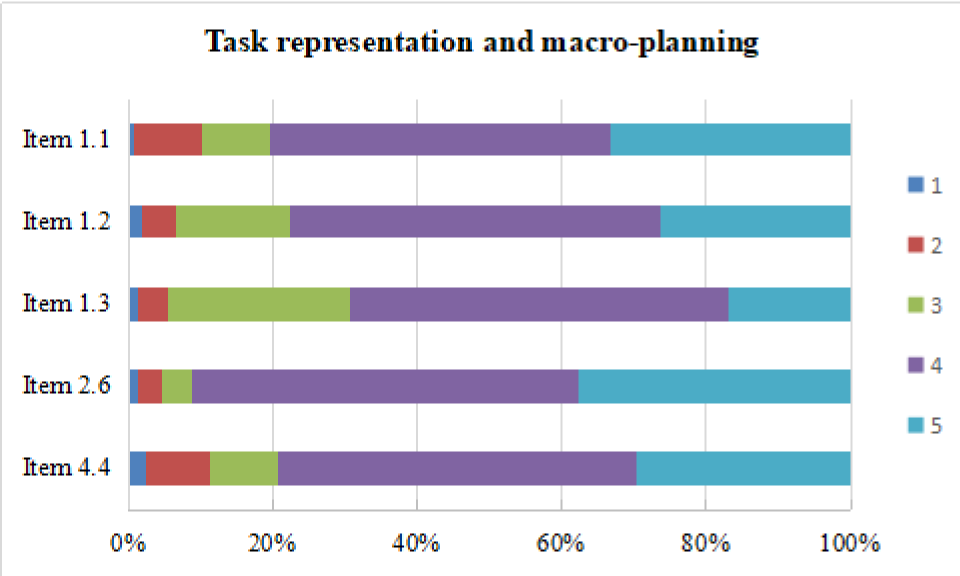

Figure 6.1: Stacked bar chart illustrating the proportion of participants' responses to each item in *task representation and macro-planning* (5=strongly agree; 4=agree; 3=no view; 2=disagree; 1=strongly disagree)

With 80.3 percent agreement rate for Item 1.1, it seems that the majority of participants did read the directions carefully to comprehend the demands of the TBEM-8 reading-to-write task. However, the percentage of participants who claimed that they understood the directions very well (Item 1.3) fell to 69.5 percent, and it is interesting to note that more than 20 percent of participants chose "no view" on this item. This may be because of the complexity and difficulty of the reading-to-write task that participants may not have been able to create a complete representation of it after reading the instructions for the first several times, and they may not have spent more effort re-reading the instructions to gain a clearer understanding of the task as they went on completing the task. This seems to be particularly true as evidenced by the findings in Item 2.6 that 91.1 percent of participants read the directions again while reading the source materials, and in Item 4.4 that about 80 percent of students reported that they re-read the instructions while writing the first draft. For Item 1.2, the only item designed to measure the process of macro-planning, a total of 77.7 percent of participants considered how to write the essay in accordance with task demands in the instructions to make it relevant and adequate to the task.

It should also be noted that the distribution of participants' responses to Items 1.1 and 4.4 are almost identical, with a minor difference in the proportion of participants who chose "strongly disagree", while the other three items in this group had an obviously lower percentage of participants who disagreed that they engaged in certain task representation and macro-planning processes. This may help to explain why Items 1.1 and 4.4 did not yield satisfactory item-total correlations of 0.30 within this group.

6.3.2 Text interpretation

Four items were meant to measure the participants' text interpretation process, which, in this questionnaire, referred specifically to participants' activities relating to reading the source materials. Table 6.3 presents these four items and the agreement rate for each item. Figure 6.2 shows the percentage of participants choosing each number for these items.

Table 6.3: Agreement with items measuring the process of *text interpretation*

Items	agree or strongly agree (n=170)
2.1 I read through the whole of each source material carefully	82.2%
2.2 I searched quickly for the ideas which might help me to write the essay.	89.3%
2.3 I read some relevant part(s) of the materials carefully.	93.5%
4.5 I selectively re-read the source materials.	87.1%

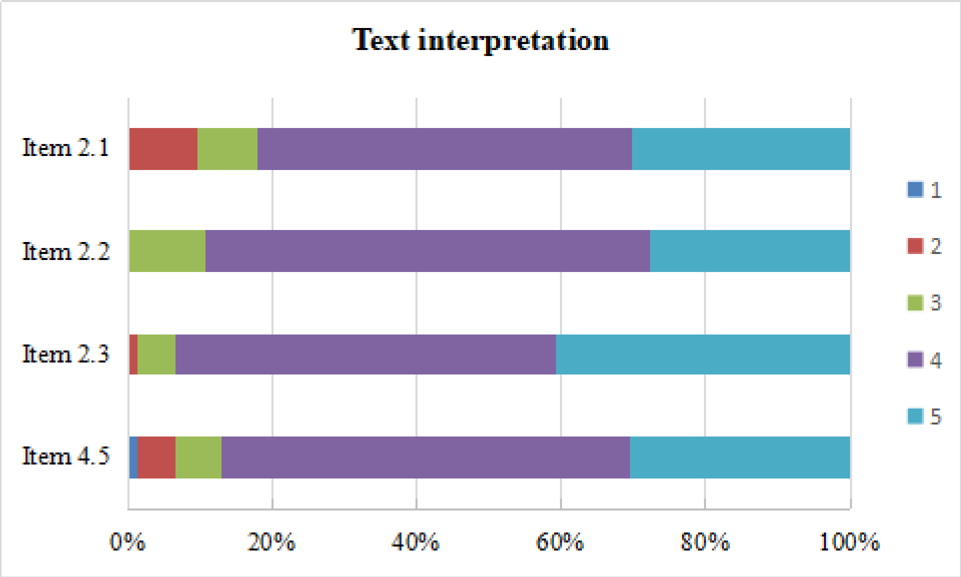

Figure 6.2: Stacked bar chart illustrating the proportion of participants' responses to each item in *text interpretation* (5=strongly agree; 4=agree; 3=no view; 2=disagree; 1=strongly disagree)

Overall, as shown in Table 6.3, more than 80 percent of participants chose either "agree" or "strongly agree" in response to the four items. Items 2.1 and 2.3 investigated participants' careful reading approaches. For Item 2.3, a very large proportion of participants (93.5 percent) claimed that they read some relevant part(s) of the materials carefully, while Item 2.1 had a lower agreement rate (82.2 percent) among participants, and about 10 percent of participants did not agree that they read through the whole of each source material, which was distinct from the other items in this group (this may be the reason that Item 2.1 achieved a low item-total correlation).

Items 2.2 and 4.5 elicited participants' responses to activities relating to expeditious reading. For Item 2.2, 89.3 percent of participants agreed that they searched quickly for the ideas which might be helpful in writing the essay, with 10.7 percent of participants choosing "no view" on this item. Items 2.1 and 2.2 seem to be opposites in some way, although it may be due to the fact that participants are thinking more along the lines of a sequence of behaviour (e.g., they read carefully first, and then expeditiously later). But the questionnaire does not necessarily help to differentiate between the different types of behaviour that were able to be identified in the eye-tracking study. A similar high percentage of participants (87.1 percent) claimed that they selectively re-read the source materials while writing the first draft (Item 4.5), but there were also about three percent of participants who thought they did not engage in this type of reading during writing.

6.3.3 Selecting

Five items were designed to measure the process of selecting. It is through this process that participants select information and ideas from source materials either for connecting purposes or to support their translation (transcribing abstract ideas into linguistic forms) process. Results of frequency analysis are presented in Table 6.4 and Figure 6.3.

Table 6.4: Agreement with items measuring the process of *selecting*

Items	**agree or strongly agree (n=170)**
2.4 I used the materials to help me get ideas on the topic	80.0%
2.5 The materials helped me choose an opinion on the issue.	75.1%
4.2 I used some of the ideas from the source materials in my essay.	95.3%
4.6 I paraphrased part(s) of the source materials in my writing.	86.5%
4.7 I copied phrases and sentences directly from the source materials into my essay.	76.9%

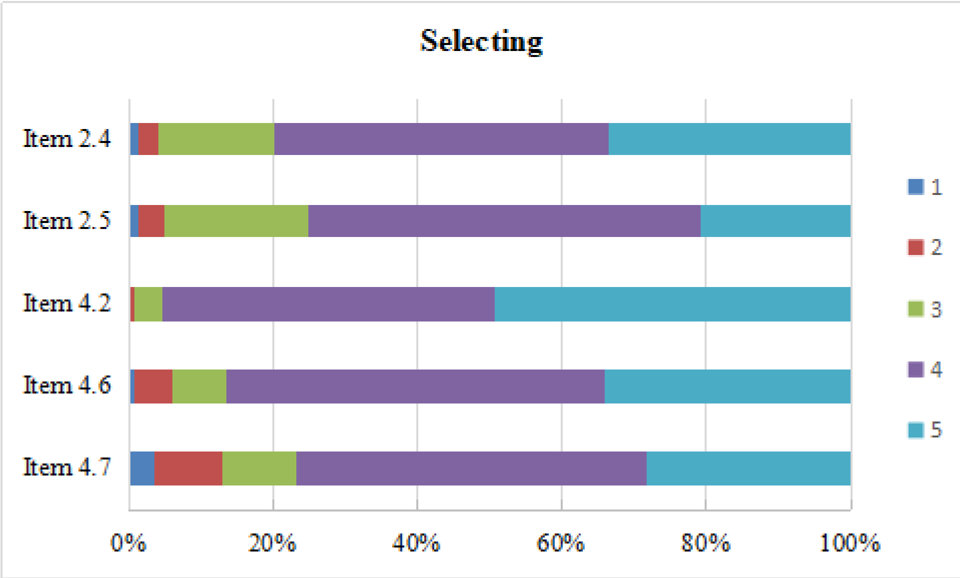

Figure 6.3: Stacked bar chart illustrating the proportion of participants' responses to each item in *selecting* (5=strongly agree; 4=agree; 3=no view; 2=disagree; 1=strongly disagree)

At the stage of reading the source materials, 80 percent of participants reported that they used the materials to help them get ideas on the topic (Item 2.4), and a slightly lower proportion (75.1 percent) of participants agreed that the materials were helpful in choosing an opinion on the issue they were going to write about (Item 2.5). However, it should be noted that a relatively high amount of participants chose "no view" on these two items, although few participants disagreed that they engaged in the selecting process while reading source materials before writing.

A markedly higher level of agreement was found in Item 4.2, with a total of 95.2 percent of participants using some of the ideas from the source materials in their essays while writing the first draft. For Item 4.6, a large proportion of participants (86.5 percent) said that they did paraphrase part(s) of the source materials in their essays. This is in line with the findings in the stimulated recall analysis that 14 of the 16 (87.5

percent) participants reported that they engaged in this type of selecting during writing. A slightly lower proportion (76.9 percent) of participants claimed that they copied phrases and sentences directly from the source materials (Item 4.7); more than 10 percent of participants said they did not copy from the materials. This does not conform to the data on copying in the stimulated recalls that only one participant reported copying directly from source materials (see Section 5.2.6 for details), although it suggests that the notion of direct copying of material may in fact be avoided by a proportion of participants. This item had the highest level of disagreement among the responses to the five items in this group, which may be the reason that it did not achieve a satisfactory item-total correlation of 0.30.

6.3.4 Connecting and generating

Three items were meant to measure the process of connecting and generating, during which participants were expected to connect ideas they selected from the source materials with their own knowledge, and generate links between ideas or new meaning in their essays. Table 6.5 and Figure 6.4 present the results of frequency analysis.

Table 6.5: Agreement with items measuring the process of *connecting and generating*

Items	**agree or strongly agree** (n=170)
2.8 I linked the important ideas in the source materials to what I know already	72.6%
2.10 I developed new ideas or a better understanding of existing knowledge.	59.4%
4.3 I developed new ideas while I was writing.	61.5%

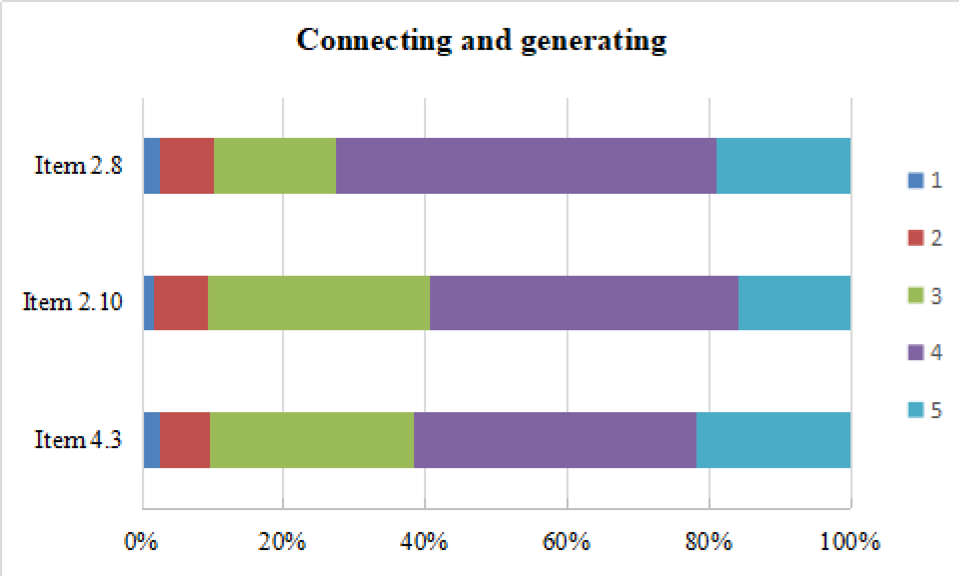

Figure 6.4: Stacked bar chart illustrating the proportion of participants' responses to each item in *connecting and generating* (5=strongly agree; 4=agree; 3=no view; 2=disagree; 1=strongly disagree)

Overall, lower levels of agreement were found in these three items (ranging from 59.4 to 72.6) compared with other items that have been presented thus far; the disagreement rate for these three items were almost identical (about 10 percent). For Item 2.8, 72.6 percent of participants reported that they linked ideas in the source materials with their own knowledge while reading these materials. However, the percentage of participants who claimed that they developed new ideas or a better understanding of existing knowledge (Item 2.10) fell to 59.4 percent. It is interesting to note that more than 15 percent of participants chose "no view" on this item, which had the second highest percentage among different responses to this item. For Item 4.3, the percentage of participants who developed new ideas while writing the first draft slightly increased to 61.5 percent; again, a relatively high proportion of participants (about 15 percent) said they had no view towards this item. These frequencies suggest that a sizable minority of participants relied a lot on the source texts for ideas to be put in their written products, rather than generating new ideas as they might have to in a more traditional independent writing task.

6.3.5 Organising

Seven items (see Table 6.6) were designed to measure the process of organising, during which participants were expected to think about the structure of the source materials, and organise or prioritise the ideas to be put in the text. Results of frequency analysis are presented in Table 6.6 and Figure 6.5.

Table 6.6: Agreement with items measuring the process of *organising*

Items	agree or strongly agree (n=170)
2.7 I prioritised the important ideas in the source materials in my mind.	27.1%
2.9 I worked out how the main ideas across the source materials relate to each other.	68.8%
2.11 I used the materials to help me organise my essay.	68.3%
3.1 I organised the ideas I planned to include in my essay.	68.4%
3.2 I recombined or reordered the ideas to fit the structure of my essay.	67.8%
3.3 I removed some ideas I planned to write.	45.7%
4.1 While I was writing, I sometimes paused to organise my ideas.	81.8%

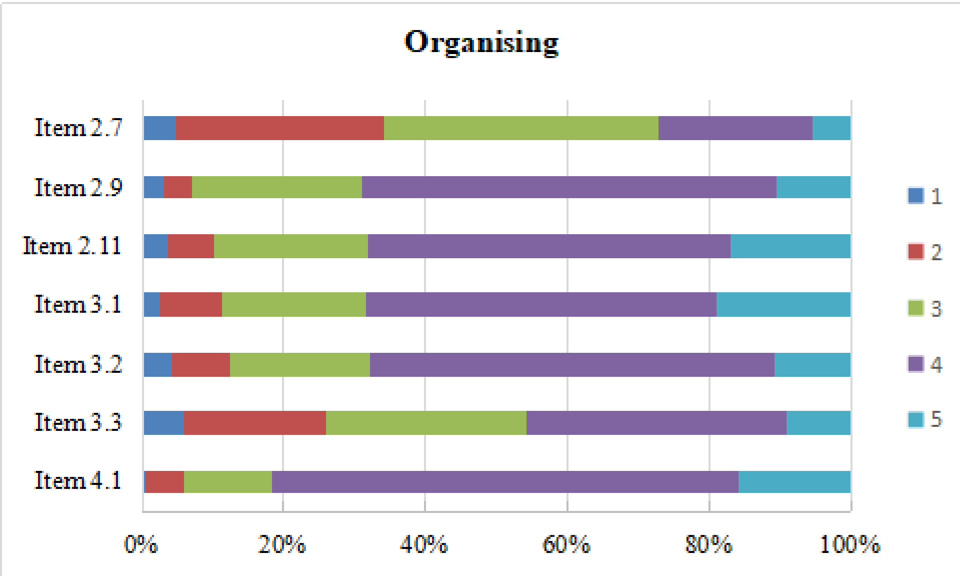

Figure 6.5: Stacked bar chart illustrating the proportion of participants' responses to each item in *organising* (5=strongly agree; 4=agree; 3=no view; 2=disagree; 1=strongly disagree)

Items 2.7, 2.9 and 2.11 look at participants' organising process while reading the source materials. A low level of agreement was found in answers responding to Item 2.7 with only 27.1 percent of the participants reporting that they prioritised the important ideas in the source materials in their mind. This compares with the 34.1 percent of participants who claimed not to engage in this organising process. There was also a strikingly high percentage of participants (38.8 percent) who chose "no view" on this item. Unlike Item 2.7, Items 2.9 and 2.11 had a similar distribution of different responses, with 68.8 percent of participants reporting that they thought about how the main ideas across the source materials relate to each other, and 68.3 percent of participants claiming that they used the materials to help them organise the essays.

Items 3.1 to 3.3 were meant to investigate other types of organising processes used by participants before they started to compose. The proportions of responses to Items 3.1 and 3.2 were almost identical; 68.4 percent of participants said that they organised the ideas they planned to include in the essays (Item 3.1), while 67.8 percent of participants did re-combine or re-order the ideas generated to fit the structure of their essays (Item 3.2). Unlike the above two items, Item 3.3 saw a lower level of agreement; 45.7 percent of participants reported that they removed some ideas they planned to write before starting to compose; also, a relatively high percentage (28.3 percent) of participants chose "no view" on this item.

For Item 4.1, unlike the other six items in this group, a large proportion of participants (81.8 percent) thought that they sometimes paused during writing to organise their ideas (the difference in the proportion of agreement may be the reason that Item 4.1 had a low item-total correlation). This agrees with the findings in the stimulated recall analysis that the participants were more actively engaged in organising processes while they were producing text.

6.3.6 Low-level editing

Six items (see Table 6.7) were designed to measure the process of low-level editing. It is through this process that participants check the accuracy of spelling, punctuation and syntax, etc. Results of the frequency analysis are presented in Table 6.7 and Figure 6.6.

Table 6.7: Agreement with items measuring the process of *low-level editing*

Items	**agree or strongly agree (n=170)**
4.13 I checked that I had put the ideas of the source materials into my own words.	81.7%
4.14 I checked the grammatical accuracy and range of the sentence structures.	59.6%
4.15 I checked the spelling, usage and range of the vocabulary.	69.4%
5.6 I checked that I had put the ideas of the source materials into my own words.	81.9%
5.7 I checked the grammatical accuracy and range of the sentence structures.	71.5%
5.8 I checked the spelling, usage and range of the vocabulary.	75.0%

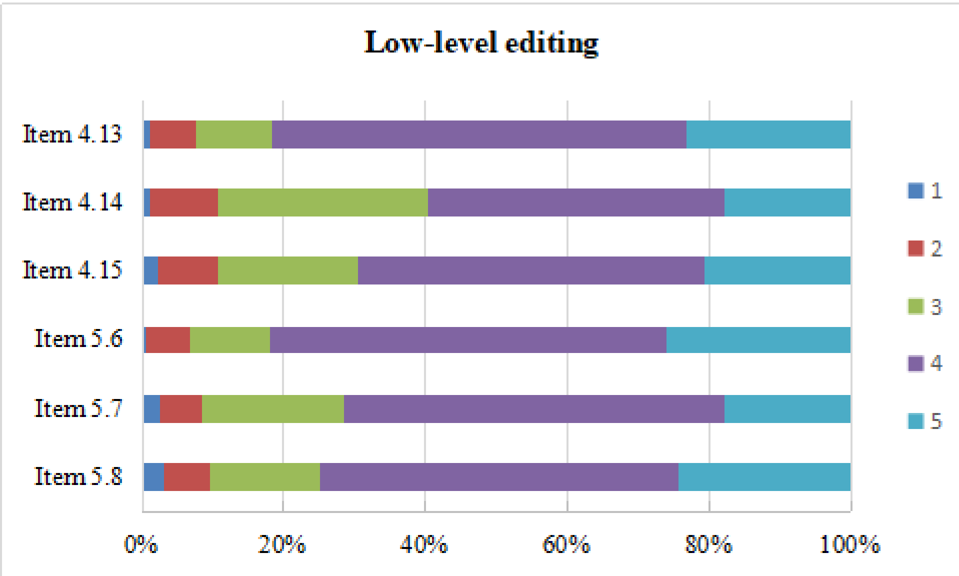

Figure 6.6: Stacked bar chart illustrating the proportion of participants' responses to each item in *low-level editing* (5=strongly agree; 4=agree; 3=no view; 2=disagree; 1=strongly disagree)

Items 4.13, 4.14 and 4.15 related to the participants' low-level editing processes while writing the first draft. More than 80 percent of participants reported that they had put the ideas drawn from the source materials into their own words (Item 4.13), indicating that most of them had a good understanding of the task and integrated information from the materials into their own writing. For Item 4.14, the percentage of agreement fell to 59.6, and about 30 percent of participants claimed that they had no idea whether they checked the grammatical accuracy and range of the sentence structures during writing. This suggests that the participants did not necessarily think about the grammatical range and accuracy of their texts as they were writing. Perhaps they were more focused on getting ideas into the text and writing something coherent. A slightly higher level of agreement was found in Item 4.15 where 69.4 percent of participants reported that they checked the spelling, usage and range of the vocabulary whilst writing.

Items 5.6, 5.7 and 5.8 investigate the participants' low-level editing process after finishing the first draft. For Item 5.6, the figure for agreement rate remains stable, with 81.9 percent of participants agreeing that they checked if they had put ideas from the source materials into their own words. The proportion of agreement in Items 5.7 and 5.8 increased to 71.5 and 75.0 percent, indicating that the participants engaged more often in low-level editing process after they had finished the first draft than they did whilst writing. Again, it is interesting to note that the participants seemed to be a bit more focused on using the source material and integrating it well (see Tables 6.4 and 6.5) than they were with achieving formal accuracy in their writing. This suggests that they may prioritise the use of selecting and connecting processes (particularly during writing) in completing an integrated writing task.

6.3.7 High-level editing

Ten items were designed to measure the process of high-level editing. At this level of editing, the concern is mainly with the extent to which the text produced so far fits in with participants' writing goals established in the previous stages, its relevance to the task set and the development of the structure of the text. Results of the frequency analysis are presented in Table 6.8 and Figure 6.7.

Table 6.8: Agreement with items measuring the process of *high-level editing*

Items	**agree or strongly agree (n=170)**
4.8 I checked that the content was relevant.	76.2%
4.9 I checked that the essay was well-organised.	70.6%
4.10 I checked that the essay was coherent, e.g., appropriate use of topic sentences, connectives, etc.	67.5%
4.11 I checked that I included all appropriate main ideas from all the source materials.	81.8%
4.12 I checked that I included my own viewpoint on the topic.	83.4%
5.1 I checked that the content was relevant.	80.9%
5.2 I checked that the essay was well-organised.	72.0%
5.3 I checked that the essay was coherent, e.g., appropriate use of topic sentences, connectives, etc.	70.7%
5.4 I checked that I included all appropriate main ideas from all the source materials.	83.9%
5.5 I checked that I included my own viewpoint on the topic.	91.0%

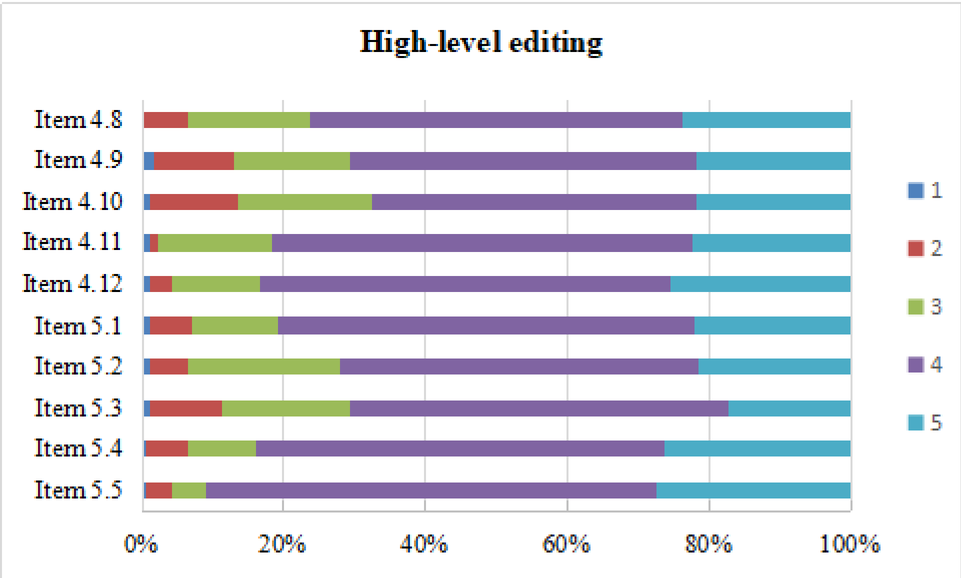

Figure 6.7: Stacked bar chart illustrating the proportion of participants' responses to each item in *high-level editing* (5=strongly agree; 4=agree; 3=no view; 2=disagree; 1=strongly disagree)

Items 4.8 to 4.12 look at the participants' high-level editing process while writing the first draft. Overall, it seems that editing at both levels exhibits a similar extent of agreement (the levels of agreement are all roughly around 60 to 80 percent). The highest levels of agreement were found in Items 4.11 and 4.12, with 81.8 percent of participants claiming that they did check whether they included all appropriate main ideas from all the source materials (Item 4.11) and 83.4 percent of participants reporting that they checked if they included their own viewpoint on the topic (Item 4.12). Some slightly lower levels of agreement were found in Items 4.8 to 4.10. 76.2 percent of participants checked if the content of their essays was relevant (Item 4.8) and 70.6 percent checked if the essays were well-organised (Item 4.9). For Item 4.10, 67.5 percent of participants checked the coherence of the essays. It looks like task achievement (e.g., relevance of content, inclusion of viewpoint) was more strongly endorsed than more linguistic foci (e.g. checking topic sentences, connectives)

Items 5.1 to 5.5 investigate the participants' high-level editing process after finishing the first draft. The proportion of agreement in these items all increased compared with items 4.8 to 4.12. Item 5.5 had the largest increase that 91.0 percent of participants reported that they had checked if they included their own opinions on the topic. It may be safely concluded that a relatively large number of participants were well aware of the importance of editing at both the basic and advanced levels, and the participants seemed to engage more in the high-level editing process than the low-level editing process, and edit more often after finishing the first draft than whilst writing.

6.3.8 Results of Mann-Whitney U tests

In order to investigate if higher- and lower-scoring participants responded to each item differently (RQ1a), a set of Mann-Whitney U tests was performed as Shapiro-Wilk tests of normality indicated that some sets of responses to the questionnaire items were non-normally distributed (p>.05). Items to which the responses were significantly different between the two groups of participants are displayed in Table 6.9 below (see Appendix J for the results of all items).

Table 6.9: Significant differences in responses to items between the higher- and lower-scoring participants

Items	Mann-Whitney U	Wilcoxon W	Z	*p*	Effect size
1.2 I thought of what I might need to write to make my essay relevant and adequate to the task.	2992.500	6647.500	-2.104	.035	.026
2.1 I read through the whole of each source material carefully.	2947.000	6433.000	-2.019	.043	.024
2.9 I worked out how the main ideas across the source materials relate to each other.	2925.500	6580.500	-2.413	.016	.034
3.1 I organised the ideas I planned to include in my essay.	2927.500	6413.500	-2.048	.041	.025
3.3 I removed some ideas I planned to write.	2844.500	6414.500	-2.020	.043	.024

As shown in Table 6.9, the higher- and lower-scoring participants responded to six items significantly differently, but the effect sizes for these items are trivial (0.024 to 0.034). The percentage of participants from each group choosing each number (1 to 5: strongly disagree, disagree, no view, agree, strongly agree) for these six items are presented in Figures 6.8 to 6.13 and looked at individually.

For Item 1.2 (see Figure 6.8), which reads as "*I thought of what I might need to write to make my essay relevant and adequate to the task*", it seems that the proportion of participants choosing "agree" were similar in higher- and lower-scoring groups, but participants who scored higher on the TBEM-8 reading-to-write task chose "strongly agree" more than their counterparts, indicating that the higher-scoring participants may be more actively engaged in the macro-planning process while reading the task instructions before writing.

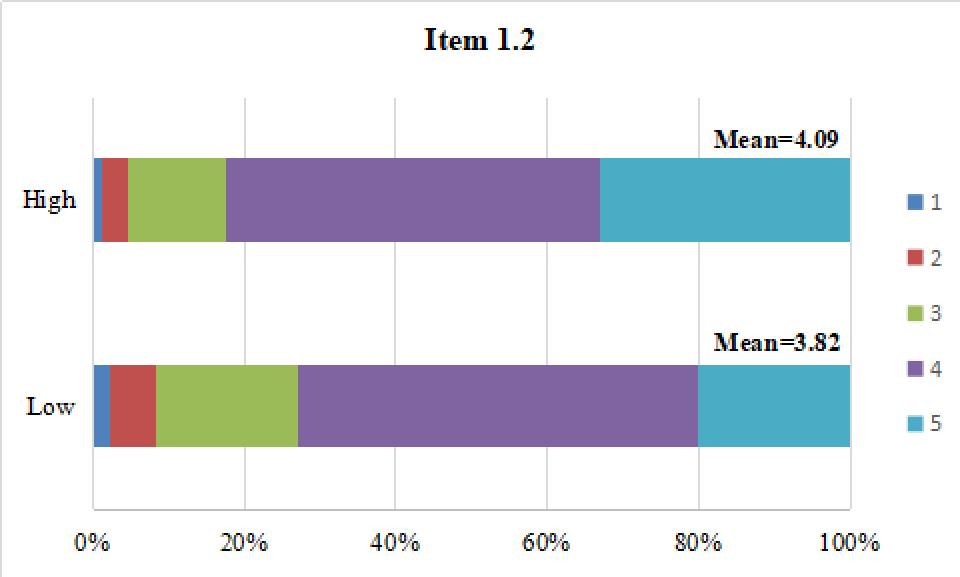

Figure 6.8: Stacked bar chart illustrating the proportion of responses to Item 1.2 for higher- and low-scoring participants (5=strongly agree; 4=agree; 3=no view; 2=disagree; 1=strongly disagree)

For Item 2.1 (see Figure 6.9), which reads as "*I read through the whole of each*

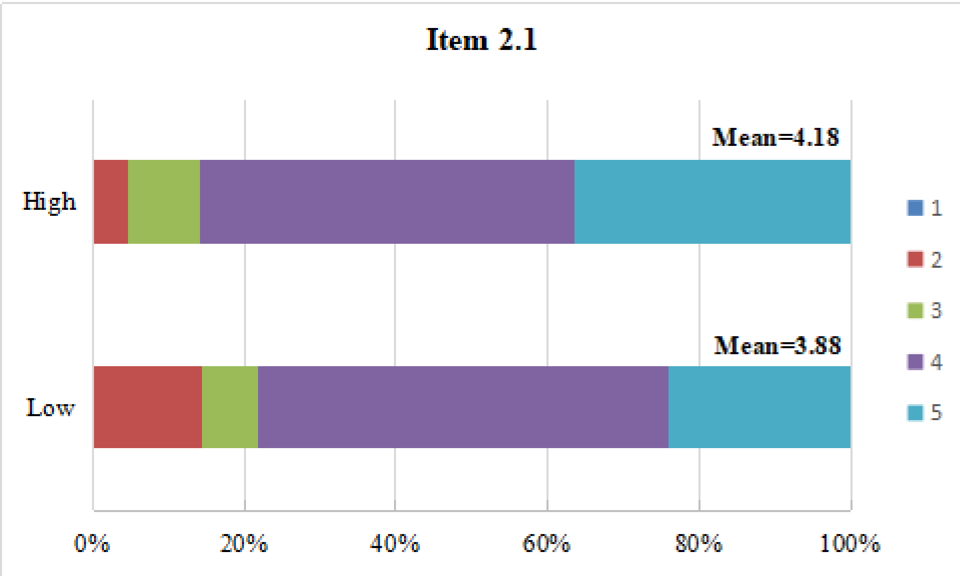

Figure 6.9: Stacked bar chart illustrating the proportion of responses to Item 2.1 for higher- and low-scoring participants (5=strongly agree; 4=agree; 3=no view; 2=disagree; 1=strongly disagree)

source material carefully”, the higher-scoring participants chose “strongly agree” more than their counterparts, and the level of disagreement in lower-scoring group was higher than that in the better-performing group. This indicates that the participants who scored higher in the TBEM-8 reading-to-write task may adopt more often a careful reading approach while reading the source materials and spend more time reading through the whole of each source material carefully.

For Item 2.9 (see Figure 6.10), which reads as “*I worked out how the main ideas across the source materials relate to each other*”, the better-performing group of participants had a higher level of agreement than their counterparts, and it is interesting to note that more than 30 percent of participants in the lower-scoring group said that they had no view on this item. This may suggest that the participants who scored higher in the reading-to-write task were more aware of the importance of creating a complete representation of the source materials, and were willing to spend more time thinking about the structure of the materials in order to help them organise their own essays.

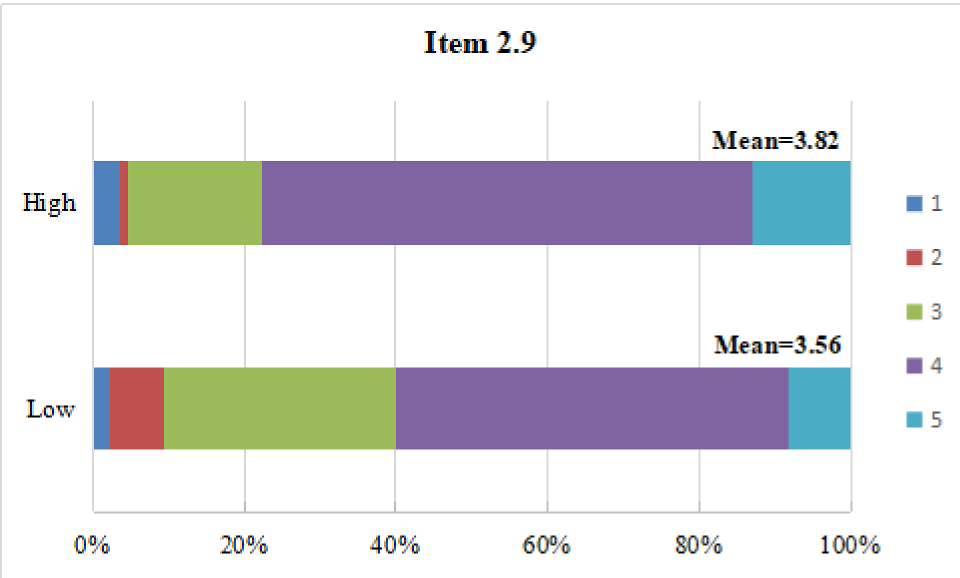

Figure 6.10: Stacked bar chart illustrating the proportion of responses to Item 2.9 for higher- and low-scoring participants (5=strongly agree; 4=agree; 3=no view; 2=disagree; 1=strongly disagree)

For Item 3.1 (see Figure 6.11), which reads as "*I organised the ideas I planned to include in my essay*", it seems that a larger proportion of participants in the higher-scoring group agreed that they engaged in this organising process before starting to write. Again, like Item 2.9, a relatively large proportion (about 30 percent) of lower-scoring participants chose "no view" on this item.

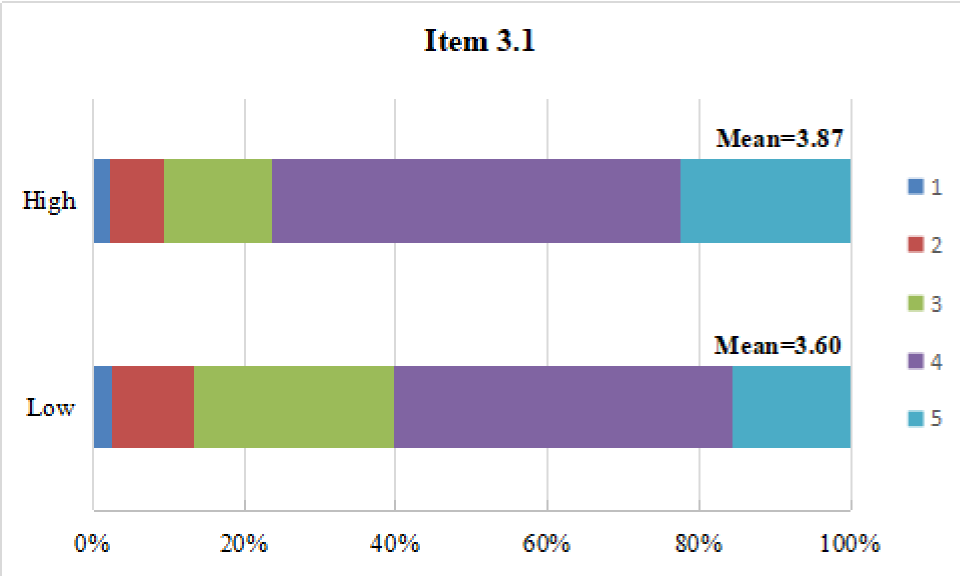

Figure 6.11: Stacked bar chart illustrating the proportion of responses to Item 3.1 for higher- and low-scoring participants (5=strongly agree; 4=agree; 3=no view; 2=disagree; 1=strongly disagree)

For Item 3.3 (see Figure 6.12), which reads as "*I removed some ideas I planned to write*", a higher level of agreement was found in the group of lower-scoring participants, indicating that they were more likely to remove some ideas they planned to write during the organising process before writing. This may be because the better-performing group of participants may be more efficient in organising the ideas to be put in the text and so their plans for the overall structure of the essays may be clearer and more complete.

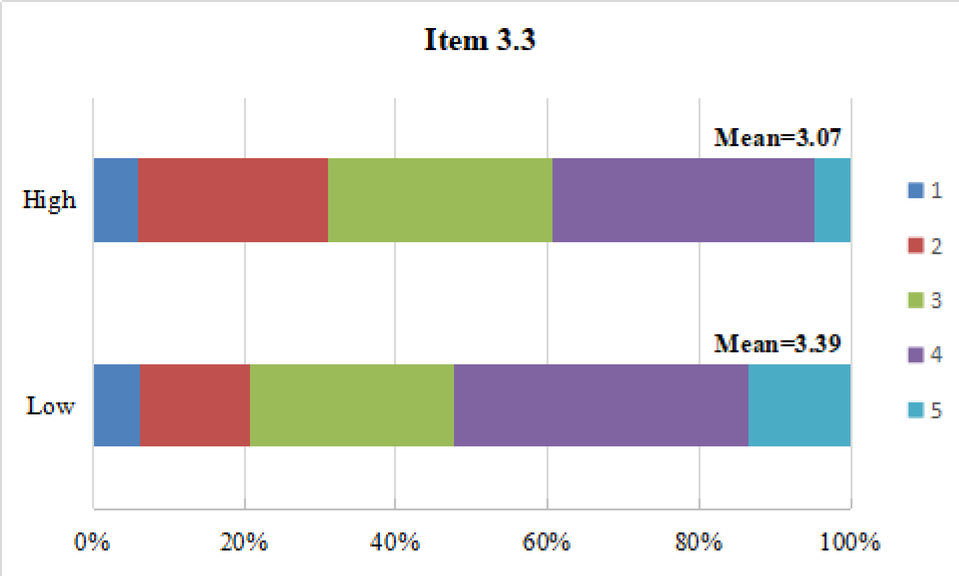

Figure 6.12: Stacked bar chart illustrating the proportion of responses to Item 3.3 for higher- and low-scoring participants (5=strongly agree; 4=agree; 3=no view; 2=disagree; 1=strongly disagree)

6.4 Summary

This chapter has presented results from analysis of the reading-to-write questionnaire data. It adds to the results presented in the eye-tracking and stimulated recall study (Study I). First, a reliability analysis showed that all the five hypothesised writing phases yielded a Cronbach's alpha of 0.50 or above, suggesting a moderate to high level of internal consistency for each writing phase. Section 6.3 then presents the results of a frequency analysis of the participants' responses to the 40 items in the questionnaire. There was a high level of agreement found in these items, with only four items achieving an agreement rate below 60 percent, and more than 70 percent of participants choosing either "agree" or "strongly agree" in 28 items. Thus it can be safely concluded that the participants underwent a variety of essential cognitive processes while completing the TBEM-8 reading-to-write task. A set of Mann-Whitney U tests was also performed to investigate if there were any differences in responses to the 40 items between the higher- and lower-scoring participants, and the results show that participants who performed better in the TBEM-8 reading-to-write task engaged more often in some types of macro-planning and organising process before starting to write, which to some extent may have impact on their reading-to-write performance.

In the next chapter, the findings from Study I and Study II will be summarised and triangulated to answer the research questions posed in the literature review, and how these results relate to the published literature will also be discussed.

CHAPTER 7 DISCUSSION

7.1 Introduction

This chapter focuses on three topics. First, Section 7.2 triangulates the findings from all different sources of data to answer the first set of research questions, and discusses the cognitive processes employed while completing the TBEM-8 reading-to-write task and how they fit with previously published knowledge. Second, in Section 7.3, the findings are triangulated to answer the second set of research questions, and how the participants engaged with the source materials in the task is discussed. Third, Section 7.4 proposes a model for reading-to-write process that emerges from the findings of this study. Finally, a summary of this chapter is provided in Section 7.5.

7.2 Cognitive processes involved in completing the TBEM-8 reading-to-write task (RQ1 & RQ1a)

Findings from all different sources of data are triangulated and discussed in the two subsections below to address the first set of research questions.

7.2.1 RQ1

The first overarching research question (RQ1) was:

> *What cognitive processes do test-takers employ while completing the TBEM-8 reading-to-write task?*

Overall, findings from the analysis of stimulated recall data indicate that the majority of participants (*n*=16) employed a wide range of cognitive processes (1,956 instances) specified in Shaw and Weir's (2007) model of writing and Spivey's (1997) discourse synthesis model, with varying frequencies. Text interpretation, selecting and micro-planning were the three most frequently reported processes according to participants' recalls, and macro-planning and translating were the two least reported processes. In addition, findings from an analysis of the reading-to-write process questionnaire data were, to a large extent, consistent with the stimulated recall results. A high level of agreement was found in participants' responses to the 40 items in the questionnaire, with more than 70 percent of participants choosing either "agree" or "strongly agree" in 28 items, and only four items achieving an agreement rate below 60 percent.

I. Text-interpretation

It appears almost self-evident that text-interpretation is an essential process in completing an integrated reading-to-write task, because not only the text in the task instructions and the text that has been produced are going to be accessed by writers, but the text in the source materials also needs to be interpreted.

A total of 668 instances (34.2 percent of all instances) of text-interpretation processes were reported by the 16 participants in the eye-tracking and stimulated recall study; meanwhile, in the questionnaire study, a large proportion (more than 80 percent) of participants claimed that they adopted both careful and expeditious reading approaches when reading the source materials. Previous studies (e.g., Chan, 2013) showed similar results for writers' reading activities in integrated writing tasks. This study goes on to further explore how these activities differ at various stages of writing, benefiting from the eye-tracking technique which allows an online investigation of participants' eye-movements during task completion.

Before writing, the participants typically started responding to the TBEM-8 reading-to-write task by quickly browsing all the components of this task, and then went back to read the task instructions and source materials one after another in a slow and careful manner (see Sections 4.4.1 and 5.2.1 for details). During writing, the participants most often used an expeditious form of reading: scanning, to locate specific information in either the instructions or source materials that they considered useful in their writing. Another interesting finding was that the participants tended to read Chinese source materials much faster than the English texts, as Participant 4 reported, "*the first paragraph, because it is in Chinese, so I read it very fast, because the second paragraph is in English, so I read it relatively slowly*". This suggests that the language of the text appeared to influence the degree and nature of writers' interaction with the source texts, which has not been studied in previous research. After writing, the participants reported relatively less instances of text-interpretation process, most of which were devoted to reading the text that has been produced for monitoring and revising purposes.

II. Task representation

Findings from all sources of data have proved that the majority of participants read the task instructions carefully to create an initial understanding of the task before writing, which aligns with previous research on this process (Allen, 2004; Chan, 2013; Flower et al., 1990). However, it was also found in this study that the task representation process occurred both during and after writing. Participants regularly revisited the instructions for either guidance on the text to be produced or to check the progress and/or quality of their writing (see Section 5.2.2 for details). This indicates that the task representation process is not a single, simple act, but an extended, repetitive interpretive process that is employed throughout task completion.

Another finding was that the use of source materials in an integrated writing task may, to some extent, complicate test-takers' task representation process. This agrees with findings in some other studies (Plakans, 2010; Wolfersberger, 2007). Participants' eye-tracking traces demonstrated that they tended to read back and forth between the instructions and source materials in the first five minutes of task completion (see Section 5.2.1 for details). This may be because the extra cognitive load involved in reading the source materials taxed test-takers' working memory to the extent that they may not have been capable of holding the content of task instructions in mind. As Participant 14 reported, "*After reading (the source materials), I forgot what I was required to do, and then I went back there (the instructions) and checked it out again*". These findings suggest that working memory may play an important role in coordinating integrated writing processes (Plakans, 2010; Purpura, 2014).

III. Macro-planning

Twelve of the sixteen participants in Study I reported using macro-planning processes (27 instances) while completing the TBEM-8 reading-to-write task. This percentage is similar to that in the questionnaire analysis where a total of 77.7 percent of participants reported that they considered how to write the essay to make it relevant and adequate to the task.

Almost all of the 27 instances of macro-planning processes occurred before writing and they were mainly concerned with goal setting and content consideration. No instances of consideration of the genre and target readership, which were considered as important aspects of macro-planning according to Field (2005) and Shaw and Weir (2007), were found in participants' recalls. This may indicate that writers prioritise considering task achievement when macro-planning; or, it may suggest that the participants were not aware of the importance of considering the genre and intended readers (relevant information has been pointed out in the task instructions) in successful writing.

It should also be noted that the macro-planning process in completing integrated writing tasks may be more complicated than in a traditional independent writing task, because writers need to comprehend the source materials provided and build connections between the texts as part of their macro-planning for successful task completion. A notable finding in the questionnaire study was that participants who scored higher on the TBEM-8 reading-to-write task tended to be more actively engaged in macro-planning than the lower-scoring participants (see Section 6.3.8 for details). This suggests that macro-planning may be an important component of a participant's reading-to-write proficiency.

IV. Organising

Fourteen participants in the stimulated recall study reported that they considered either the structure of the source materials (42 instances) or the structure of their own writing (18 instances) during task completion, which aligns with previous studies on the organising process in integrated reading-to-write tasks (Plakans, 2009b; Stein, 1990). However, it is interesting to note that, although the participants did appear to spend time pondering the structure of source materials *before* writing, few had worked out a plan of ideas to be included in the text. Instead, they seemed more prone to using strategies to understand source materials *during* writing, while concurrently structuring their essays (see Section 5.2.4 for details). This may be due to the complexity and difficulty of the organising process in integrated writing tasks which involves comprehension of source materials. Put simply, test-takers may not be able to create a complete representation of the sources instantly, which impedes the progress of building their own text on the basis of these sources.

Another noteworthy finding of the questionnaire analysis is that the better-performing group of participants had a higher level of agreement on items related to the organising process than their lower-scoring counterparts. This may suggest that participants who performed better in the TBEM-8 reading-to-write task were more aware of the importance of creating a complete representation of the sources provided, and were more willing to spend time thinking about the structure of the input texts in order to help to organise their own essays.

V. Connecting and generating

As a central process in completing reading-to-write tasks, 91 instances of connecting and generating processes were reported by participants in the stimulated recall study. It should be noted that the majority of these instances occurred during writing, although previous studies showed that this process was very likely to happen when writers were reading source texts and brainstorming before they started to write (Chan, 2013; Plakans, 2009b; Stein, 1990). This may be because of the limitation of the stimulated recall methodology that participants may not be able to recollect adequately, due to memory decay, what they had done at earlier stages of task completion. On the other hand, however, it may be due to the difficulty of building a complete representation of the source materials within a limited amount of time before writing, and thus participants tended to go on writing with a rough understanding of the sources, but constantly referring back to them (evidenced by the eye-tracking data) during writing to strengthen the understanding of the sources, on which participants' essays were based. This, in some way, may also suggest that connecting and generating is an ongoing process of meaning building rather than a one-off act.

Findings from the analysis of the questionnaire data provided an interesting contrast. Although more than 70 percent of participants claimed that they linked ideas from source materials with their own knowledge while reading the sources, the percentage of participants who reported that they developed new ideas during task completion fell to about 60 percent (see Section 6.3.4 for details). These frequencies suggest that a sizable minority of participants may have relied mostly on the source texts for their ideas, rather than generating new ideas as they might have to do in a more traditional writing-only task. Or, they tended not to connect ideas in the source texts with their own knowledge deeply, as Cumming et al. (2005) pointed out, so that neither the ideas presented in the sources nor their prior knowledge were likely to be reconstructed.

VI. Selecting

Findings form the analysis of both the stimulated recall and questionnaire data have demonstrated that all the participants were actively engaged in selecting processes while completing the TBEM-8 reading-to-write task, during which they (1) selected ideas from both source materials and memory for connecting; (2) selected specific information from sources to support writing; and (3) selected sentences for paraphrasing or translating (see Sections 5.2.6 and 6.3.3 for details). Most frequently, the participants were found to seek lexical and syntactical support from source texts when they were in the middle of text production. This is consistent with the eye-tracking findings which showed that the participants constantly switched their attention between different AOIs during writing. This moving from written text to source text, and across source texts, to select specific words or phrases appears to be one of the most fundamental characteristics of the reading-to-write process in this study.

One difference in the findings from different sources of data is that: in the stimulated recall study, only one participant reported that she copied (more than just a single word or phrase) directly from source materials, however, the questionnaire results showed that 76.9 percent of participants claimed that they copied phrases and sentences from the source materials. John and Mayes (1990) found that lower-proficiency writers tended to copy more directly, but no significant difference in direct copying was found between higher- and lower-scoring participants in this study, based on their self-report data.

VII. Micro-planning and translating

Micro-planning and translating processes have largely been under-researched in previous studies. The 16 participants in Study I reported 164 instances of micro-planning processes; 62 instances were devoted to micro-planning at the paragraph level, when participants planned for the content of a particular paragraph; 102 instances were found at the sentence level, when participants planned for the content and structure of an upcoming sentence. It is important to note that the participants were found often going back to read the task instructions (TI-1), source materials (TI-2) and the text-written-so-far (TI-3) when they were making micro-plans in order to generate ideas to be put in the text.

The output of the micro-planning process may be, to a certain degree, stored in writers' minds in the form of abstract ideas, which were then likely to be translated into linguistic forms in the translating process. A limited number of translating processes were reported by the 16 participants (see Section 5.2.7 for details); however, the number was very likely to be underestimated because of the automatized nature of this process and the limitations of stimulated recall methodology.

VIII. Monitoring and revising

There is ample evidence of the use of monitoring and revising processes at either a basic level or a more advanced level in participants' stimulated recalls and the questionnaire data. One notable difference between the findings from these two sources of data is that: in the eye-tracking and stimulated recall study, most of the reported monitoring processes occurred during writing; while the questionnaire data showed that the participants monitored more often *after* they finished the first draft than whilst writing. This may be due to the fact that participants (Study I) who were eye-tracked while completing the TBEM-8 reading-to-write task may have spent more time accommodating themselves to the equipment they were working with (e.g., the eye-tracker and the keyboard) than those who (Study II) took the test in a normal classroom setting, and so they (Study I) tended to spend more time on the task and had relatively less time devoted to monitoring after completing the draft.

Another interesting finding in the questionnaire study was that the participants seemed to be more focused on using the source material and integrating it well than they were with achieving formal accuracy in their writing (see Section 6.3.6 for details); in other words, task achievement (e.g., relevance of content, inclusion of viewpoint) was more strongly endorsed than more linguistic foci (e.g., checking grammatical accuracy, checking topic sentences). This suggests that the participants may prioritise the use of connecting and other relevant processes such as monitoring in completing an integrated writing task.

7.2.2 RQ1a

The sub-question of the first overarching research question (RQ1a) was:

> *Are there any relationships between the use of cognitive processes and test-takers' performance on the task?*

In **Study I**, no statistically significant correlations (at the 0.05 or 0.01 level) were found between the number of times participants used a specific cognitive process and their performance on the TBEM-8 reading-to-write task. However, there was one moderate positive correlation found if the p-value was set to 0.1 (Brunfaut and McCray, 2015), between the counts of text interpretation-2 process and the task performance; that is, the more the participants reported reading the source materials, the better they performed in the task. In addition, the correlation analysis yielded several significant correlations among the cognitive processes themselves, for example, the correlations among the processes of text interpretation-2, monitoring-1 and revising were all statistically significant at the 0.01 level. This suggests that these processes were very likely to co-occur during task completion; participants used the processes together to achieve specific goals, for example, to search for correct information (e.g., correct spelling of a certain word) in the source materials when monitoring, and/or make corresponding revisions. In Study II, the results of a set of Mann-Whitney U tests showed that the higher-scoring participants engaged more often in some types of macro-planning and organising processes before they started to write, but the effect sizes of these findings were trivial.

These correlation results may lead us to conclude that there were no relationships between the use of cognitive processes and test-takers' performance on the TBEM-8 reading-to-write task; in other words, their reading-to-write proficiency may not be indicative of their uses of specific cognitive processes. This aligns with the findings in Flower et al.'s (1990) study, in which they concluded that it is not the writers' proficiency but the contexts, including the task set, the participants' writing goals, and their prior knowledge, that affect the use of particular cognitive processes.

7.3 The extent to which test-takers engaged with source materials in the TBEM-8 reading-to-write task (RQ2, RQ2a & RQ2b)

Findings from all different sources of data (particularly the eye-tracking data) are triangulated and discussed in the three subsections below to address the second set of research questions.

7.3.1 RQ2

The second overarching research question (RQ2) was:

> *To what extent do test-takers engage with the source materials in the TBEM-8 reading-to-write task?*

In previous studies, test-takers' use of the source materials was largely investigated through examining the written products, this study, benefiting from the eye-tracking technique, is perhaps one of the few studies that looks at test-takers' real-time source use while completing an integrated reading-to-write task.

I. Heat map output

The heat map output (see Section 4.3 for details) showed that the 16 participants' attention covered all the seven parts of the TBEM-8 reading-to-write task (few fixations were found outside these areas), with various focus points within each part. Overall, the majority of participants' attention was on the answer sheet, on which they typed the response to the task. Source 4, the only graphic input, received remarkably less attention from participants than other sources. One interesting point to note is that the face of the cartoon image in Source 4 received a relatively high amount of attention compared to other areas in the picture. Several participants claimed that it was because they tended to look at places where there was no text while thinking about what to write, or the face was so striking that they could not help looking at it.

II. Time to first fixation

The participants approached the task quite differently, but one major pattern emerged from the measures of *time to first fixation*. They started responding to the task by quickly browsing each part of the task (with one exception; see Section 4.4.1 for details), and then went back to read the task instructions and source materials in a slow and careful manner, during which recursions among these materials were often found. This metric, together with participants' eye-movement recordings, demonstrated a vivid presentation of participants' looking behaviour at the start of task completion, and may, to some extent, map onto the potential cognitive processes (e.g., task representation) that were likely to happen during this period. One possible conclusion that can be drawn about participants' cognitive processing at this stage is that developing a task representation emerges as a demanding activity in an integrated reading-to-write task, involving repeated reading of the task instructions and source materials to create an understanding of the task.

III. Total visit duration

Overall, the participants spent, on average, over a quarter (26.4 percent; 580.8

seconds) of their time in reading, and 73.6 percent (1623.1 seconds) in writing. As few studies have investigated the proportion of time spent on reading and writing in a reading-to-write task, little was known about the extent of an appropriate proportion of time devoted to reading (or writing) in such a task. But it may be argued that this proportion is very likely to be determined by the specific type of reading-to-write task, that is, whether it elicits a high intensity of reading, for example, in a summary writing task, or whether it involves a moderate engagement with the source materials, such as in a writing from sources task.

In terms of time spent on each source material, Source 2, which contained several short excerpts of English texts, received the most attention from participants among the five sources. This may be because, as reported by several participants, they spent relatively more time on processing English texts than Chinese texts (with word number controlled), which suggests that the language of the text may, to some extent, influence the degree and nature of the interaction with the source texts. Source 4, the picture, received the least attention from the participants; for example, Participant 3 spent only 1.8 seconds looking at the image. Another point to note is that Source 5 (key concepts and expressions), although having considerably fewer words than any other source text, received a markedly high amount of attention (with word number controlled) from the participants. This once again supports the prominence of the selecting process (discussed above), indicating that test-takers frequently looked for either lexical support or ideas to be produced in the text while completing the task.

IV. Visit count

Although the participants spent more time writing than reading, on average they visited the task instructions and source materials more frequently (331.1 counts) than the answer sheet (229.6 counts). This is interesting as the higher number of visit counts for the reading (versus the writing) indicates that test-takers were not simply moving from the writing to one text and back to the writing. Instead, they were moving between the source texts, and then going back to the writing. This is important because it may provide evidence for test-takers' selecting and connecting processes; when they were synthesising the information they were reading/discovering, and not just filling in the writing with discrete pieces of information.

Another point worth discussing is that Source 4, the picture, was the only material on which the visit counts (M=22.3) outnumbered the total visit duration (M=18.8); Test-takers were looking at the picture quite frequently, but for very short periods of time. This might again indicate that the picture was a "safe" resting place for their eyes as they thought about the task. Or that they naturally gravitated towards the picture because it appears more interesting to them than the printed words.

V. Visit duration

Overall, the mean visit duration on each part of the TBEM-8 reading-to-write task was less than three seconds, except for on the answer sheet, where the mean visit duration was 7.5 seconds. This, in some way, indicates that the participants tended to constantly switch their attention among different parts of the task, no matter whether they were reading or writing. Most of the median visit durations on each source text were around a second, which may imply that the participants adopted a more expeditious reading approach, for example, searching for specific information they considered useful in their writing. Another interesting point to note is that the participants' median visit durations on the answer sheet were much less than the corresponding mean visit duration: ten participants' median visit durations were less than three seconds. This, again, demonstrated that test-takers frequently referred back to the source materials for various purposes (e.g., idea generation) in the process of text production.

7.3.2 RQ2a

The first sub-question of the second research question (RQ2a) was:

> *Are there any differences in eye-tracking measures among different source materials?*

Participants spent the most time on Source 5, followed by Source 2, Instructions, Source 1, and finally Source 3 (with number of words controlled; see Section 4.4.2 for details), with significant differences in *total visit duration* among each source (medium to large effect sizes), except between the Instructions and Source 2. The high amount of attention spent on Source 5, again, indicates the importance of this material in the participants' reading-to-write process; as discussed earlier, they went there for either lexical support or idea generation. For Source 2, it may be argued that it was the language of this text that led participants to spend more time processing the information in it, but this conclusion must be taken with caution because it might also be due to the fact that participants extracted more useful information from Source 2 and thus spending more time on it. The relatively high amount of time spent on the Instructions may, to some extent, again support the complexity of the task representation process in completing an integrated writing task.

In terms of the mean visit count, Source 2 was visited most frequently, followed by Source 1, Source 5, Source 3 and finally Instructions, but the Mann-Whitney tests confirmed that no significant differences were found between the number of visits in

Source 1, Source 2, and Source 5, nor between Source 3 and Instructions, though with the latter two sources having significantly fewer visits. One important point to note is that although the participants spent relatively more time on the AOI of Instructions, they had fewer visits. Longer dwell time on the instructions may suggest that participants were reading this information carefully, as opposed to moving around quickly to find useful information for the writing task.

7.3.3 RQ2b

The second sub-question of the second research question (RQ2b) was:

> *Are there any relationships between eye-tracking measures and test-takers' performance on the task?*

The correlation analysis between the eye-tracking measures (*total visit duration, mean visit duration, max visit duration, and visit count*) and test-takers' performance on the TBEM-8 reading-to-write task yielded no statistically significant results (at the 0.05 level) except for a moderate positive correlation (ρ=.499, p=.049) between the participants' *max visit duration* on Source 5 and their performance on the task. Another point to note is that, for Source 4, the picture, almost all the correlations between the eye-tracking measures and the performance showed a negative direction, although they did not reach statistical significance. Higher-scoring participants tended to spend less time on it (they were more likely to quickly disregard the picture as less relevant overall) and visit it less often than the lower-scoring participants.

7.4 A model of the reading-to-write process

Based on the findings of this study and relevant models of reading and writing in the literature (Field, 2004; Hayes, 1996, 2012; Hayes and Flower, 1980; Kellogg, 1996; Shaw and Weir, 2007; Spivey, 1984, 1990, 1997, 2001; Spivey and King, 1989), a model of the reading-to-write process can be proposed (see Figure 7.1), attempting to illustrate what cognitive processes are likely to be involved in completing an integrated reading-to-write task and how these cognitive processes may interact with each other for successful task completion.

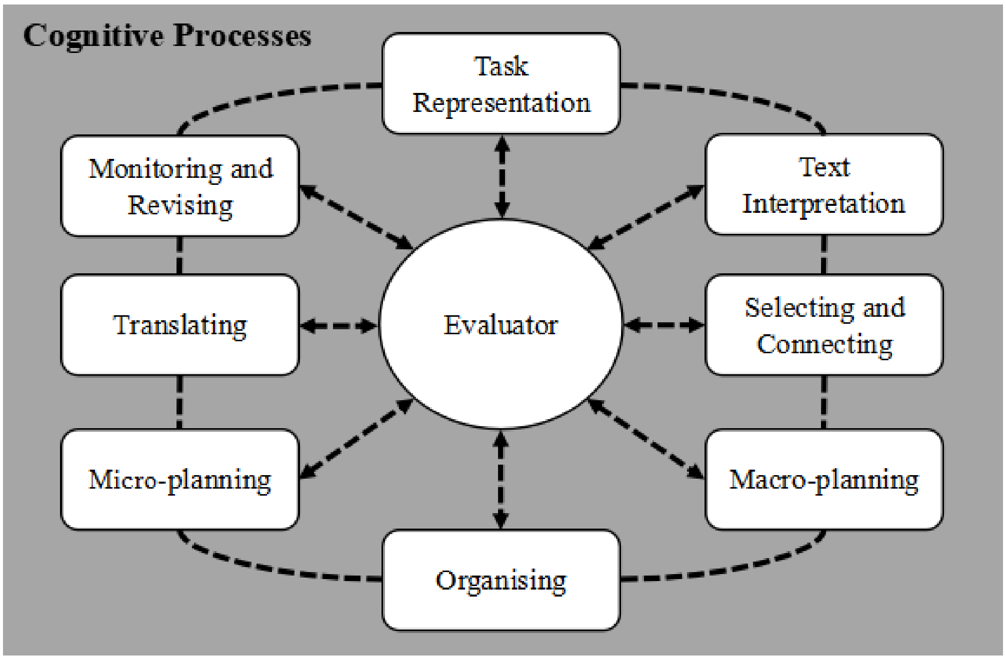

Figure 7.1: A model of the reading-to-write process

It should be noted that the central focus of this model is test-takers' internal cognitive processing; task environment (e.g., intended readership, text-written-so-far) proposed in Hayes' (1996, 2012) model is not included here for simplicity. Also, as this model has been arrived at for the TBEM-8 task type, it may or may not apply to other types of reading-to-write tasks which contain differing types of source materials and/or have different reading/writing demands.

This model consists of the ten categories of cognitive processes that have been proposed and looked at in this study, and an evaluator, which examines the outputs of any of the processes, evaluates their adequacy to the task, and determines if more types of processes are needed to fulfill certain goals. It was found, on the basis of findings in this study, that the reading-to-write process is not a linear process, but involves multiple recursions of different cognitive processes. And these processes are very likely to co-occur under various contexts during reading and writing. For example, when a writer is making micro-plans (micro-planning) for the text to be produced, they may need to go to the source materials searching for ideas to be put in the text (text-interpretation, selecting and connecting), or if they are making a rough outline for the essay (macro-planning), they may be engaged in identifying the relationships between different source materials (organising) in order to create a good representation of the sources. Another important point to note is that most of the cognitive processes may take place at any stage of writing. For example, task representation, which has been considered as creating an initial understanding of a task at an early stage of writing (Chan, 2013; Flower et al., 1990; Plakans, 2010), was found to be a circular process that occurred throughout the whole task completion in this study.

The relationships between test-takers' performance on the TBEM-8 reading-to-write task and their use of cognitive processes seem to be obscure (very few statistically significant findings were found), based on the findings of the current study. Thus it might not be test-takers' language proficiency that determines their cognitive processing during task completion, but their interpretation of the various contexts (e.g., the task set, writing goals), as well as their working memory and prior knowledge that affect the use of cognitive processes (Stein, 1990).

7.5 Summary

This chapter has discussed three topics. First, major findings from all sources of data were gathered and triangulated to answer the first set of research questions, and cognitive processes employed while completing the task were discussed in relating to relevant studies on these processes (e.g., Chan, 2013; Field, 2005; Flower et al., 1990; Plakans, 2009b; Shaw and Weir, 2007). Second, how test-takers engaged with the source materials in the TBEM-8 reading-to-write task (mainly benefiting from the eye-tracking data) were discussed to answer the second set of research questions. Third, a reading-to-process model was proposed to illustrate how different cognitive processes interact with each other while reading from sources.

In the next chapter, a summary of the current study will be provided; implications of the findings, limitations of this study, and recommendations for future research will also be discussed.

CHAPTER 8 CONCLUSION

8.1 Summary of study

This study has investigated test-takers' cognitive processing while completing the TBEM-8 reading-to-write task. Two separate studies were conducted in sequence. In Study I, 16 participants (Master's students based at Lancaster University) completed a prototypical task while their eye movements were recorded by a Tobii TX300 eye-tracker. These eye traces then formed the stimuli for a stimulated recall session to elicit their cognitive processes during task completion. In Study II, another 172 participants (Business English major undergraduate students based at two universities in China) responded to a reading-to-write process questionnaire after completing the task. This questionnaire was initially developed and validated by Chan (2013), and adapted for the TBEM-8 reading-to-write task. In addition, findings resulting from Study I fed into several revisions of the questionnaire and a pilot study was conducted to finalise the main study questionnaire, in which 40 items were grouped into five writing phases (*conceptualisation, meaning and discourse construction, organisation, low-level monitoring and revising, and high-level monitoring and revising*) to reflect the cognitive processes that test-takers were hypothesised to undergo.

The combined use of eye-tracking, stimulated recall, and questionnaire methods proved to be particularly useful in looking at test-takers' cognitive processing while completing the TBEM-8 reading-to-write task. Findings from stimulated recalls showed that the participants engaged in a wide range of cognitive processes specified in Shaw and Weir's (2007) model of writing and Spivey's (1990, 1997, 2001) discourse synthesis model, with varying frequencies. *Selecting* and *micro-planning* were the two most frequently used processes according to the participants, and *macro-planning* and *translating* were the two least reported processes (the low frequencies might be due to the limitations of the stimulated recall methodology and the automatized nature of the translating process).

Heat map outputs demonstrated that the participants' attention covered all the main parts of the TBEM-8 reading-to-write task, with various foci within each part. The eye-tracking metric of *time to first fixation*, together with participants' eye-movement recordings, illustrated how they approached the task within the first several minutes of task completion. Other metrics, including *total visit duration, visit count* and *visit duration* illustrated in detail the extent to which participants engaged with the source materials. Overall, they spent, on average, 26.4 percent (580.8 seconds) of the total processing time in reading, and 73.6 percent (1623.1 seconds) in writing. In terms of time spent on each source material, Source 2, the English source texts, received the most attention from participants. It is argued that the language of this text may be the factor that increased their processing time on this material. Source 4, the only graphic input, received the least attention from participants; the majority of test-takers reported little interest in this material, and the limited amount of time spent on it was likely due to unconscious looking behaviour. Source 5 (key concepts and expressions), though having much fewer words in it compared to other source texts, received markedly high amounts of attention from the participants, which indicated that they frequently looked for lexical support or idea hints in this material. For the participants' reading behaviour, their eye-tracking traces and the measures showed that they read both carefully at times (e.g., while reading the task instructions and source materials for the first several times before writing), and expeditiously at other times (e.g., while searching for specific information during writing).

The correlation analysis between cognitive processes/eye-tracking measures and test-takers' performance on the TBEM-8 reading-to-write task yielded no statistically significant results (at the 0.05 or 0.01 levels), except for a moderate positive correlation (ρ=.499, p=.049) between the participants' max visit duration on Source 5 (key concepts and expressions) and their reading-to-write performance, and one (ρ=.432, p=.098) between the counts of text interpretation-2 process (reading source materials) and the task performance if the p-value was set to 0.1. The correlation results, together with the stimulated recalls, may lead us to conclude that the participants' reading-to-write proficiency may not be indicative of their uses of certain cognitive processes, and how they would engage with the source materials. Rather, it might be the contexts, for example, the task set, test-takers' writing goals and prior knowledge, that affect the use of particular cognitive processes.

In the discussion, findings from the three sources of data were triangulated to answer the research questions proposed at the end of the literature review chapter, with discussions of how these results fit with previously published knowledge. Based on these findings and discussions, a reading-to-write process model was proposed, attempting to illustrate how the ten categories of processes examined in this study interact with each other for successful task completion. It was argued that the reading-to-write process is not a linear act, but involves multiple recursions of cognitive processes, which frequently co-occur and overlap with each other, looping back and forth for different purposes at different writing stages.

8.2 Implications

The findings of this study have several implications for (1) cognitive processes involved in completing an integrated writing task; (2) the socio-cognitive validation framework;

(3) the validity and development of the TBEM-8 reading-to-write task; and (4) research methodology for investigating test-takers' cognitive processes during task completion.

8.2.1 Cognitive processes involved in completing an integrated reading-to-write task

The major outcome for this study is increased understanding of cognitive processes involved in completing an integrated reading-to-write task. A total of ten categories of cognitive processes were proposed and investigated; these processes were driven from relevant reading and writing models in the established literature (see Section 2.2 for a review of these models), as well as findings of the current study which examined how test-takers employed different cognitive processes while completing the TBEM-8 reading-to-write task. Thus a more complete representation of reading-to-write processes was built, and an example process study was provided for researchers who are interested in looking at test-takers' cognitive processing while completing a reading-to-write task.

A reading-to-write process model (see Section 7.4 for details) was proposed at the end of discussion chapter, attempting to illustrate how the ten cognitive processes interact with each other while reading from sources. This fills in a research gap in previous writing models (e.g., Hayes, 1996); although the role of reading has been pointed out as important in writing processes, when and how reading interacts with other processes in writing remained largely under-represented. In addition, it extends Spivey's (1990, 1997, 2001) discourse synthesis model by introducing processes that are essential in completing a reading-to-write task (e.g., task representation, macro-planning) but were not included in that model. Also, this study demonstrated that test-takers employed a unique set of cognitive processes which were different from that used in traditional writing-only tasks (Gebril and Plakans, 2013; Plakans, 2008; Shaw and Weir, 2007). It suggests that these two types of writing tasks may be used in complementary ways, but not as substitutes for each other (Yu, 2013).

8.2.2 The socio-cognitive validation framework

Weir's (2005) socio-cognitive validation framework is regarded as having "direct relevance and value to an operational language testing/assessment context" and "to be both theoretically sound and practically useful" when developing and validating language tests (Taylor, 2011, p.2). However, few studies have applied this framework to tests that measure integrated language skills, for example, integrated reading-to-write tasks (Chan, 2013) and listening-to-summarize tasks (Rukthong, 2016).

The current study has extended the application of the socio-cognitive framework to integrated reading-to-write task, by introducing a total of ten cognitive parameters/processes for looking at the cognitive validity of a reading-to-write task. It has shown how cognitive validity may be examined for this task type, thus providing, to some extent, both theoretical and practical values for development and validation of integrated reading-to-write tasks, and assistance to test developers and researchers who intend to develop valid reading-to-write tasks and do further validation studies based on the findings of this study.

8.2.3 The TBEM-8 reading-to-write task

It is a challenging task to conduct validation studies on the cognitive validity of an integrated reading-to-write task, because a model of reading-to-write processes was lacking in the literature (Hirvela, 2004). Although this task type is generally regarded as having good cognitive validity (Plakans, 2010; Weigle, 2002; Weir, Chan &

Nakatsuhara, 2013), empirical evidence supporting this claim was scarce (the number of studies on reading-to-write processes is considerably smaller than that on independent writing tasks) and not comprehensive (few studies looked at the entire variety of reading-to-write processes).

Based on the ten cognitive parameters/processes proposed above, the current study examined the cognitive validity of the TBEM-8 reading-to-write task. Findings revealed that this task successfully elicited these processes from the participants during task completion, thus providing strong evidence for the cognitive validity of the task as a tool to assess test-takers' reading-to-write ability. Some practical implications for the development of the TBEM-8 reading-to-write task include: (1) reconsidering the use of graphic input in the source materials because the findings showed that almost all participants spent relatively little attention on this material and considered it as less useful though it took up much space in the task; and (2) the language of the input texts (English versus Chinese) may influence the degree and nature of test-takers' interaction with the text, which should be taken into consideration if a certain degree of test-takers' interacting with the sources needs to be reflected when designing the tasks.

8.2.4 Methodological implications

The methodology utilised in this study - including a detailed analysis of eye-tracking data, of the stimulated recall protocols participants produced, and of the questionnaire data - proved to be useful. Eye-tracking visualisations and metrics revealed several overall processing patterns (e.g., how participants approached the TBEM-8 reading-to-write task) and provided quantitative data on participants' looking behaviour while completing the task, whereas stimulated recalls reported more qualitative data on participants' cognitive processing during task completion. The combined use of these two methods generated rich data from participants, but it is particularly time-consuming and labour-intensive, and can only be applied to a limited number of participants in a sole-researcher project. Therefore, a reading-to-write process questionnaire developed by Chan (2013) was also utilised to elicit participants' (N=172) cognitive processes in order to offset the drawbacks of the limited sample used in the eye-tracking and stimulated recall study.

The findings resulting from the analysis of eye-tracking, stimulated recall, and questionnaire data were triangulated and provided a solid basis on which conclusions could be drawn about test-takers' cognitive processing while completing the TBEM-8 reading-to-write task. It is believed that this methodology could be of value as part of test validation studies. For example, it could be used to collect *a priori* cognitive validity evidence based not on "what the test constructors believe an item to be testing" (Alderson, 2000, p. 97), but on what processes test-takers employ for successful task completion (Brunfaut and McCray, 2015). Therefore, it could help test developers to determine whether a task accurately measure the construct intended to be measured, thus helping to minimise the two major threats to validity: construct under-representation and construct-irrelevant variance (Messick, 1989, 1992).

8.3 Limitations

Several limitations exist for the current study. Firstly, it is limited to only one integrated reading-to-write task. This is because, as stated earlier, that the TBEM-8 reading-to-write task is still at an early stage of development, and further tasks could not be provided by the TBEM-8 testing committee due to confidentiality and the small number of existing tasks. Although the study generated rich quantitative and qualitative data on test-takers' cognitive processing during task completion, findings from this study should be interpreted with caution in that test-takers may employ a varied use of cognitive processes when responding to different prompts/source materials (Yang, 2009). Nevertheless, the use of the TBEM-8 task allowed for the collection of validity evidence early in the cycle of test use, and thus will have impact on the production of future tasks.

Secondly, the combined use of eye-tracking and stimulated recall methods is quite time-consuming and labour-intensive, and it could be only applied to a relatively small number of participants (16 in this study), so that interpreting the results too broadly would pose risks, and any conclusions drawn should be seen as tentative. This limitation was, to some extent, offset by administering a reading-to-write process questionnaire to another 172 participants, however, as the number of questionnaire items is 40, which is relatively large considering the number of participants, statistical analyses such as factor analysis yielded no meaningful results.

Another limitation relates to the dependence on the stimulated recall technique to collect participants' cognitive processes while completing the TBEM-8 reading-to-write task. Although this method has been strengthened by using participants' eye movement recordings as stimuli, which, to a large extent, minimised the threat of memory decay and potential fabrication to the accuracy of the stimulated recall data, participants reported that they sometimes were not be able to recollect what they were doing when watching the replay of their eye traces. A process such as macro-planning, which may happen at an earlier stage of task completion, is less likely to be recollected and reported by participants. This may lead to an under-representation of the cognitive processing involved in completing an integrated reading-to-write task, and thus creating an incomplete picture of the task construct.

Finally, as the central focus of this study is on test-taker's cognitive processing during task completion, little attention was paid to the features of their written products. Also, although test-takers' use of cognitive processing was looked at, whether that use was successful or unsuccessful was not examined, and the resulting writing performances were not analysed (e.g., to cross-check the reported cognitive processes or find further explanations for the processing findings). Although this study may have strengthened the current understanding of the reading-to-write process by using a triangulation of three research methods, looking at test-takers' written products could help to gain further insights into the findings that emerged from this process-oriented study, as demonstrated in previous studies that focused on the discourse features of participants' written products (Cumming et al., 2005, 2006; Gebril and Plakans, 2009, 2013, 2016; Plakans and Gebril, 2017).

8.4 Recommendations for future research

The current study has uncovered important findings about test-takers' cognitive processing involved in completing an integrated reading-to-write task. However, considering the limitations of this study given above, more work is needed to build a more comprehensive understanding of this task type.

Firstly, although it might not be ideal to increase the sample size for the eye-tracking and stimulated recall study, there are still some other ways to improve the interpretability of the stimulated recall data. In this study, the stimulated recall session was unstructured, and primarily led by the participants to recollect their thoughts, though at some point researcher interventions of asking questions to clarify certain issues occurred during the session. This may still generate a rich data set of participants' cognitive processing activities, but the data could be overwhelming and displayed in an unorganised and unsystematic manner. For future research, therefore, it is recommended to carry out structured or semi-structured stimulated recall sessions to elicit participants' cognitive processes. This could be achieved by designing some preset questions to be asked for the session, or taking notes of interesting points observed while watching participants' real-time eye movements displayed on the monitoring screen during experiment, and asking them about these noted points afterwards.

Secondly, although the reading-to-write process questionnaire used in this study has largely been validated in Chan's (2013) study, there is still need to conduct further investigation on the quality of this questionnaire, because several revisions have been made to it. Therefore, it would be ideal if a larger number of participants could be recruited, and statistical analyses such as factor analysis can thus be performed to look at the quality of the questionnaire and provide further insights to the reading-to-write construct. Also, as mentioned in the limitations above, more types of integrated reading-to-write tasks should be looked at to increase the generalisability of the results.

Another important aspect that needs to be further investigated relates to verbatim source use; more specifically, whether the verbatim source use has an impact on test-takers' performance on a reading-to-write task. In this study, findings on participants' verbatim source use showed somewhat contradictory results; direct copying was not found to be a significant issue in the stimulated recalls, but questionnaire data demonstrated that quite a number of participants reported that they did copy directly from the source materials. It is of interest to find out how the use of source materials may inform test-takers' writing in terms of the language and the format, by examining their written products.

Finally, features of the written products (e.g., grammatical accuracy) also deserve more attention. As revealed in some previous studies on test-takers' written products, discourse features exert an important influence on the score assigned (Cumming et al., 2005, 2006). This is vital to a better interpretation and use of the scores on an integrated reading-to-write task. Also, raters' perception of the reading-to-write processes is of interest, that is, whether or to what extent they interpret how test-takers employ the proposed cognitive processes in their ratings.

REFERENCES

Ackerman, J. M. (1991). Reading, writing, and knowing: The role of disciplinary knowledge in comprehension and composing. *Research in the Teaching of English, 25*(2), 133-178.

Alamargot, D., & Chanquoy, L. (2001). *Through the models of writing*. Dordrecht-Boston London: Kluwer Academic Publishers.

Alderson, J.C. (2000). *Assessing reading*. Cambridge: Cambridge University Press.

Alderson, J. C., Clapham, C., & Wall, D. (1995). *Language test construction and evaluation*. Cambridge: Cambridge University Press.

Allen, S. (2004). Task representation of a Japanese L2 writing and its impact on the usage of source text information. *Journal of Asian Pacific Communication, 14*(1), 77-89.

Rukthong, A. (2016) *Investigating the listening construct underlying listening-to-summarize tasks*. Unpublished PhD thesis, Lancaster University, UK.

Asabi, A. R., Akbari, N., & Graves, B. (2006). Discourse appropriations, construction of identities, and the complex issue of plagiarism: ESL students writing in graduate school. *Journal of Second Language Writing, 15*(2), 102-117.

Ascensión, Y. (2005). *Validation of reading-to-write assessment tasks performed by second language learners*. Unpublished doctoral dissertation, Northern Arizona University, Flagstaff, Arizona.

Ascensión-Delaney, Y. (2008). Investigating the reading-to-write construct. *Journal of English for Academic Purposes, 7*(3), 140-150.

Baba, K. (2009). Aspects of lexical proficiency in writing summaries in a foreign language. *Journal of Second Language Writing*, *18*(3), 191-208.

Bachman, L. F. (2005). Building and supporting a case for test use. *Language Assessment Quarterly, 2*(1), 1-34.

Bachman, L., & Palmer, A. (2010). *Language assessment in practice*. Oxford: Oxford University Press.

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

Ball, L.J., Eger, N., Stevens, R. & Dodd, J. (2006). Applying the PEEP method in usability testing. *Interfaces, 67*, 15-19.

Bax, S. (2013). The cognitive processing of candidates during reading tests: Evidence from eye-tracking. *Language Testing*, *30*(4), 441-465.

Bax, S., & Weir, C. (2012). Investigating learners' cognitive processes during a computer-based CAE Reading test. *Cambridge Research Notes, Cambridge ESOL, 47* (February 2012), 3-14.

Bereiter, C., & Scardamalia, M. (1987). *The psychology of Written Composition*. Hillside, NJ: Lawrence Erbaum Associates.

Berninger, V., Cartwright, A., Yates, C., Swanson, H. L., & Abbott, R. (1994). Developmental skills related to writing and reading acquisition in the intermediate grades: Shared and unique variance. *Reading and Writing: An Interdisciplinary Journal, 6*(2), 161-196.

Birch, S., & Rayner, K. (1997). Linguistic focus affects eye movements during reading. *Memory & Cognition*, *25*(5), 653-660.

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.

Brown, J. D., Hilgers, T., & Marsella, J. (1991). Essay prompts and topics: Minimizing the effect of mean differences. *Written Communication, 8*(4), 533-556.

Brunfaut, T. (2016). *Looking into reading II: A follow-up study on test-takers' cognitive processes while completing Aptis B1 reading tasks*. British Council Validation Series, VS/2016/001. London: The British Council.

Brunfaut, T., & McCray, G. (2014). *Looking into reading: Pilot study*. Unpublished report. British Council Assessment Research Grant 2012.

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

Burtis, P. J., Bereiter, C., Scardamalia, M., & Tetroe, J. (1983). The development of planning in writing. In B. M. Kroll & G. Wells (Eds.), *Explorations in the development of writing* (pp. 153-174). New York: Wiley.

Campbell, C. (1990). Writing with other's words: Using background reading texts in academic compositions. In B. Kroll (Ed.), *Second language writing* (pp. 211-230). Cambridge, UK: Cambridge University Press.

Carrell, P. L. (2005). Metacognitive awareness and second language reading. *The Modern Language Journal, 73*(2), 121-134.

Chan, S. H. C. (2011). Demonstrating the cognitive validity and face validity of PTE Academic Writing items. *Pearson External Research Notes*. Retrieved from: http://pearsonpte.com/research/research-summaries-notes/.

Chan, S. H. C. (2013). *Establishing the validity of reading-into-writing test tasks for the UK academic context.* Unpublished PhD thesis, The University of Bedfordshire, UK.

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

Chan, S. H. C., Wu, R. Y. F., & Weir, C. J. (2014). Examining the context and cognitive validity of the GEPT Advanced Writing Task 1: A comparison with real-life academic writing tasks. *LTTC-GEPT Research Report, 3,* 1-91.

Chapelle, C. A., Grabe, W. & Berns, M. (1993). *Communicative language proficiency: Definitions and implications for TOEFL 2000*. [ETS International Report.] Princeton, NJ: Educational Testing Service.

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.

Cherkes-Julkowski, M., Sharp, S., & Stolzenberg, J. (1997). *Rethinking Attention Deficit Disorders*. Cambridge, MA: Brookline Books.

Connelly, V., Gee, D., & Walsh, E. (2007). A comparison of keyboarded and hand-written compositions and the relationship with transcription speed. *British Journal of Educational Psychology, 77*(2), 479-492.

Cowen, L., Ball, L. J., & Delin, J. (2002). An Eye Movement Analysis of Webpage Usability. In X. Faulkner, J. Finlay, & F. Detienne (Eds.), *People and Computers XVI - Memorable yet Invisible: Proceedings of the HCI* 2002 London: Springer-Verlag Ltd.

Cronbach, L. J. (1988). Five perspectives on validity argument. In H. Wainer & H. Braun (Eds.), *Test validity* (pp. 3-17). Hillsdale, NJ: Lawrence Erlbaum.

Currie, P. (1998). Staying out of trouble: Apparent plagiarism and academic survival. *Journal of Second Language Writing, 7*(1), 1-18.

Cumming, A. (2013). Assessing Integrated Writing Tasks for Academic Purposes: Promises and Perils. *Language Assessment Quarterly*, *10*(1), 1-8.

Cumming, A., Kantor, R., Baba, K., Erdosy, U., Eouanzoui, K., & James, M. (2005). Differences in written discourse in independent and integrated prototype tasks for next generation TOEFL. *Assessing Writing, 10*(1), 5-43.

Cumming, A., Kantor, R., Baba, K., Erdosy, U., Eouanzoui, K., & James, M. (2006). *Analysis of discourse features and verification of scoring levels for independent and integrated tasks for the new TOEFL*. (TOEFL Monograph No. MS-30 Rm 05-13) Princeton, NJ: ETS.

Daneman, M., & Reingold, E. (1993). What eye fixations tell us about phonological recoding during reading. *Canadian Journal of Experimental Psychology*, *47*(2), 153-178.

Das, J. P., & Misra, S. B. (2015). *Cognitive Planning and Executive Functions: Applications in Management and Education*. New Delhi: Sage.

Ehmke, C., & Wilson, S. (2007). Identifying web usability problems from eye-tracking data. *Proceedings of HCI 2007*.

Ellis, R. (1994). *The study of second language acquisition*. Oxford: Oxford University Press.

Elling, S., Lentz, L., & de Jong, M. (2011). Retrospective think-aloud method: using eye movements as an extra cue for participants' verbalizations. *In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 1161-1170). ACM.

Esmaeili, H. (2002). Integrated reading and writing tasks and ESL students' reading and writing performance in an English language test. *The Canadian Modern Language Review, 58*(4), 599-622.

Eysenck, M. W., & Keane, M. T. (2005). *Cognitive psychology: A student's handbook (5th ed.)*. New York, NY, US: Psychology Press.

Feak, C. & Dobson, B. (1996). Building on the impromptu: a source-based academic writing assessment. *College ESL, 6*(1), 73-84.

Field, J. (2004). *Psycholinguistics: the key concepts*. London: Routledge.

Field, J. (2008). *Listening in the Language Classroom*. Cambridge: Cambridge University Press.

Field, J. (2011). Cognitive validity. In L. Taylor (Ed.), *Examining Speaking: Research and Practice in Assessing Second Language Speaking, Studies in Language Testing 30* (pp. 65-111). Cambridge: UCLES/Cambridge University Press.

Field, J. (2013). Cognitive validity. In A. Geranpayeh & L. Taylor (Eds.), *Examining Listening: Research and Practice in Assessing Second Language Listening, Studies in Language Testing 35* (pp. 77-151). Cambridge: UCLES/Cambridge University Press.

Fischer, P. M., Richards, J. W., Berman, E. J., & Krugman, D. M. (1989). Recall and eye tracking study of adolescents viewing tobacco advertisements. *JAMA, 261*(1), 84-89.

Fitzgerald, J. (1987). Research on revision in writing. *Review of Educational Research, 57*(4), 481-506.

Flower, L., Stein, V., Ackerman, J., Kantz, M. J., McCormick, K., & Peck, W. C. (1990). *Reading-to-write: Exploring a Cognitive and Social Process*. New York: Oxford University Press.

Gass, S. M., & Mackey, A. (2000). *Stimulated recall methodology in second language research*. Routledge.

Graham, S., & Harris, K. R. (1996). Self-regulation and strategy instruction for students who find writing and learning challenging. In C. M. Levy & S. Ransdell (Eds.), *The science of writing: Theories, methods, individual differences and applications*. (pp. 347-360). Mahwah, New Jersey: Lawrence Erbaum Associates.

Graham, S., & Harris, K. (2000). The role of self-regulation and transcription skills in writing and writing development. *Educational Psychologist, 35*(1), 3-12.

Gebril, A. (2006). *Independent and integrated academic writing tasks: A study in generalizability and test method*. Unpublished doctoral dissertation, The University of Iowa, Iowa City.

Gebril, A. (2009). Score generalizability of academic writing tasks: Does one test method fit it all? *Journal of Language Testing, 26*(4), 507-531.

Gebril, A., & Plakans, L. (2009). Investigating source use, discourse features, and process in integrated writing tests. *Spaan Working Papers in Second or Foreign Language Assessment, 7*, 47-84.

Gebril, A., & Plakans, L. (2013). Toward a Transparent Construct of Reading-to-Write Tasks: The Interface Between Discourse Features and Proficiency. *Language Assessment Quarterly*, *10*(1), 9-27.

Gebril, A. & Plakans, L. (2016). Source-based tasks in academic writing assessment: Lexical diversity, textual borrowing and proficiency. *Journal of English for Academic Purposes, 24*, 78-88.

Grabe, W. (2003). Reading and writing relations: Second language perspectives on research and practice. In B. Kroll (Ed.), *Exploring dynamics of second language writing* (pp.242-262). Cambridge: Cambridge University Press.

Grabe, W., & Kaplan, F. L. (1996). *Theory and Practice of Writing: An applied linguistic perspective*. London: Longman.

Graham, S., & Harris, K. R. (1996). Self-regulation and strategy instruction for students who find writing and learning challenging. In C. M. Levy & S. Ransdell (Eds.), *The science of writing: Theories, methods, individual differences, and applications* (pp. 347-360). Hillsdale, NJ, US: Lawrence Erlbaum Associates, Inc.

Graham, S., & Harris, K. R. (2000). The role of self-regulation and transcription skills in writing and writing development. *Educational Psychologist, 35*(1), 3-12.

Green, A. (1998). *Verbal Protocol Analysis in Language Testing Research: a handbook*. Studies in Language Testing, 5. Cambridge: Cambridge University Press.

Greene, S. (1993). The role of task in the development of academic thinking through reading and writing in a college history course. *Research in the Teaching of English, 27*(1), 46-75.

Guion, R.M. (1980). On Trinitarian doctrines of validity. *Professional Psychology, 11*(3), 385-398.

Guo, L., Crossley, S., & McNamara, D. (2013). Predicting human judgments of essay quality in both integrated and independent second language writing samples: a comparison study. *Assessing Writing, 18*(3), 218-238.

Hamp-Lyons, L., & Kroll, B. (1997). *TOEFL 2000: Writing: Composition, community and assessment*. Princeton, NJ: Educational Testing Service.

Hansen, J. P. (1991). The use of eye mark recordings to support verbal retrospection in software testing. *Acta Psychologica, 76*(1), 31-49.

Harwood, N. (2009). An interview-based study of the functions of citations in academic writing across two disciplines. *Journal of Pragmatics, 41*(3), 497-518.

Hayes, J. R. (1996). *A new Framework for understanding cognition and affect in writing*. In S. Ransdell (Ed.), (pp. 1-28). Mahwah, New Jersey: Lawrence Erbaum Associates.

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

Hayes, J. R. & Berninger, V. W. (2009). Relationships between idea generation and transcription: How the act of writing shapes what children write. In: Braverman RK, Lunsford K, McLeod S, Null S, Rogers ASP, editors. *Traditions of writing research* (pp. 166-180). New York: Taylorand Frances/Routledge.

Hayes, J. R., & Berninger, V. (2014). Cognitive processes in writing: A framework. In B. Arfé, J. Dockrell, & V. Berninger (Eds.), *Writing development and instruction in children with hearing, speech, and language disorders* (pp. 3-15). New York, NY: Oxford University Press.

Hayes, J. R., & Chenoweth, N. A. (2006). Is working memory involved in the transcribing and editing of texts? *Written Communication*, *23*(2), 135-149.

Hayes, J. R., & Flower, L. (1980). Identifying the organisation of writing processes. In L. W. Gregg & E. R. Steinberg (Eds.), *Cognitive processes in writing* (pp. 3-30). Hillsdale, N. J.: Lawrence Erbaum Associates.

Hirvela, A. (2004). *Connecting reading and writing in second language writing instruction*. Ann Arbor, MI: The University of Michigan Press.

Holmqvist, K., Nystrom, M., Andersson, R., Dewhurst, R., Jarodzka, H., & Van de Weijer, J. (2011). *Eye Tracking. A comprehensive guide to methods and measures.* Oxford: Oxford University Press.

Holzknecht, F., Eberharter, K., Kremmel, B., Zehentner, M. McCray, G. Konrad, E., & Spöttl, C. (2017). *Looking into Listening: Using eye-tracking to establish the cognitive validity of the Aptis Listening Test*. London: The British Council (ARAGs Research Reports Online; vol. AR-G/2017/3).

Huey, E.B. (1908). *The psychology and pedagogy of reading*. New York: Macmillan.

Hyland, K. (2002). *Teaching and Researching Writing, Applied Linguistics in Action Series*. London: Longman.

Hyrskykari, A., Ovaska, S., Majaranta, P., Räihä, K. J., & Lehtinen, M. (2008). Gaze path stimulation in retrospective think-aloud. *Journal of Eye Movement Research*, *2*(4), 1-18.

Inhoff, A.W., Topolski, R., Vitu, F., & O'Regan, J.K. (1993). Attention demands during reading and the occurrence of brief (express) fixations. *Perception & Psychophysics*, *54*(6), 814-823.

Jacob, R.J., & Karn, S.K. (2003). Eye tracking in human-computer interaction and usability research: Ready to deliver the promises. In J. H. Radach & H.Deubel (Eds.), *In the mind's eye: Cognitive and applied aspects of eye movement research.* (pp. 573-605). Amsterdam: Elsevier Science.

Jacobs, A.M. (1986). Eye movement control in visual search: How direct is visual span control? *Perception & Psychophysics, 39*(1), 47-58.

Jacobs, A.M. (1987). On localization and saccade programming. *Vision Research, 27*(11), 1953-1966.

Jarvis, S., Grant, L., Bikowski, D., & Ferris, D. (2003). Exploring multiple profiles of highly rated learner compositions. *Journal of Second Language Writing*, *12*(4), 377-403.

Jennings, M., Fox, J., Graves, B., & Shohamy, E. (1999). The test-takers' choice: An investigation of the effect of topic on language-test performance. *Language Testing, 16*(4), 426-456.

Johns, A. M., & Mayes, P. (1990). An analysis of summary protocols of university ESL students. *Applied Linguistics, 11*(3), 253-271.

Jones, D. A., & Christensen, C. A. (1999). Relationship between automaticity in handwriting and students' ability to generate written text. *Journal of Educational Psychology, 91*(1), 44-49.

Just, M. A., & Carpenter, P. A. (1976). Eye fixations and cognitive processes. *Cognitive psychology, 8*(4), 441-480.

Kane, M. T. (1992). An argument-based approach to validity. *Psychological Bulletin, 122*(3), 527-535.

Kane, M. T. (2001). Current concerns in validity theory. *Journal of Educational Measurement, 38*(4), 319-342.

Kane, M. T. (2002). Validating high-stakes testing programs. *Educational Measurement: Issues and Practices, 21*(1), 31-41.

Kane, M. T., Crooks, T. J., & Cohen, A. S. (1999). Validating measures of performance. *Educational Measurement: Issues and Practice, 18*(2), 5-17.

Kellogg, R. T. (1994). *The psychology of writing*. New York: Oxford University Press.

Kellogg, R. T. (1996). A model working memory in writing. In C. M. Levy & S. Ransdell (Eds.), *The science of writing: Theories, methods, individual differences and applications*. (pp. 57-71). Mahwah, New Jersey: Lawrence Erbaum Associates, Publishers.

Kellogg, R.T., Whiteford, A.P., Turner, C.E., Cahill, M., & Mertens, A. (2013). *Journal of Writing Research, 5*(2), 159-190.

Khalifa, H., & Weir, C. J. (2009). *Examining reading: Research and practice in assessing second language reading, Studies in Language Testing 29*. Cambridge: UCLES/Cambridge University Press.

Kim, B., Dong, Y., Kim, S., & Lee, K-P. (2007). Development of Integrated Analysis System and Tool of Perception, Recognition, and Behavior for Web Usability Test: With Emphasis on Eye-Tracking, Mouse-Tracking, and Retrospective Think Aloud, *Usability and Internationalization. HCI and Culture*, 113-121.

Koda, K. (2005). *Insights into Second Language Reading: A Cross-Linguistic Approach*. Cambridge: Cambridge University Press.

Kucer, S. (1985). The Making of Meaning. *Written Communication*, *2*(3), 317-336.

Kunnan, A. J. (Ed.) (1998). *Validation in Language Assessment*. Mahwah, NJ: Lawrence Erlbaum Associates.

Lado, R. (1961). *Language Testing*. London: Longman.

Lee, H. K., & Anderson, C. (2007). Validity and topic generality of a writing performance test. *Language Testing, 24*(3), 307-330.

Lee, Y., & Kantor, R. (2005). Dependability of new ESL writing test scores: Tasks and alternative rating schemes (TOEFL Monograph Series No. 31). Princeton, NJ: ETS.

Lenski, S. D. (1998). Strategic knowledge when reading in order to write. *Reading Psychology: An International Quarterly, 19*(3), 287-315.

Leow, R. P., & Morgan-Short, K. (2004). To think aloud or not to think aloud: The issue of reactivity in SLA research methodology. *Studies in second language acquisition*, *26*(1), 35-57.

Levelt, W. J. M. (1989). *Speaking. From intention on articulation*. Cambridge, MA-London: ACL-MIT Press.

Levin, J. R. (1988). Elaboration-based learning strategies: Powerful theory = powerful application. *Contemporary Educational Psychology, 13*(3), 191-205.

Lewkowicz, J. (1994). *Writing from sources: does source material help or hinder students? performance? Language and learning*. Papers presented at the Annual International Language in Education Conference ed. by Nancy Bird. et al. ERIC Document (ED 386 050).

MacArther, C. A., Graham, S., & Harris, K. R. (2004). Insights from instructional research on revision with struggling writers. In L. Allal, L. Chanquoy, & P. Largy (Eds.), *Revision: Cognitive and instructional processes* (pp. 125-137). Amsterdam, Netherlands: Kluwer Academic Press.

Marsella, J., Hilgers, T. L., & McLaren, C. (1992). How students handle writing assignments: A study of eighteen responses in six disciplines. In A. Herrington & C. Moran (Eds.), *Writing teaching and learning in the disciplines* (pp. 174-188). New York: Modern Language Associates.

Mathison, M. A., & Spivey, N. N. (1993). *Writing from academic sources: Authorship in writing the critique*. Berkeley, CA.: National Center for the Study of Writing and Literacy.

McConkie, G. W., Zola, D., Grimes, J., Kerr, P.W., Bryant, N.R., & Wolff, P.M. (1991). Children's eye movements during reading. In J.F. Stein (Ed.), *Vision and visual dyslexia* (pp. 251-262). London: Macmillan Press.

McCray, G. (2013). *Statistical modelling of cognitive processing in reading comprehension in the context of language testing*. Unpublished PhD thesis, Lancaster University, UK.

McCray, G., Alderson, J. C., & Brunfaut, T. (2012). *Validity in reading comprehension items: triangulation of eye-tracking and stimulated recall data*. Paper presented at the EALTA conference, University of Innsbruck, Austria.

McCray, G., & Brunfaut, T. (2018). Investigating the construct measured by banked gap-fill items: Evidence from eye-tracking. *Language Testing, 35*(1), 51-73.

Messick, S. A. (1988). The once and future issues of validity: assessing the meaning and consequences of measurement. In H. Wainer & H. I. Braun (Eds.), *Test Validity*. Hillsdale, NJ: Lawrence Erlbaum.

Messick, S. A. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement*. New York: Macmillan.

Messick, S. (1992). Validity of test interpretation and use. In M. C. Alkin (Ed.), *Encyclopedia of Educational Research* (6th ed.) (pp. 1487-1495). NY: Macmillan.

Messick, S. A. (1994). The interplay of evidence and consequences in the validation of performance assessments. *Educational Researcher, 23*(2), 13-23.

Messick, S. A. (1995). Validity of psychological assessment: validation of inferences from person's responses and performances as scientific inquiry into score meaning. *American Psychologist, 50*(9), 741-749.

Mislevy, R.J., Steinberg, L.S., & Almond, R.G. (2002). On the roles of task model variables in assessment design. In S. Irvine & P. Kyllonen (Eds.), *Generating items for cognitive tests: Theory and practice* (pp. 97-128). Hillsdale, NJ: Erlbaum.

Mislevy, R. J., Steinberg, L. S., & Almond, R. G. (2003). On the structure of educational assessments. *Measurement: Interdisciplinary Research and Perspectives, 1*(1), 3-62.

Moore, T., Morton, J., & Price, S. (2010). *Construct validity in the IELTS academic reading test: A comparison of reading requirements in IELTS test items and in university study*. Unpublished IELTS Research Report, British Council/IDP Australia.

Murray, W.S., & Kennedy, A. (1998). Spatial coding in the processing of anaphor by good and poor readers: Evidence from eye movement analyses. *Quarterly Journal of Experimental Psychology*, *40*(4), 693-718.

Namahn (2001). Using eye tracking for usability testing. Retrieved June 1, 2016, from www.namahn.com/resources/documents/note-eyetracking.pdf.

Olsen, A. (2012). The Tobii I-VT fixation filter. Retrieved from www.acuity-ets.com/downloads/Tobii%20I-VT%20Fixation%20Filter.pdf (accessed 15 June 2016).

Ortega, L. (2003). Syntactic complexity measures and their relationship to L2 proficiency: A research synthesis of college-level L2 writing. *Applied Linguistics*, *24*(4), 492-518.

O'Sullivan, B., & Weir, C.J. (2011). Test development and validation. In B. O'Sullivan (Ed.), *Language testing theories and practices* (pp. 13-32). Basingstoke: Palgrave Macmillan.

Pennycook, A. (1996). Borrowing others? words: Text, ownership, memory, and plagiarism. *TESOL Quarterly, 30*(2), 201-230.

Perl, S. (1979). The composing processes of unskilled college writers. *Research in the Teaching of English, 13*(4), 317-336.

Peyrichoux, I., & Robillard-Bastien, A. (2006). Maximize Usability Testing Benefits with Eye Tracking. In *Getting a Measure of Satisfaction from Eyetracking in Practice-Workshop at CHI* (Vol. 2006).

Plakans, L. (2008). Comparing composing processes in writing-only and reading-to-write test tasks. *Assessing Writing, 13*(2), 79-150.

Plakans, L. (2009a). The role of reading strategies in integrated L2 writing tasks. *Journal of English For Academic Purposes*, *8*(4), 252-266.

Plakans, L. (2009b). Discourse synthesis in integrated second language writing assessment. *Language Testing, 26*(4), 561-587.

Plakans, L. (2010). Independent versus integrated writing tasks: A comparison of task representation. *TESOL Quarterly, 44*(1), 185-194.

Plakans, L. (2015). Integrated second language writing assessment: why? what? how? *Language and Linguistics Compass, 9*(4), 2015, 159-167.

Plakans, L., & Gebril, A. (2017). Exploring the relationship of organization and connection with scores in integrated writing assessment. *Assessing Writing, 31*, 98-112.

Pollatsek, A., Rayner, K., & Collins, W.E. (1984). Integrating pictorial information across eye movements. *Journal of Experimental Psychology: General, 113*(3), 426-442.

Purpura, J. E. (2014). Cognition and language assessment. In A. J. Kunnan(Ed.), *The companion to language assessment* (pp. 1452-1476). Oxford, UK: John Wiley & Sons.

Rayner, K. (1978). Eye movements in reading and information processing. *Psychological Bulletin, 85*(3), 618-660.

Rayner, K. (1985). The role of eye movements in learning to read and reading disability. *Remedial and Special Education*, *6*(6), 53-60.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin, 124*(3), 372-422.

Rayner, K. (2009). Eye movements and attention in reading, scene perception, and visual search. *The Quarterly Journal of Experimental Psychology*, *62*(8), 1457-1506.

Rayner, K., & McConkie, G.W. (1976). What guides a reader's eye movements? *Vision Research*, *16*(8), 829-837.

Rayner, K., & Morrison, R.M. (1981). Eye movements and identifying words in parafoveal vision. *Bulletin of the Psychonomic Society, 17*(3), 135-138.

Rayner, K., Raney, G.E., & Pollatsek, A. (1995). Eye movements and discourse processing. In R.F. Lorch & E.J. O'Brien (Eds.), *Sources of coherence in reading* (pp. 9-36). Hillsdale, NJ: Erlbaum.

Reichle, E.D., Pollatsek, A., Fisher D.L., & Rayner, K. (1998). Toward a model of eye movement control in reading. *Psychological Review, 105*(1), 125-157.

Ruiz-Funes, M. (1999). Writing, reading, and reading-to-write in a foreign language: A critical review. *Foreign Language Annals, 32*(4), 514-526.

Ruiz-Funes, M. (2001). Task representation in foreign language reading-to-write. *Foreign Language Annals, 34*(1), 226-234.

Russo, J. E. (1979). A software system for the collection of retrospective protocols prompted by eye fixations. *Behavior Research Methods & Instrumentation*, *11*(2), 177-179.

Salvucci, D.D., & Goldberg, J.H. (2000). Identifying fixations and saccades in eye-tracking protocols. *Proceedings of the Eye Tracking Research and Applications Symposium 2000.* (pp. 71-78). NY: ACM Press.

Sasaki, M. (2000). Toward an empirical model of EFL writing processes: An exploratory study. *Journal of Second Language Writing*, *9*(3), 259-291.

Sanders, A.F. (1993). Processing information in the functional visual field. In G. d'Ydewalle & J. Van Rensbergen (Eds.), *Perception and cognition: Advances in eye movement research* (pp. 3-22). Amsterdam: North Holland.

Scardamalia, M., & Bereiter, C. (1987). Knowledge telling and knowledge transforming in written composition. In S. Rosenberg (Eds.), *Advances in Applied Psycholinguistics, Volume 2: Reading, writing and language learning*. Cambridge: Cambridge University Press.

Scardamalia, M., & Bereiter, C. (1996). Adaptation and understanding: A case for new cultures of schooling. In S. Vosniadou, E. DeCorte, R. Glaser, & H. Mandl (Eds.), *International perspectives on the design of technology-supported learning environments* (pp. 149-163). Mahwah, NJ: Erlbaum.

Sereno, S.C., & Rayner, K. (2003). Measuring word recognition in reading: eye movements and event-related potentials. *Trends in Cognitive Science, 7*(11): 489-493.

Shaw, S., & Weir, C.J. (2007). *Examining writing: Research and practice in assessing second language writing, Studies in Language Testing 26*. Cambridge: UCLES/Cambridge University Press.

Shi, L. (2004). Textual borrowing in second-language writing. *Written Communication, 21*(2), 171-200.

Smagorinsky, P. (1994). *Speaking about writing: Reflections on research methodology*. USA: Sage Publications, Inc.

Spivey, N. N. (1984). *Discourse synthesis: Constructing texts in reading and writing. Outstanding Dissertation Monograph*. Newark, DE: International Reading Association.

Spivey, N. N. (1987). Construing constructivism: Reading research in the United States. *Poetics, 16*(2), 169-192.

Spivey, N. N. (1990). Transforming texts: Constructive processes in reading and writing. *Written Communication, 7*(2), 256-87.

Spivey, N. N. (1991). The shaping of meaning: options in writing the comparison. *Research in the Teaching of English, 25*(4), 390-418.

Spivey, N. N. (1997). *The constructivist metaphor: Reading, writing and the making of meaning*. San Diego, CA: Academic Press.

Spivey, N. N. (2001). Discourse synthesis: Process and product. *Discourse synthesis: Studies in historical and contemporary social epistemology* (pp. 379-396). Westport, CT: Praeger.

Spivey, N., & King, J. R. (1989). Readers as writers composing from sources. *Reading Research Quarterly, 24*(1), 7-26.

Stein, V. (1990). Exploring the cognition of reading-to-write. In L. Flower, V. Stein, J. Ackerman, M. J. Kantz, K. McCormick, & W. C. Peck (Eds.), *Reading to write. Exploring a cognitive & social process* (pp. 119-43). New York: Oxford University Press.

Stratman, J. F., & Hamp-Lyons, L. (1994). Reactivity in concurrent think-aloud protocols: Issues research. In P. Smagorinsky (Ed.), *Speaking about writing: Reflections on research methodology* (pp. 89-112). USA: Sage Publications, Inc.

Taylor, S.E. (1965). Eye movements while reading: facts and fallacies. *American Educational Research Journal*, *2*(4), 187-202.

Taylor, L. (Ed). (2011). *Examining Speaking: Research and practice in assessing second language speaking, Studies in Language Testing 30*. Cambridge: UCLES/Cambridge University Press.

Toulmin, S. (2003). *The uses of argument*. Cambridge, UK: Cambridge University Press.

Urquhart, A. H. & Weir, C. J. (1998). *Reading in a second language: process, product and practice*. Harlow: Longman.

Uttal, W.R., & Smith, P. (1968). Recognition of alphabetic characters during vocabulary eye movements. *Perception & Psychophysics, 3*(4A), 257-264.

Vernon, M.D. (1928). Methods of recording eye movements. *The British Journal of ophthalmology*, 12, 113-130.

Watanabe, Y. (2001). *Read-to-write tasks for the assessment of second language academic writing skills: Investigating text features and rater reactions*. Unpublished doctoral dissertation, University of Hawaii, Manoa.

Weigle, S. C. (2002). *Assessing Writing*. Cambridge: Cambridge University Press.

Weigle, S. C. (2004). Integrating reading and writing in a competency test for non-native speakers of English. *Assessing Writing, 9*(1), 27-55.

Weir, C J. (2005). *Language Testing and Validation: An Evidence-Based Approach*. Basingstoke: Palgrave Macmillan.

Weir, C J, Chan, S. H. C., & Nakatsuhara, F. (2013). Examining the Criterion-related Validity of the GEPT Advanced Reading and Writing Tests: Comparing GEPT with IELTS and Real-life Academic Performance. *LTTC-GEPT Research Reports, RG-01*, 1-43.

Weir, C., Vidakovic, I., & Dimitrova-Galaczi, E. (2013). *Measured constructs*. Cambridge: Cambridge University Press.

Wolfersberger, M. A. (2007). *Second language writing from sources: An ethnographic study of an argument essay task*. Unpublished doctoral dissertation, University of Auckland, Auckland, NZ.

Yang, H. (2009). *Exploring the complexity of second language writers' strategy use and performance on an integrated writing test through SEM and qualitative approaches*. Unpublished doctoral dissertation, The University of Texas at Austin, Austin, Texas.

Yoshida, M. (2007). *The effects of task, text, and proficiency on second language reading*. Unpublished doctoral dissertation. Temple University.

Yu, G. (2013). From Integrative to Integrated Language Assessment: Are We There Yet? *Language Assessment Quarterly*, *10*(1), 110-114.

Yu, G., He, L., & Isaacs, T. (2017). *The cognitive processes of taking IELTS Academic writing task one: An eye-tracking study*. IELTS Research Reports Online Series.

Yu, G., Rea-Dickins, P. M. & Kiely, R. (2011). *The cognitive processes of taking IELTS Academic Writing Task 1, IELTS Research Reports Volume 11* (pp. 373-449). Canberra: IDP: IELTS Australia & London: British Council.

APPENDIX A

PARTICIPANT INFORMATION SHEET

Investigating the TBEM-8 Writing Construct

Researcher: Pucheng Wang

General description

I would like to invite you to participate in a research study that investigates second language speakers' thoughts whilst completing English reading-to-write tasks. I am a PhD student in the Department of Linguistics and English Language at Lancaster University and I am carrying out this study as part of my Doctoral studies. Please take time to read the following information carefully before you decide whether or not you wish to take part.

What your participation would involve

For the study's purpose, I will make use of eye-tracking software. If you decide to participate, I will first assess your suitability for research involving eye tracking (e.g. some types of glasses impact on the accuracy of recording eye movements). You will be seated in front of a computer screen and be asked to follow the movements of a dot on the screen. This screening will take approximately 5 minutes.

If the screening is satisfactory, I will ask you to attend the following session. In this session, you will be seated in front of a computer screen and will be asked to complete the TBEM-8 (Test for Business English Majors, Band 8) Writing Test. While performing the writing task, your eye movements will be registered by the computer. After you finish the task, a stimulated recall session (where the recorded information from your writing task will be used) will be conducted. You will be shown the recordings of your eye movements and asked to recollect what you were thinking whilst completing the writing task. The whole experiment will last approximately 1.5 hours.

In return for your participation and your time in this study, I am able to offer you **£10**. If you were so kind to take part in the eye-tracking screening, but this was not successful, I am able to offer you a small reward (chocolates) for your time.

How the data will be handled

Your participation in this study is not related to your studies at Lancaster University, and your performance will not affect your evaluation on your degree programme. At every stage, your name will remain confidential, and the data will be made anonymous for reporting purposes. The data will be kept securely and will be used for research purposes only (academic publications, conference presentations). Only the researcher and the research assistants will have access to the data. Any paper-based data will be kept in a locked cupboard in the researcher's and research assistants' office, or in the eye-tracking lab. Electronic data will be stored on a password protected computer and files containing personal data will be encrypted. The data will be retained for 10 years.

The results of the study will be used for academic purposes only. This will include my PhD thesis and other publications, for example journal articles. I am also planning to present the results of my study at academic conferences.

Withdrawing from the study

You are free to withdraw from the study at any time by e-mailing me, Pucheng Wang, at p.wang1@lancaster.ac.uk. In that case, the data will be destroyed and not used. However, please be informed that if you withdraw more than 2 weeks after your participation, the data will be used in the study.

Any questions?

If you have any queries about the study, please feel free to contact me, at p.wang1@lancaster.ac.uk or my supervisor, Dr. Luke Harding, at l.harding@lancaster.ac.uk, 01524 593034. If at any stage of the study you wish to speak to an independent person about this project, you are welcome to contact the Head of Department, Prof. Elena Semino, at e.semino@lancaster.ac.uk, 01524 594176. All enquiries will be treated confidentially.

This study has been reviewed by and approved by members of Lancaster University Research Ethics Committee.

I would be very grateful if you would agree to take part!

Pucheng Wang
Email: p.wang1@lancaster.ac.uk
Linguistics and English Language
County South, Lancaster University
LA1 4YL, United Kingdom

Lancaster University
Lancaster LA1 4YL
United Kingdom
Tel: +44 (0)1524 593045
Fax: +44 (0)1524 843085
http://www.ling.lancs.ac.uk

APPENDIX B

Lancaster University
Linguistics and English Language

CONSENT FORM

Project title: ***Investigating the TBEM-8 Writing Construct***

1. I have read and had explained to me by **Pucheng Wang** the Information Sheet relating to this project.

2. I have had explained to me the purposes of the project and what will be required of me, and any questions have been answered to my satisfaction. I agree to the arrangements described in the Information Sheet in so far as they relate to my participation.

3. I understand that my participation is entirely voluntary and that I have the right to withdraw from the project within the time indicated on the Information Sheet.

4. I have received a copy of this Consent Form and of the accompanying Information Sheet.

Name:

Signed:

Date:

APPENDIX C

Task 2: Essay writing (20%) **(40 minutes)**

Directions: *In this task, you are required to write an essay of 250-280 words as an assignment for your professor of Strategic Management. You will be given 40 minutes to write an essay entitled* ***The Post Jobs' Era of Apple.*** *In the essay, you should describe the event, analyze the situation and comment on the impact of Jobs' resignation on Apple. Your essay is to be based on the source materials given below. But you should not simply copy and translate the source materials. Your essay will be judged according to how well you develop your ideas and how coherent your essay is. The task is to be completed on* ***Answer Sheet 3****.*

SOURCE MATERIALS:

乔布斯与苹果公司:

史蒂夫·乔布斯（Steve Jobs）是苹果公司的前任首席兼创办人之一，他是 iMac、iPad、iPod、iTunes Store 及 iPhone 等产品的缔造者。2007 年他被《财富》杂志评为年度最伟大商人，2009 年入选《时代周刊》年度风云人物之一。

2011 年 8 月 10 日，苹果公司股票收盘价为每股 363.69 美元，公司市值约 3372 亿美元，成为全球市值最高的上市公司。

2011 年 8 月 25 日，苹果公司宣布乔布斯辞去首席执行官一职。

美国苹果公司 5 日宣布，该公司前首席执行官史蒂夫·乔布斯已经去世（新华网旧金山 2011 年 10 月 5 日电）。(216)

Aug 25, 2011

Video News

Steve Jobs: Shoppers' views on changes at Apple
--BBC

Steve Jobs Resignation: Impact on Tech Industry
http://online.wsj.com

Steve Jobs' Resignation and Apple Stocks
http://abcnews.go.com

Internet News

After 14 years as Apple's CEO, Steve Jobs resigned his post on Wednesday and was replaced by Tim Cook, who previously was the company's Chief Operating Officer. Jobs, in turn, was elected as chairman of Apple's board of directors.
http://www.macworld.com

No man is irreplaceable, and Apple is packed with brilliant engineers, designers and managers. The question now is whether it can continue to "think different" without the man who made that into a personal and professional credo.
http://www.bbc.co.uk/news/

国内媒体相关报道：

苹果公司网站发布的消息说：“苹果失去了一位富有远见和创造力的天才，世界失去了一个不可思议之人。”

— 新华网旧金山 2011 年 10 月 5 日

“没错，乔布斯卸任后的确会出现一些变化，但这并不意味着苹果会发生根本性的转变。”他说，“如果苹果的竞争对手认为这是一个可以利用的机会，那就太愚蠢了。”

—《金融投资报》2011 年 8 月 26

乔布斯并不是什么创意高手，只是他恰巧雇用了一些能够慧眼识才的人，而他们能够透过创意看到创意本身具备的颠覆个人娱乐领域的能力。乔布斯也是个聪明人，毕竟，他的身家净值已达到数十亿，而且正是在他的带领下，苹果实现了持续快速增长。

— http://www.sina.com.cn 2011 年 9 月 3 日

Key concepts and expressions (*For reference only*):

Charisma	Differential competitive advantage
Innovation	Leadership
Listed company	Market share
Market value	Strategic management
Sustainable growth	Vision

APPENDIX D

Task 2: Essay writing (20%) **(40 minutes)**

Directions: *In this task, you are required to write an essay of 250-280 words as an assignment for your professor of Strategic Management. You will be given 40 minutes to write an essay entitled* ***The Post Jobs' Era of Apple.*** *In the essay, you should describe the event, analyze the situation and comment on the impact of Jobs' resignation on Apple. Your essay is to be based on the source materials given below. But you should not simply copy and translate the source materials. Your essay will be judged according to how well you develop your ideas and how coherent your essay is. The task is to be completed on* ***Answer Sheet 3****.*

SOURCE MATERIALS:

Steve Jobs and Apple:

Steve Jobs was the chief executive officer (CEO), and a co-founder of Apple Inc. He created products including iMac, iPad, iPod, iTunes and iPhone. In 2007, he was named as 'the greatest entrepreneur' by Fortune Magazine; In 2009, he was selected as Time magazine's Person of the Year.

On August 10, 2011, Apple became the world's most-valuable company, with a market capitalisation of $337.2 billion and $363.69 per share.

On August 25, 2011, Apple announced that Jobs has resigned as chief executive of Apple.

On October 5, 2011, Apple announced that its former CEO, Steve Jobs has passed away. (Xinhua News, October 5, 2011 reported)

Media reports in China：

Official obituary on the website of Apple says “Apple has lost a visionary and creative genius, and the world has lost an amazing human being”.
-Xinhua News, San Francisco, October 5, 2011

“Changes may happen after Jobs’ resignation, but it does not mean Apple will experience a fundamental alteration”, he says, “It would be naive if its competitors think they can take this opportunity to beat Apple”
-Financial Investment, August 26, 2011

Jobs is not a creative person, it is that he happens to hire some employees that are able to find the talents, who are creative and able to identify the power of originality. But Jobs is a smart person, who has a net worth of about $1 billion, and helps Apple achieving a sustainable high growth.

- http://www.sina.com.cn, September 3, 2011

Key concepts and expressions (*For reference only*):

Charisma	Differential competitive advantage
Innovation	Leadership
Listed company	Market share
Market value	Strategic management
Sustainable growth	Vision

APPENDIX E

Writing Process Questionnaire

Name: _________ Gender: ___ TEM-4 result: ___ TEM-8: result ___

IELTS results (if any): Overall band ____ Reading ____ Writing ____

In this questionnaire, there are some statements about how you might complete the test you have just taken. Please answer all the questions, thinking about what you did

- **While reading the task prompt**
- **While reading the source texts**
- **Before writing**
- **While writing the first draft**
- **After writing the first draft**

in the test taking experience you have just had.

Please circle the extent of your agreement or disagreement to each statement below, using the following 5-point scale, for Example:

I find academic writing to be easy.
1.Strongly disagree 2.Disagree 3.No view 4.Agree 5. Strongly agree

I. While reading the task prompt

1.1 I read the whole task prompt (i.e. instructions) carefully

1.Strongly disagree 2.Disagree 3.No view 4.Agree 5.Strongly agree

1.2 I thought of what I might need to write to make my text relevant and adequate to the task.

1.Strongly disagree 2.Disagree 3.No view 4.Agree 5.Strongly agree

1.3 I thought of how my text would suit the expectations of the intended reader.

1.Strongly disagree 2.Disagree 3.No view 4.Agree 5.Strongly agree

1.4 I was able to understand the instructions for this writing task very well.

1.Strongly disagree 2.Disagree 3.No view 4.Agree 5.Strongly agree

1.5 After reading the prompt, I thought about the purpose of the task.

1.Strongly disagree 2.Disagree 3.No view 4.Agree 5.Strongly agree

What else did you do while reading the prompt?

II. While reading the source texts

2.1 I read through the whole of each source text carefully.

1.Strongly disagree 2.Disagree 3.No view 4.Agree 5.Strongly agree

2.2 I read the whole of each source text more than once.

1.Strongly disagree 2.Disagree 3.No view 4.Agree 5.Strongly agree

2.3 I searched quickly for part(s) of the texts which might help complete the task.

1.Strongly disagree 2.Disagree 3.No view 4.Agree 5.Strongly agree

2.4 I read some relevant part(s) of the texts carefully.

1.Strongly disagree 2.Disagree 3.No view 4.Agree 5.Strongly agree

2.5 I read the task prompt again.

1.Strongly disagree 2.Disagree 3.No view 4.Agree 5.Strongly agree

2.6 I took notes on or underlined the important ideas in the source texts.

1.Strongly disagree 2.Disagree 3.No view 4.Agree 5.Strongly agree

2.7 I prioritised important ideas in the source texts in my mind.

1.Strongly disagree 2.Disagree 3.No view 4.Agree 5.Strongly agree

2.8 I linked the important ideas in the source texts to what I know already.

1.Strongly disagree 2.Disagree 3.No view 4.Agree 5.Strongly agree

2.9 I worked out how the main ideas across the source texts relate to each other.

1.Strongly disagree 2.Disagree 3.No view 4.Agree 5.Strongly agree

2.10 I developed new ideas or a better understanding of existing knowledge.

1.Strongly disagree 2.Disagree 3.No view 4.Agree 5.Strongly agree

2.11 I changed my writing plan (e.g. structure, content etc.).

1.Strongly disagree 2.Disagree 3.No view 4.Agree 5.Strongly agree

What else did you do while reading the source texts?

III. Before writing

3.1 I organised the ideas for my text before starting to write.

1.Strongly disagree 2.Disagree 3.No view 4.Agree 5.Strongly agree

3.2 I recombined or reordered the ideas to fit the structure of my essay.

1.Strongly disagree 2.Disagree 3.No view 4.Agree 5.Strongly agree

3.3 I removed some ideas I planned to write.

1.Strongly disagree 2.Disagree 3.No view 4.Agree 5.Strongly agree

What else did you do before writing?

IV. While writing the first draft

4.1 While I was writing, I sometimes paused to organise my ideas.

1.Strongly disagree 2.Disagree 3.No view 4.Agree 5.Strongly agree

4.2 I developed new ideas while I was writing.

1.Strongly disagree 2.Disagree 3.No view 4.Agree 5.Strongly agree

4.3 I re-read the task prompt.

1.Strongly disagree 2.Disagree 3.No view 4.Agree 5.Strongly agree

4.4 I selectively re-read the source texts.

1.Strongly disagree 2.Disagree 3.No view 4.Agree 5.Strongly agree

4.5 I changed my writing plan (e.g. structure, content etc.).

1.Strongly disagree 2.Disagree 3.No view 4.Agree 5.Strongly agree

4.6 I checked that the content was relevant.

1.Strongly disagree 2.Disagree 3.No view 4.Agree 5.Strongly agree

4.7 I checked that the text was well-organised.

1.Strongly disagree 2.Disagree 3.No view 4.Agree 5.Strongly agree

4.8 I checked that the text was coherent.

1.Strongly disagree 2.Disagree 3.No view 4.Agree 5.Strongly agree

4.9 I checked that I included all appropriate main ideas from all the source texts.

1.Strongly disagree 2.Disagree 3.No view 4.Agree 5.Strongly agree

4.10 I checked that I included my own viewpoint on the topic.

1.Strongly disagree 2.Disagree 3.No view 4.Agree 5.Strongly agree

4.11 I checked that I had put the ideas of the source texts into my own words.

1.Strongly disagree 2.Disagree 3.No view 4.Agree 5.Strongly agree

4.12 I checked the possible effect of my writing on the intended reader.

1.Strongly disagree 2.Disagree 3.No view 4.Agree 5.Strongly agree

4.13 I checked the accuracy and range of the sentence structures.

1.Strongly disagree 2.Disagree 3.No view 4.Agree 5.Strongly agree

4.14 I checked the appropriateness and range of vocabulary.

1.Strongly disagree 2.Disagree 3.No view 4.Agree 5.Strongly agree

What else did you do while writing the first draft?

V. After writing the first draft

5.1 I checked that the content was relevant.

1.Strongly disagree 2.Disagree 3.No view 4.Agree 5.Strongly agree

5.2 I checked that the text was well-organised.

1.Strongly disagree 2.Disagree 3.No view 4.Agree 5.Strongly agree

5.3 I checked that the text was coherent.

1.Strongly disagree 2.Disagree 3.No view 4.Agree 5.Strongly agree

5.4 I checked that I included all appropriate main ideas from all the source texts.

1.Strongly disagree 2.Disagree 3.No view 4.Agree 5.Strongly agree

5.5 I checked that I included my own viewpoint on the topic.

1.Strongly disagree 2.Disagree 3.No view 4.Agree 5.Strongly agree

5.6 I checked that I had put the ideas of the source texts into my own words.

1.Strongly disagree 2.Disagree 3.No view 4.Agree 5.Strongly agree

5.7 I checked the possible effect of my writing on the intended reader.

1.Strongly disagree 2.Disagree 3.No view 4.Agree 5.Strongly agree

5.8 I checked the accuracy and range of the sentence structures.

1.Strongly disagree 2.Disagree 3.No view 4.Agree 5.Strongly agree

5.9 I checked the appropriateness and range of vocabulary.

1.Strongly disagree 2.Disagree 3.No view 4.Agree 5.Strongly agree

What else did you do after writing the first draft?

The end

APPENDIX F

1. Item 1.1
I read the task prompt (i.e. instructions) carefully to understand each word in it.

2. Item 1.2
I thought of what I might need to write to make my essay relevant and adequate to the task.

3. Item 1.3
I thought of how my essay would suit the expectations of the intended reader.

4. Item 3.1
I organised the ideas I plan to include in my essay.

5. Item 4.7
I checked that the essay was well-organised.

6. Item 4.8
I checked that the essay was coherent, e.g. appropriate use of topic sentences, connectives, etc.

7. Item 4.13
I checked the grammatical accuracy and range of the sentence structures (while writing the first draft).

8. Item 4.14
I checked the spelling, usage and range of the vocabulary (while writing the first draft).

9. Item 5.2
I checked that the essay was well-organised (while writing the first draft).

10. Item 5.3
I checked that the essay was coherent, e.g. appropriate use of topic sentences, connectives, etc. (after writing the first draft)

11. Item 5.8
I checked the grammatical accuracy and range of the sentence structures (after writing the first draft).

12. Item 5.9
I checked the spelling, usage and range of the vocabulary (after writing the first draft).

APPENDIX G

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Q1	Equal variances assumed	1.194	.281	2.512	39	.016	.695	.277	.135	1.255
	Equal variances not assumed			2.546	31.333	.016	.695	.273	.138	1.252
Q2	Equal variances assumed	.018	.894	2.184	39	.035	.598	.274	.044	1.151
	Equal variances not assumed			2.198	37.571	.034	.598	.272	.047	1.148
Q3	Equal variances assumed	.735	.396	5.119	39	.000	1.260	.246	.762	1.757
	Equal variances not assumed			5.149	37.680	.000	1.260	.245	.764	1.755
Q4	Equal variances assumed	.083	.775	2.494	39	.017	.498	.200	.094	.901
	Equal variances not assumed			2.516	35.617	.017	.498	.198	.096	.899
Q5	Equal variances assumed	1.046	.313	4.980	39	.000	1.200	.241	.713	1.687
	Equal variances not assumed			4.959	37.317	.000	1.200	.242	.710	1.690
Q6 (Item 2.1)	Equal variances assumed	6.990	.012	1.938	39	.060	.586	.302	-.026	1.197
	Equal variances not assumed			1.964	31.469	.058	.586	.298	-.022	1.194
Q7	Equal variances assumed	4.860	.033	2.492	39	.017	.852	.342	.160	1.544
	Equal variances not assumed			2.505	37.964	.017	.852	.340	.164	1.541

Q8	Equal variances assumed	.513	.478	3.094	39	.004	.774	.250	.268	1.280
	Equal variances not assumed			3.109	38.187	.004	.774	.249	.270	1.278
Q9	Equal variances assumed	.429	.516	4.686	39	.000	.938	.200	.533	1.343
	Equal variances not assumed			4.710	38.121	.000	.938	.199	.535	1.341
Q10	Equal variances assumed	4.619	.038	2.996	39	.005	.926	.309	.301	1.552
	Equal variances not assumed			3.034	31.970	.005	.926	.305	.304	1.548
Q11	Equal variances assumed	.312	.580	2.480	39	.018	.871	.351	.161	1.582
	Equal variances not assumed			2.479	38.876	.018	.871	.351	.160	1.582
Q12	Equal variances assumed	.240	.627	3.137	39	.003	.686	.219	.244	1.128
	Equal variances not assumed			3.158	37.272	.003	.686	.217	.246	1.126
Q13	Equal variances assumed	.093	.762	3.535	39	.001	.829	.234	.355	1.303
	Equal variances not assumed			3.535	38.860	.001	.829	.234	.354	1.303
Q14	Equal variances assumed	.513	.478	4.611	39	.000	.924	.200	.519	1.329
	Equal variances not assumed			4.625	38.769	.000	.924	.200	.520	1.328
Q15	Equal variances assumed	1.574	.217	4.639	39	.000	1.150	.248	.649	1.651
	Equal variances not assumed			4.615	36.550	.000	1.150	.249	.645	1.655
Q16 (Item 2.11)	Equal variances assumed	6.050	.018	1.544	39	.131	.481	.311	-.149	1.111
	Equal variances not assumed			1.527	31.194	.137	.481	.315	-.161	1.123
Q17	Equal variances assumed	1.135	.293	3.256	39	.002	.643	.197	.244	1.042
	Equal variances not assumed			3.262	38.975	.002	.643	.197	.244	1.041

Q18	Equal variances assumed	2.080	.157	3.835	39	.000	.969	.253	.458	1.480
	Equal variances not assumed			3.850	38.628	.000	.969	.252	.460	1.478
Q19	Equal variances assumed	4.302	.045	2.684	39	.011	.714	.266	.176	1.252
	Equal variances not assumed			2.693	38.772	.010	.714	.265	.178	1.251
Q20	Equal variances assumed	.139	.711	3.625	39	.001	.836	.231	.369	1.302
	Equal variances not assumed			3.620	38.560	.001	.836	.231	.369	1.303
Q21	Equal variances assumed	1.818	.185	2.040	39	.048	.521	.256	.005	1.038
	Equal variances not assumed			2.039	38.819	.048	.521	.256	.004	1.039
Q22	Equal variances assumed	4.841	.034	3.674	39	.001	1.160	.316	.521	1.798
	Equal variances not assumed			3.706	35.787	.001	1.160	.313	.525	1.794
Q23	Equal variances assumed	3.647	.064	2.163	39	.037	.398	.184	.026	.769
	Equal variances not assumed			2.163	38.918	.037	.398	.184	.026	.769
Q24 (Item 4.5)	Equal variances assumed	.970	.331	1.988	39	.054	.567	.285	-.010	1.143
	Equal variances not assumed			1.982	37.886	.055	.567	.286	-.012	1.146
Q25	Equal variances assumed	.028	.867	3.777	39	.001	.786	.208	.365	1.206
	Equal variances not assumed			3.793	38.484	.001	.786	.207	.367	1.205
Q26	Equal variances assumed	1.073	.307	5.893	39	.000	1.393	.236	.915	1.871
	Equal variances not assumed			5.934	36.975	.000	1.393	.235	.917	1.868
Q27	Equal variances assumed	.418	.522	5.297	39	.000	1.148	.217	.709	1.586
	Equal variances not assumed			5.316	38.662	.000	1.148	.216	.711	1.584

Q28	Equal variances assumed	1.066	.308	4.329	39	.000	1.138	.263	.606	1.670
	Equal variances not assumed			4.327	38.780	.000	1.138	.263	.606	1.670
Q29	Equal variances assumed	.450	.506	4.161	39	.000	.979	.235	.503	1.454
	Equal variances not assumed			4.179	38.486	.000	.979	.234	.505	1.452
Q30	Equal variances assumed	4.353	.044	4.680	37	.000	.976	.209	.554	1.399
	Equal variances not assumed			4.632	30.983	.000	.976	.211	.546	1.406
Q31	Equal variances assumed	1.970	.168	3.568	39	.001	.986	.276	.427	1.544
	Equal variances not assumed			3.584	38.395	.001	.986	.275	.429	1.542
Q32	Equal variances assumed	8.718	.005	3.362	39	.002	.957	.285	.381	1.533
	Equal variances not assumed			3.394	35.255	.002	.957	.282	.385	1.530
Q33	Equal variances assumed	.606	.441	4.017	39	.000	1.198	.298	.595	1.801
	Equal variances not assumed			4.022	38.998	.000	1.198	.298	.595	1.800
Q34	Equal variances assumed	.552	.462	2.713	39	.010	.724	.267	.184	1.263
	Equal variances not assumed			2.716	38.999	.010	.724	.266	.185	1.263
Q35	Equal variances assumed	1.590	.215	4.821	39	.000	1.050	.218	.609	1.491
	Equal variances not assumed			4.850	37.588	.000	1.050	.216	.612	1.488
Q36	Equal variances assumed	.000	.983	3.801	39	.000	.950	.250	.445	1.455
	Equal variances not assumed			3.795	38.503	.001	.950	.250	.444	1.456
Q37	Equal variances assumed	9.196	.004	3.853	39	.000	.900	.234	.428	1.372
	Equal variances not assumed			3.896	33.565	.000	.900	.231	.430	1.370

Q38	Equal variances assumed	1.103	.300	3.764	39	.001	.926	.246	.428	1.424
	Equal variances not assumed			3.786	37.680	.001	.926	.245	.431	1.422
Q39	Equal variances assumed	9.307	.004	4.055	39	.000	.917	.226	.459	1.374
	Equal variances not assumed			4.097	34.307	.000	.917	.224	.462	1.371
Q40 (Item 5.7)	Equal variances assumed	.000	.991	1.851	39	.072	.495	.268	-.046	1.037
	Equal variances not assumed			1.854	38.965	.071	.495	.267	-.045	1.035
Q41	Equal variances assumed	.364	.550	4.458	39	.000	1.055	.237	.576	1.533
	Equal variances not assumed			4.462	38.985	.000	1.055	.236	.577	1.533
Q42	Equal variances assumed	.931	.341	4.313	39	.000	1.107	.257	.588	1.626
	Equal variances not assumed			4.322	38.960	.000	1.107	.256	.589	1.625

APPENDIX H

读写结合写作过程问卷

姓名: ....　　性别:　☐ 男　☐ 女　☐ 不方便透露

专四成绩: ....　☐ 无成绩

雅思成绩: 总分 ....　阅读 ....　写作 ....　☐ 无成绩

说明：

下面的问卷是对读写结合写作过程基本情况的调查。请你根据自己的实际写作情况，回忆在以下各个阶段

- **阅读试题导语时**
- **阅读素材时**
- **开始写作前**
- **写作初稿时**
- **完成初稿后**　　　你都做了些什么。

请根据每个数字所代表的含义选出最能代表你的看法的一个数，在这个数上画圈。

1=非常不同意　2=不同意　3=不确定　4=同意　5=非常同意

举例：

我觉得这篇作文的难度比较大。　1　2　3　4　5

每道题的选项均无对错好坏之分，你的回答也不会影响你的写作成绩，所收集的数据只用作学术研究。问卷中涉及到的所有信息都会严格保密。谢谢合作。

. . . .

1. 阅读试题导语时 请回忆你在阅读试题导语时做了些什么。		非常不同意	不同意	不确定	同意	非常同意
1.1	我仔细阅读了导语，理解其中每个单词的意思。	1	2	3	4	5
1.2	我考虑了需要写什么内容使文章切合题意，且内容充实。	1	2	3	4	5
1.3	我完全理解题目的要求。	1	2	3	4	5

阅读导语时，你还做了些什么？						
2. 阅读素材时 请回忆你在阅读素材时做了些什么。		非常不同意	不同意	不确定	同意	非常同意
2.1	我仔细阅读了每个材料。	1	2	3	4	5
2.2	我快速搜寻了材料中我需要的观点。	1	2	3	4	5
2.3	我仔细阅读了材料中对写作有帮助的相关部分。	1	2	3	4	5
2.4	阅读材料启发了我关于文章主题的一些想法。	1	2	3	4	5
2.5	阅读材料帮助我确定了立场。	1	2	3	4	5
2.6	我回去读了试题导语部分。	1	2	3	4	5
2.7	我把材料中的主要观点按重要性排了序。	1	2	3	4	5
2.8	我把材料中的重要观点和我已有的知识联系起来。	1	2	3	4	5
2.9	我把各个材料中的主要观点联系起来。	1	2	3	4	5
2.10	我产生了新的观点或者对已有的知识了解更深刻了。	1	2	3	4	5
2.11	阅读材料帮助我组织文章结构。	1	2	3	4	5
阅读素材时，你还做了些什么？						
3. 开始写作前 请回忆你在开始写作前做了些什么。		非常不同意	不同意	不确定	同意	非常同意
3.1	我整理了要写的观点。	1	2	3	4	5

3. 2	根据文章的结构，我重新组合或排序要写的观点。	1	2	3	4	5
3. 3	我放弃了一些计划要写的观点。	1	2	3	4	5
开始写作前，你还做了些什么？						

4. 写作初稿时 请回忆你在写作初稿时做了些什么。		非常不同意	不同意	不确定	同意	非常同意
4. 1	我有时会停下来整理观点。	1	2	3	4	5
4. 2	我在文章中用到了材料中的一些观点。	1	2	3	4	5
4. 3	我产生了新的观点。	1	2	3	4	5
4. 4	我又回去读了试题导语部分。	1	2	3	4	5
4. 5	我选择性地重新阅读了一些材料。	1	2	3	4	5
4. 6	我在文章中改述了材料中的一些话。	1	2	3	4	5
4. 7	我在文章中摘抄了材料中的一些短语和句子。	1	2	3	4	5
4. 8	我检查了文章内容是否贴切。	1	2	3	4	5
4. 9	我检查了文章是否结构严谨，条理清晰。	1	2	3	4	5
4. 10	我检查了文章是否连贯通顺，例如，是否运用了主题句、连接词等。	1	2	3	4	5
4. 11	我检查了文章是否包含材料中的主要观点。	1	2	3	4	5
4. 12	我检查了文章是否包含我的个人观点。	1	2	3	4	5
4. 13	我检查了是否有用自己的文字表达材料中的观点。	1	2	3	4	5

4. 14	我检查了句子的结构是否准确，类型是否丰富。	1	2	3	4	5
4. 15	我检查了单词拼写和用法是否准确，类型是否丰富。	1	2	3	4	5
写作初稿时，你还做了些什么？						

5. 完成初稿后 请回忆你在完成初稿后做了些什么。		**非常不同意**	**不同意**	**不确定**	**同意**	**非常同意**
5. 1	我检查了文章内容是否贴切。	1	2	3	4	5
5. 2	我检查了文章是否结构严谨，条理清晰。	1	2	3	4	5
5. 3	我检查了文章是否连贯通顺，例如，是否运用了主题句、连接词等。	1	2	3	4	5
5. 4	我检查了文章是否包含材料中的主要观点。	1	2	3	4	5
5. 5	我检查了文章是否包含我的个人观点。	1	2	3	4	5
5. 6	我检查了是否有用自己的文字表达材料中的观点。	1	2	3	4	5
5. 7	我检查了句子的结构是否准确，类型是否丰富。	1	2	3	4	5
5. 8	我检查了单词拼写和用法是否准确，类型是否丰富。	1	2	3	4	5
完成初稿后，你还做了些什么？						

--The end--

Thank you!

APPENDIX I

Writing Process Questionnaire

Name: .... Gender: □ male □ female

TEM-4 result (if any): ....

IELTS result (if any): Overall band .... Reading .... Writing ....

In this questionnaire, there are some statements about how you might complete the writing task you have just taken. Please answer all the questions, thinking about what you did

- **While reading the directions**
- **While reading the source materials**
- **Before writing**
- **While writing the first draft**
- **After writing the first draft**

in the task taking experience you have just had.

Please circle the extent of your agreement or disagreement to each statement below, using the following 5-point scale,

FOR EXAMPLE

I found academic writing to be easy.
1.Strongly disagree 2.Disagree 3.No view 4.Agree 5.Strongly agree
..

I. While reading the directions

1.1 I read the directions carefully to understand each word in it.

1.Strongly disagree 2.Disagree 3.No view 4.Agree 5.Strongly agree

1.2 I thought of what I might need to write to make my essay relevant and adequate to the task.

1.Strongly disagree 2.Disagree 3.No view 4.Agree 5.Strongly agree

1.3 I was able to understand the directions for this writing test very well.

1.Strongly disagree 2.Disagree 3.No view 4.Agree 5.Strongly agree

What else did you do while reading the directions?

II. While reading the source materials

2.1 I read through the whole of each source material carefully.

1.Strongly disagree 2.Disagree 3.No view 4.Agree 5.Strongly agree

2.2 I searched quickly for the ideas which might help me to write the essay.

1.Strongly disagree 2.Disagree 3.No view 4.Agree 5.Strongly agree

2.3 I read some relevant part(s) of the materials carefully.

1.Strongly disagree 2.Disagree 3.No view 4.Agree 5.Strongly agree

2.4 I used the readings to help me get ideas on the topic.

1.Strongly disagree 2.Disagree 3.No view 4.Agree 5.Strongly agree

5 The readings helped me choose an opinion on the issue.

1.Strongly disagree 2.Disagree 3.No view 4.Agree 5.Strongly agree

2.6 I read the directions again.

1.Strongly disagree 2.Disagree 3.No view 4.Agree 5.Strongly agree

2.7 I prioritised the important ideas in the source materials in my mind.

1.Strongly disagree 2.Disagree 3.No view 4.Agree 5.Strongly agree

2.8 I linked the important ideas in the source materials to what I know already.

1.Strongly disagree 2.Disagree 3.No view 4.Agree 5.Strongly agree

2.9 I worked out how the main ideas across the source materials relate to each other.

1.Strongly disagree 2.Disagree 3.No view 4.Agree 5.Strongly agree

2.10 I developed new ideas or a better understanding of existing knowledge.

1.Strongly disagree 2.Disagree 3.No view 4.Agree 5.Strongly agree

2.11 I used the readings to help me organise my essay.

1.Strongly disagree 2.Disagree 3.No view 4.Agree 5.Strongly agree

What else did you do while reading the source materials?

III. Before writing

3.1 I organised the ideas I plan to include in my essay.

1.Strongly disagree 2.Disagree 3.No view 4.Agree 5.Strongly agree

3.2 I recombined or reordered the ideas to fit the structure of my essay.

1.Strongly disagree 2.Disagree 3.No view 4.Agree 5.Strongly agree

3.3 I removed some ideas I planned to write.

1.Strongly disagree 2.Disagree 3.No view 4.Agree 5.Strongly agree

What else did you do before writing?

IV. While writing the first draft

4.1 While I was writing, I sometimes paused to organise my ideas.

1.Strongly disagree 2.Disagree 3.No view 4.Agree 5.Strongly agree

4.2 I used some of the ideas from the readings in my essay.

1.Strongly disagree 2.Disagree 3.No view 4.Agree 5.Strongly agree

4.3 I developed new ideas while I was writing.

1.Strongly disagree 2.Disagree 3.No view 4.Agree 5.Strongly agree

4.4 I re-read the directions.

1.Strongly disagree 2.Disagree 3.No view 4.Agree 5.Strongly agree

4.5 I selectively re-read the source materials.

1.Strongly disagree 2.Disagree 3.No view 4.Agree 5.Strongly agree

6 I paraphrased the reading in my writing.

1.Strongly disagree 2.Disagree 3.No view 4.Agree 5.Strongly agree

7 I copied phrases and sentences directly from the reading into my essay.

1.Strongly disagree 2.Disagree 3.No view 4.Agree 5.Strongly agree

4.8 I checked that the content was relevant.

1.Strongly disagree 2.Disagree 3.No view 4.Agree 5.Strongly agree

4.9 I checked that the essay was well-organised.

1.Strongly disagree 2.Disagree 3.No view 4.Agree 5.Strongly agree

4.10 I checked that the essay was coherent, e.g. appropriate use of topic sentences, connectives, etc.

1.Strongly disagree 2.Disagree 3.No view 4.Agree 5.Strongly agree

4.11 I checked that I included all appropriate main ideas from all the source materials.

1.Strongly disagree 2.Disagree 3.No view 4.Agree 5.Strongly agree

4.12 I checked that I included my own viewpoint on the topic.

1.Strongly disagree 2.Disagree 3.No view 4.Agree 5.Strongly agree

4.13 I checked that I had put the ideas of the source materials into my own words.

1.Strongly disagree 2.Disagree 3.No view 4.Agree 5.Strongly agree

4.14 I checked the grammatical accuracy and range of the sentence structures.

1.Strongly disagree 2.Disagree 3.No view 4.Agree 5.Strongly agree

4.15 I checked the spelling, usage and range of the vocabulary.

1.Strongly disagree 2.Disagree 3.No view 4.Agree 5.Strongly agree

What else did you do while writing the first draft?

V. After writing the first draft

5.1 I checked that the content was relevant.

1.Strongly disagree 2.Disagree 3.No view 4.Agree 5.Strongly agree

5.2 I checked that the essay was well-organised.

1.Strongly disagree 2.Disagree 3.No view 4.Agree 5.Strongly agree

5.3 I checked that the essay was coherent, e.g. appropriate use of topic sentences, connectives, etc.

1.Strongly disagree 2.Disagree 3.No view 4.Agree 5.Strongly agree

5.4 I checked that I included all appropriate main ideas from all the source materials.

1.Strongly disagree 2.Disagree 3.No view 4.Agree 5.Strongly agree

5.5 I checked that I included my own viewpoint on the topic.

1.Strongly disagree 2.Disagree 3.No view 4.Agree 5.Strongly agree

5.6 I checked that I had put the ideas of the source materials into my own words.

1.Strongly disagree 2.Disagree 3.No view 4.Agree 5.Strongly agree

5.7 I checked the grammatical accuracy and range of the sentence structures.

1.Strongly disagree 2.Disagree 3.No view 4.Agree 5.Strongly agree

5.8 I checked the spelling, usage and range of the vocabulary.

1.Strongly disagree 2.Disagree 3.No view 4.Agree 5.Strongly agree

What else did you do after writing the first draft?

--The end--

Thank you!

APPENDIX J

Test Statistics[a]

	Mann-Whitney U	Wilcoxon W	Z	Asymp. Sig. (2-tailed)
Q1	3039.000	6525.000	-1.674	.094
Q2	2992.500	6647.500	-2.104	.035
Q3	3113.500	6768.500	-1.702	.089
Q4	2947.000	6433.000	-2.019	.043
Q5	3086.500	6656.500	-1.762	.078
Q6	3363.500	7018.500	-.732	.464
Q7	3581.000	7236.000	-.106	.916
Q8	3243.500	6898.500	-1.133	.257
Q9	3139.500	6794.500	-1.655	.098
Q10	3299.000	6954.000	-1.027	.305
Q11	3428.500	6998.500	-.346	.730
Q12	2925.500	6580.500	-2.413	.016
Q13	3484.000	7139.000	-.426	.670
Q14	3030.500	6685.500	-1.967	.049
Q15	2927.500	6413.500	-2.048	.041
Q16	3485.500	6971.500	-.149	.882
Q17	2844.500	6414.500	-2.020	.043
Q18	3411.000	7066.000	-.746	.455
Q19	3525.500	7180.500	-.306	.759
Q20	3401.500	7056.500	-.558	.577
Q21	3536.500	7106.500	-.114	.909
Q22	3447.500	7102.500	-.578	.563
Q23	3567.500	7222.500	-.155	.877
Q24	3351.500	7006.500	-.740	.459
Q25	3019.000	6589.000	-1.765	.078
Q26	3431.500	7086.500	-.606	.545
Q27	3489.000	7144.000	-.271	.787
Q28	3390.500	7045.500	-.786	.432
Q29	3502.000	7072.000	-.241	.810
Q30	3252.500	6822.500	-1.127	.260
Q31	3502.000	7072.000	-.087	.931
Q32	3213.500	6868.500	-1.336	.182
Q33	3212.000	6867.000	-1.131	.258
Q34	3459.500	6945.500	-.234	.815
Q35	3474.000	7129.000	-.038	.969

Q36	3346.000	7001.000	-.648	.517
Q37	3467.000	7037.000	-.072	.943
Q38	3155.000	6810.000	-1.035	.301
Q39	3268.500	6923.500	-.901	.368
Q40	3317.500	6803.500	-.722	.470