Scientific Research Publishing

Open Journal of Social Sciences, 2022, 10, 442-457
https://www.scirp.org/journal/jss
ISSN Online: 2327-5960
ISSN Print: 2327-5952

Investigating EFL Test-Takers' Engagement with Source Materials in Reading-to-Write: Evidence from Eye-Tracking

Pucheng Wang

School of Applied Foreign Languages, Zhejiang International Studies University, Hangzhou, China
Email: p.wang@zisu.edu.cn

How to cite this paper: Wang, P. C. (2022). Investigating EFL Test-Takers' Engagement with Source Materials in Reading-to-Write: Evidence from Eye-Tracking. *Open Journal of Social Sciences, 10,* 442-457.
https://doi.org/10.4236/jss.2022.104032

Received: March 21, 2022
Accepted: April 24, 2022
Published: April 27, 2022

Abstract

This study investigated how English as Foreign Language (EFL) test-takers were engaged with source materials while completing a reading-to-write task. 16 participants were eye-tracked throughout task completion. Two eye-tracking metrics were examined: total visit duration, which reports how long the participants looked at each part of the task; and individual visit duration, which calculates statistics such as the mean visit duration and max visit duration of the participants' overall individual visits. Results showed that the participants spent over a quarter of their time in reading whilst reading-to-write. During task completion, they constantly switched their attention between reading source materials and writing, and they tended to adopt a more expeditious reading approach when engaging with the source text.

Keywords

EFL Writing, Reading-to-Write Tasks, Source Materials, Eye-Tracking, Careful Reading, Expeditious Reading

1. Introduction

In a typical reading-to-write task, test-takers are provided with source materials. They are required to comprehend these sources, 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, reading-to-write tasks may better contextualise writing activities, thus enhancing the connection between test-takers' performance and real language use.

A significant amount of research on reading-to-write 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 (Cumming et al., 2005; Gebril & Plakans, 2009; 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. They 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 second language (L2) writing research (Johns & Mayes, 1990; Shi, 2004). In an early study on verbatim source use, Johns and Mayes (1990) examined direct copying in 80 writing responses of higher- and lower-proficiency groups of writers, finding 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 used more combination of idea units from the source texts and were likely to distort some of these ideas. Shi (2004) compared the written products of two types of writing tasks (opinion and summary) 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.

These studies have provided a solid foundation of understanding source use in reading-to-write tasks. It is clear that source use may vary across proficiency levels, 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, and very few studies (Wang & Zhang, 2021) looked at test-takers' online source use processes.

Therefore, in order to better understand how writers read and use source text in reading-to-write, there is a need to conduct experiments using some online investigation methods. This study used eye-tracking technique to look into 16 EFL writers' engagement with source materials while completing a typical reading-to-write task. Detailed information of the participants, eye-tracker, and pro-

cedures for data collection and analysis will be presented in the following sections.

2. Literature Review

The theoretical underpinning to eye-tracking technique is that our eye movements can be used to make inferences about our cognitive processes (Peyrichoux & 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 analyze 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, & 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, & 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 technique to add more qualitative information to the data. They are commonly used in second and foreign language testing research (Ascensión, 2005, 2008; 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). Concurrent think-aloud (CTA) method allows a participant to verbalise their thoughts during task completion, while 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 & 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 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 & 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, 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.

The combination of 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, & Lee, 2007), for example, they may fixate on certain areas of the screen while verbalising their cognitive processes. RTA method is more appropriate to be used in process studies (particularly when participants have to perform tasks which require high cognitive demand, e.g., a 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 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 & Mackey, 2000: p. 11). 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 & Mackey, 2000: p. 11). The stimulated recall method has proven to be able to get more detailed information from participants, 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 & McCray, 2015; McCray & 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 eye-tracking technique and two types of think-aloud methods, it was decided to combine eye-tracking and stimulated recall method to obtain data on test-takers' cognitive processes while completing a reading-to-write task. 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.

3. Methods

3.1. Participants

20 participants took part in this study. They were all master's students studying

at a UK university. All of them were 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.

The 20 participants were invited to book a slot for taking part in the experiment on 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 may 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. Two of these 18 participants' data were excluded due to insufficient accuracy (weighted gaze samples < 50%; 50% means that at least one eye was observed throughout the entire recording). Thus, the final data set included eye traces from 16 participants: 11 females (69%) and 5 males (31%); the ages of these participants ranged from 21 to 28 yeas old (Mode = 23; Mean = 22.6; SD = 1.66). Table 1 summarizes their IELTS test overall scores, as well as their performance on the Reading and Writing components.

3.2. Equipment and Instrument

3.2.1. 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 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

Table 1. 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

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 1 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 1, 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 1). 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.

3.2.2. TBEM-8 Reading-to-Write Task

One sample task of the Test for Business English Majors-Band 8 (TBEM-8) reading-to-write task (developed in China) was used. This sample task was considered a typical integrated writing task that engages test-takers with both reading and writing skills, as well as a prototype task on which other TBEM-8 writing tasks could be built and is therefore fundamentally indicative of the future tasks which would be developed. One notable feature of the task is that it includes both English and Chinese source materials; test-takers are required to

Figure 1. Layouts of the eye-tracking lab during the experiment.

read and integrate the information in these sources into an essay on a business-related topic (see below for more details about this task). Although the task has been developed for several years, there is not much 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.

The topic of the task concerns Steve Jobs' resignation from Apple. The task contains a set of instructions, and five source materials in the prompt. Source 1 (213 words) is a short passage in Chinese, which gives some background information of Steve Jobs and Apple; Source 2 (120 words) is a collection of English material including several video news headlines and two short excerpts from some internet news, all of which are on Steve Jobs' resignation; Source 3 (275 words) is another set of material in Chinese, and contains three short excerpts from some Chinese newspaper articles, which provide different views on Steve Job's resignation; Source 4, unlike other text materials, is 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 provides test-takers a list of ten words and expressions for reference while completing the task.

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. 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 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 13 px/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. Procedures for Data Collection

During the 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. Due to the constraints of the data collection methods chosen, the data were collected from one participant at a time.

The eye-tracking session started with 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 al-

lows 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). Following successful calibrations, the participant proceeded to complete the TBEM-8 reading-to-write task, which was presented on the eye-tracker screen. The participant's eye movements were simultaneously recorded as they completed the task. A stimulated recall session was also conducted after these participants completed the task so as to get some explanatory data of their eye movements during task completion.

3.4. Data Analyses

In order to answer the research question, that is, to what extent do test-takers engage with the source materials in the TBEM-8 reading-to-write task, two eye-tracking metrics were investigated.

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 2**), 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 analyze fixation data within each individual area. Below are the two eye-tracking metrics examined in this study:

1) Total visit duration, which measures the duration of all visits within an AOI.

2) Individual visit duration, which measures the duration of each individual visit within an AOI.

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.

The two 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 the research question: 1) total visit duration reports how long the participant spent looking at each part of the task; 2)

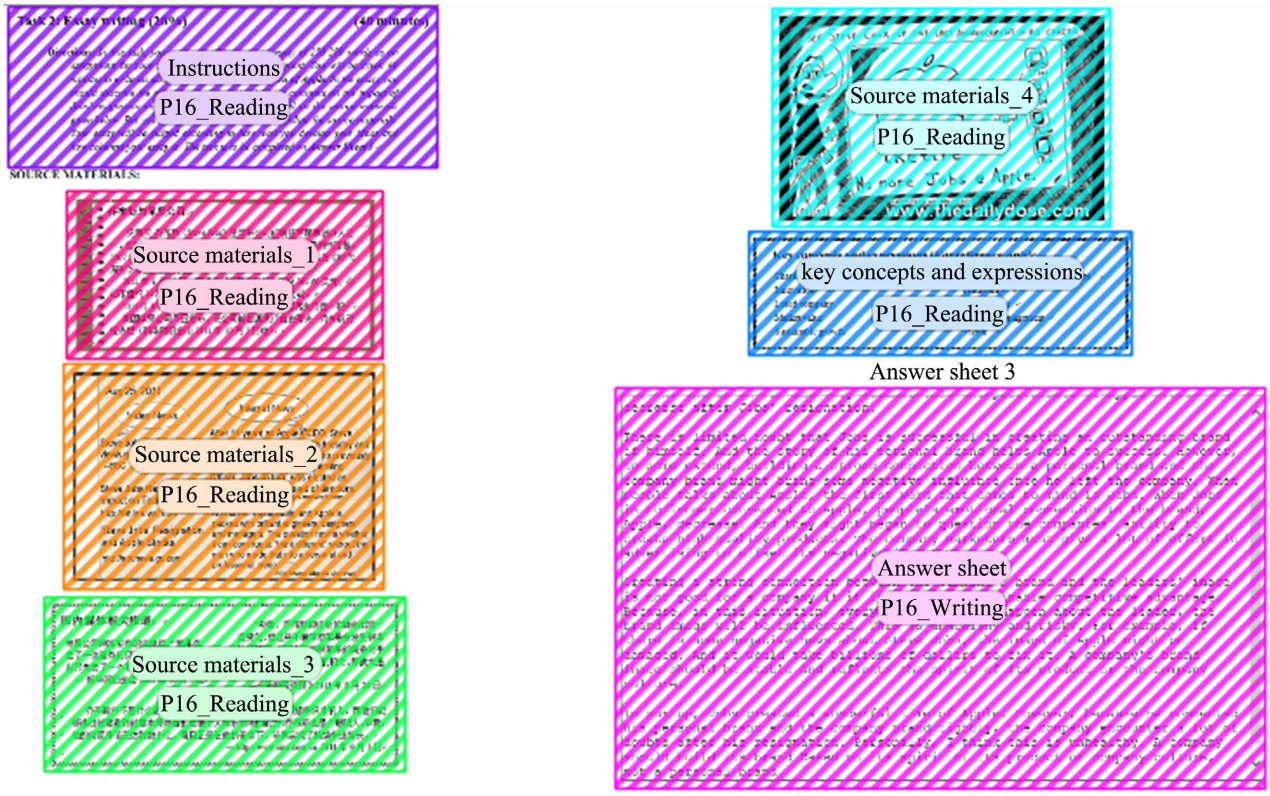

Figure 2. Areas of interest on the eye-tracker screen.

individual 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.

4. Results and Discussion

4.1. Total Visit Duration

Table 2 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 2. Total visit duration on each individual AOI in the reading group by participant.

Areas of interest	Participant																Mean	SD	Min	Max
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16				
	Sec %	Sec %	Sec %	Sec %	Sec %	Sec %	Sec %	Sec %	Sec %	Sec %	Sec %	Sec %	Sec %	Sec %	Sec %	Sec %	Sec %		Sec	Sec
Instructions	131.5	67.0	142.7	113.4	153.9	107.6	73.8	120.5	73.8	94.8	128.7	188.0	174.8	108.1	93.6	59.7	114.5	37.8	59.7	188.0
	26.8	11.5	22.6	19.3	22.4	16.1	18.5	25.1	11.2	22.4	20.5	32.1	25.2	17.6	11.9	16.1	19.7			
Source 1	37.0	124.3	186.2	174.9	129.7	93.3	68.8	59.4	234.2	93.5	189.6	79.9	166.8	100.6	135.9	60.3	120.9	56.5	37.0	234.2
	7.6	21.4	29.5	29.8	18.8	13.9	17.2	12.4	35.5	22.0	30.2	13.6	24.0	16.4	17.2	16.3	20.8			
Source 2	122.0	292.8	119.1	109.5	151.6	292.5	143.1	113.0	178.3	96.0	138.9	85.2	203.4	130.1	296.1	51.3	157.7	76.1	51.3	296.1
	24.9	50.4	18.9	18.7	22.0	43.7	35.8	23.6	27.0	22.6	22.1	14.5	29.3	21.2	37.6	13.9	27.2			
Source 3	83.0	82.5	65.9	99.9	134.9	108.0	53.4	73.5	84.3	60.6	101.0	75.1	66.1	78.4	114.7	62.7	84.0	22.3	53.4	134.9
	16.9	14.2	10.4	17.0	19.6	16.1	13.4	15.3	12.8	14.3	16.1	12.8	9.5	12.8	14.5	16.9	14.5			
Source 4 (picture)	6.1	8.0	1.8	42.7	35.3	3.7	18.4	11.4	6.4	10.0	13.5	81.0	5.2	9.4	20.6	27.9	18.8	20.3	1.8	81.0
	1.2	1.4	0.3	7.3	5.1	0.6	4.6	2.4	1.0	2.4	2.1	13.8	0.8	1.5	2.6	7.6	3.2			
Source 5	110.5	6.6	115.5	46.0	83.1	63.8	42.0	101.6	83.1	69.2	56.9	77.1	78.6	188.3	127.7	108.0	84.9	41.6	6.6	188.3
	22.5	1.1	18.3	7.8	12.1	9.5	10.5	21.2	12.6	16.3	9.0	13.1	11.3	30.6	16.2	29.2	14.6			
Total	490.2	581.1	631.2	586.5	688.6	668.8	399.5	479.3	660.1	424.1	628.6	586.3	695.0	614.9	788.6	369.9	580.8	117.8	369.9	788.6
	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100			

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, 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 in each AOI was controlled by dividing the total visit duration by the total number of words in each part of the task. Results are shown in Table 3 (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 ($\chi 2 = 56.68$, df = 4, $p < 0.001$). The Mann-Whitney tests (see Table 4) were also conducted as post-hoc

Table 3. 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 - **0.98**	44.22
Source 1	120.91	213	120.91/213 - **0.57**	26.72
Source 2	157.68	120	157.68/120 - **1.31**	50.78
Source 3	84.00	275	84.00/275 - **0.31**	12.16
Source 5	84.87	24	84.87/24 - **3.54**	68.63

Table 4. Significant differences in *total visit duration* between different AOIs.

Comparisons	Mann-Whitney U	Z	*p*	Effect size
Instructions - Source 1	43.000	−3.224	0.001	0.32
Instructions - Source 2	87.000	−1.551	0.121	0.08
Instructions - Source 3	0.500	−4.838	0.000	0.73
Instructions - Source 5	16.000	−4.225	0.000	0.56
Source 1 - Source 2	26.000	−3.857	0.000	0.46
Source 1 - Source 3	47.000	−3.116	0.002	0.30
Source 1 - Source 5	13.500	−4.324	0.000	0.58
Source 2 - Source 3	3.000	−4.746	0.000	0.70
Source 2 - Source 5	24.500	−3.904	0.000	0.48
Source 3 - Source 5	8.000	−4.559	0.000	0.65

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 ($p = 0.121$).

Table 5 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 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*

Table 5. Total visit duration on the AOI reading group and writing group by participant.

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

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 2) 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, were looking for mechanical support, such as the spelling of a particular word, for example, she recalled "*I was looking for the word 'resignation*'".

4.2. Individual Visit Duration

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

Table 6 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 between different source materials or

Table 6. Descriptive statistics for visit duration on each individual AOI by participant.

Areas of interest		Participant																Mean
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	
		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	0.1	0.1	0.2	0.1	0.1	0.1	0.2	0.1	0.2	0.2	0.2	0.1	0.3	0.1	0.1	0.1	
	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	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.0	0.1	0.1	
	Max	3.2	5.9	10.9	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	0.1	0.1	0.1	0.1	0.2	0.1	0.1	0.1	0.1	0.2	0.2	0.1	0.1	0.1	0.1	0.2	
	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	0.2	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.3	0.1	0.1	0.1	0.1	0.1	0.1	0.1	
	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	
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	

Continued

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	

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 were 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 are less than three seconds, meaning that half of the visits lasted less than three seconds. This again provided evidence for participants' looking behaviour that they constantly and frequently went to look at the instructions and source materials while writing.

5. Conclusion

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. 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 indicates that test-takers frequently looked for either lexical support or ideas to be produced in the text while completing the task.

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 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.

With the above findings, EFL instructors can better design their lessons and adjust their teaching plans with regard to the source use while reading-to-write, for example, they could devise classroom teaching activities that aim to help learners to be aware of the importance of structuring the source texts and extracting relevant information beforehand, so that the students may avoid unnecessary switches between different parts of the task, and improve their performance in this type of task. Also, this study, benefiting from the eye-tracking technique, provided quantitative data on participants' looking behaviour while completing an integrated reading-to-write task. Unlike previous studies that investigated test-takers' use of the source materials through examining the written products, this study is perhaps one of the few studies that looks at test-takers' real-time source use during task completion, thus gaining further insights into EFL writers' engagement with source materials in completing this type of task.

Conflicts of Interest

The author declares no conflicts of interest regarding the publication of this paper.

References

Ascensión Delaney, Y. (2005). *Validation of Reading-to-Write Assessment Tasks Performed by Second Language Learners*. Unpublished Doctoral Dissertation, Northern Arizona University.

Ascensión Delaney, Y. (2008). Investigating the Reading-to-Write Construct. *Journal of English for Academic Purposes*, 7, 140-150. https://doi.org/10.1016/j.jeap.2008.04.001

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. The British Council.

Brunfaut, T., & McCray, G. (2015). *Looking into Test-Takers' Cognitive Processes Whilst Completing Reading Tasks: A Mixed-Method Eye-Tracking and Stimulated Recall Study*. The British Council (ARAGs Research Reports Online; vol. AR/2015/001).

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* (pp. 317-335). Springer-Verlag Ltd. https://doi.org/10.1007/978-1-4471-0105-5_19

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,* 5-43. https://doi.org/10.1016/j.asw.2005.02.001

Gass, S. M., & Mackey, A. (2000). *Stimulated Recall Methodology in Second Language Research*. Routledge.

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.

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 University Press.

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,* 1-18.

Johns, A. M., & Mayes, P. (1990). An Analysis of Summary Protocols of University ESL Students. *Applied Linguistics, 11,* 253-271. https://doi.org/10.1093/applin/11.3.253

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. In N. Aykin (Ed.), *Usability and Internationalization. HCI and Culture* (pp. 113-121). Springer. https://doi.org/10.1007/978-3-540-73287-7_15

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, 35-57.

McCray, G., & Brunfaut, T. (2018). Investigating the Construct Measured by Banked Gap-Fill Items: Evidence from Eye-Tracking. *Language Testing, 35,* 51-73. https://doi.org/10.1177/0265532216677105

Peyrichoux, I., & Robillard-Bastien, A. (2006). Maximize Usability Testing Benefits with Eye Tracking. SIGCHI Conference Paper.

Shi, L. (2004). Textual Borrowing in Second-Language Writing. *Written Communication, 21,* 171-200. https://doi.org/10.1177/0741088303262846

Wang, P. C., & Zhang, Z. G. (2021). Constructive Processes in Completing Reading-to-Write Tasks: Selecting, Organising and Connecting. *Open Journal of Modern Linguistics, 11,* 919-930. https://doi.org/10.4236/ojml.2021.116071

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.

Yoshida, M. (2007). *The Effects of Task, Text, and Proficiency on Second Language Reading*. Unpublished Doctoral Dissertation, Temple University.

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, The British Council.

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). IDP: IELTS Australia, British Council.