

SHORT REPORT

Open Access



Using keystroke logging to understand writers' processes on a reading-into-writing test

Sathena Chan

Correspondence:
sathena.chan@beds.ac.uk
Centre for Research in English
Language Learning and Assessment
(CRELLA), University of Bedfordshire,
Luton, UK

Abstract

Background: Integrated reading-into-writing tasks are increasingly used in large-scale language proficiency tests. Such tasks are said to possess higher authenticity as they reflect real-life writing conditions better than independent, writing-only tasks. However, to effectively define the reading-into-writing construct, more empirical evidence regarding how writers compose from sources both in real-life and under test conditions is urgently needed. Most previous process studies used think aloud or questionnaire to collect evidence. These methods rely on participants' perceptions of their processes, as well as their ability to report them.

Findings: This paper reports on a small-scale experimental study to explore writers' processes on a reading-into-writing test by employing keystroke logging. Two L2 postgraduates completed an argumentative essay on computer. Their text production processes were captured by a keystroke logging programme. Students were also interviewed to provide additional information. Keystroke logging like most computing tools provides a range of measures. The study examined the students' reading-into-writing processes by analysing a selection of the keystroke logging measures in conjunction with students' final texts and interview protocols.

Conclusions: The results suggest that the nature of the writers' reading-into-writing processes might have a major influence on the writer's final performance. Recommendations for future process studies are provided.

Background

Evidence from recent research has suggested that integrated reading-into-writing tasks, or writing from sources, tap into a unique set of literacy skills which go beyond those normally required by independent writing tasks (Chan, Wu & Weir, 2014; Grabe, 2003; Weir, Vidakovic & Galaczi, 2013). If reading-into-writing is a skill that differs from reading or writing in isolation, we need to model the processes involved. Integrated test tasks which require students to transform knowledge from sources are believed to represent a more authentic performance condition and reflect academic literacy requirements, than those independent writing tasks which require students to express their opinions based on previous knowledge (Cumming, 2013; Plakans, 2009). As a result, it is believed that when engaged in reading-into-writing activities, students are more likely to adopt a knowledge transformation than knowledge telling approach to writing.

Knowledge transformation (Scardamalia & Bereiter, 1987), as compared to knowledge telling, requires a generation of new representations based on connecting the relationship between existing facts and ideas. The knowledge transformation writing approach involves high-level processes such as task representation, discourse synthesis and revision which are likely to be different from those processes involved when writers write upon memory. The decision of which writing approach to employ is largely influenced by many other writer characteristics including task type, writers' L1/L2 writing proficiency and L2 linguistic knowledge. In particular, task type is found to have a significant impact on writers' choice of writing approach (Severinson Eklundh & Kollberg, 2003). A number of studies concluded that reading-into-writing test tasks offer advantages over writing-only tasks in encouraging a knowledge transformation approach on the part of the writer (Plakans, 2009, 2010; Plakans & Gebril, 2012, 2013). To effectively define the reading-into-writing construct, more empirical evidence regarding how writers compose from sources both in real-life and under test conditions is urgently needed.

Most previous process studies used think aloud or questionnaires to collect evidence. Both methods rely on participants' perceptions of their processes as well as their ability to report them. The use of think aloud, however useful, inevitably interrupts writers' processes (Stratman & Hamp-Lyons, 1994). New technology such as keystroke logging might well provide an alternative way to investigating writers' processes. However, like most computing tools, keystroke logging programmes provide multiple analyses and a range of measures in relation to fluency, pausing behaviours and incidence of revisions. There is almost no discussion in the language testing literature regarding how these data might help us to identify and account for test takers' reading-into-writing processes.

This work is a small-scale experimental study to explore the use of keystroke logging supplemented by interview as an alternative method to investigate writers' reading-into-writing processes under test conditions. Although of limited scope, the findings provide insights into how this research method can be used to investigate test takers' cognitive processes, especially in the context of integrated reading-into-writing tests. Discussion is made with regard to the usefulness of different data points in differentiating writers' processes which might lead to different performances.

Literature review

Integrated reading-into-writing tasks

Distinctive from independent tasks, integrated tasks require students to use multiple language skills: reading and writing in this case. In pedagogical context, reading-into-writing refers to those 'instructional tasks that combine reading and writing for various education purposes' (Ascención-Delaney, 2008, p.140). Common reading-into-writing tasks include summary, argumentative essay based on multiple sources, report writing, case study, and literature review. It is more narrowly defined in the field of language testing as 'a test that integrates reading with writing by having examinees read and respond to one or more source texts' (Weigle, 2004, p.30). In the recent decade, more language test providers have incorporated integrated tasks into their writing module. For example, TOEFL iBT by ETS, PTE Academic by Pearson, Integrated Skills of English (ISE) by Trinity College London (see Chan, Inoue & Taylor, 2015), General English Proficiency Test (GEPT) by LTTC in Taiwan, and Test of English for Academic Purposes by EIKEN in Japan.

Reading-into-writing processes

As integrated reading-into-writing tasks become widely used in high-stakes English language tests, the need to establish its cognitive validity is pressing. Cognitive validity concerns the extent to which 'the mental processes that a test elicits from a candidate resemble the processes that s/he would employ in non-test conditions' (Field, 2004, p.7). There is a gap in the existing research because although models of writing have been proposed, the reading-into-writing processes are not well understood (Hirvela, 2004; Plakans, 2009).

The reading-into-writing process is regarded as 'the process of a person who reads a relevant book, an article, a letter, knowing he or she needs to write' (Flower, 1990, p. 6). Some have compared writers' processes on independent writing-only and reading-into-writing integrated tasks, leading to the conclusion that integrated tasks require an additional set of processes which is distinct from those needed to complete writing-only tasks (e.g. Chan, 2013; Plakans, 2010). A small number of studies investigated one of these reading-into-writing sub-processes such as task representation (Wolfersberger, 2013), discourse synthesis (Plakans, 2009) and revising (Barkaoui, 2016).

Task representation is 'an interpretive process that translates the rhetorical situation—as the writer reads it—into an act of composing' (Flower, 1990, p.35). Flower (1990) conducted a think aloud study to investigate four undergraduates' task representation processes on reading-into-writing tasks in an EAP classroom setting. She found that the four students had a different understanding of the same task in terms of primary sources of ideas, features of the text, organisational structure of the text and strategies to use. Students with more academic experience seemed to have a more accurate task representation of the reading-into-writing task. Using semi-structured interviews, Wolfersberger (2013) investigated how four writers' task representation shapes their final product in a classroom-based assessment. The findings indicate that writers' task representation has a noticeable impact on the form and substance of the final products. Some believe that task representation in standardised writing assessment should be stable and shared between test takers and test evaluators (Connor & Carrell, 1993), but findings from these studies, though conducted in EAP classrooms, suggest that writers' ability of creating an accurate task representation is indeed part of the reading-into-writing construct to be assessed.

Discourse synthesis is arguably the core process of the reading-into-writing construct. Spivey and King (1989) defined it as 'a process in which readers (writers) read multiple texts on a topic and synthesize them' (p.11). Their work (Spivey, 1990, 2001; Spivey & King, 1989) has revealed that writers transform a new representation of meaning through three processes: (a) selecting relevant content from sources, (b) organising the content according to the writing goals, and (c) connecting the content from various sources and generating links between these ideas. Using think aloud, Plakans (2009) investigated six L2 writers' discourse synthesis processes on two reading-into-writing argumentative essay tasks. The results revealed significant differences among the writers. High-scoring and low-scoring students transformed ideas from the sources using different strategies. These studies confirmed L2 writers' use of discourse synthesis on reading-into-writing tasks but little has been shown as in how these processes lead to their text.

Combining think aloud data from nine students and questionnaire data from other 136 students, Plakans and Gebril (2012) examined the role of reading comprehension in writers' source use. They found that while some writers may have had less comprehension of the topic, according to self-report, they were able to make use of the texts similarly as students who showed better comprehension. The results indicate that there is a threshold for reading comprehension in writers' ability of source use on integrated tasks. Another study by Plakans and Gebril (2013) investigated features of source use in 480 TOEFL iBT performances by textual analysis. They found that high-scoring texts included important ideas from both sources (i.e. reading and listening texts in the context of their study) whereas low-scoring texts included ideas mainly from the reading texts and contained direct copying of words and phrases. However, as actual evidence of students' processes was not collected, it is not possible to know whether these differences found in the final texts were a result of writers' use of discourse synthesis processes.

Revising is another important sub-process in reading-into-writing. Like task representation and discourse synthesis, it is evident that writers at different proficiency levels employ revising processes differently. Skilled writers are found to make changes in relation to the global content and organisation of their text whereas unskilled writers make changes predominantly related to linguistic accuracy (Field, 2004; Flower, Hayes, Carey, Schriver, & Stratman, 1986; Kellogg, 1996). Previous research also shows that writers' revisions might vary in different conditions, such as L1 and L2 writing (Stevenson, Schoonen & Gloppe, 2006), paper-based versus computer-based (Chan, Bax & Weir, in press).

Using keystroke logging, Severinson Eklundh and Kollberg (2003) investigated the impact of task type on ten third year students' revising processes. The students completed one independent task (narrative) and three integrated tasks (summary, comparison and argumentative). The results revealed that the integrated tasks were more demanding for most students, which resulted in longer pause times, higher revision numbers and lower text production rates (measured in number of words in the final texts per minute of writing time). The results suggested that students revised differently on each of the integrated tasks. Most students revised a great extent regarding structuring and formatting aspects of the text in the summary task but engaged in frequent revising of major content elements in the comparison task. The argumentative task, on the other hand, exhibited a mixed pattern, but the researchers noted that some skilled students were able to write the text without much high-level revision, because they already had a rather well-known schema for the argumentation task. However, they centred on addressing how writers revise to develop the discourse structure of their text. More research is needed to investigate how writers revise in relation to the sources in integrated tasks.

A recent keystroke logging study by Barkaoui (2016) further investigated the impact of task type and second language proficiency on test takers' revision patterns between the TOEFL iBT independent and integrated tasks. Confirming with the previous literature, the results suggested that L2 proficiency played a significant role on test takers' revision patterns. The low proficient group made significantly more overall revisions and precontextual revisions (i.e. revisions made at the point of inscription) than did the high proficient group. Furthermore, the low proficient group made significantly more low-level revisions (i.e. those related to typology, form and language) than did the high

proficient group. The results showed that task type also, but to a less extent, impacted on test takers’ revision patterns. On the integrated task, the participants made comparatively fewer revisions during the first segment of text production. They focused on precontextual revisions during the second segment of text production, and contextual revisions (i.e. those made already written text) during the last segment. However, task type did not have a significant impact on the focus on the test takers’ revisions. Barkaoui’s (2016) study provided useful and timely evidence of the impact of L2 proficiency and task type on writers’ revisions under test conditions through keystroke logging. However, due to a different focus of his study, the analysis did not address much about how test takers revised in response to the distinct features of integrated tasks. More discussions on how keystroke logging data were analysed and interpreted between Barkaoui’s study and this study are provided later.

The previous studies have helped to understanding writers’ processes of task representation, discourse synthesis and revising in integrated tasks. Nevertheless, to define the construct of reading-into-writing for teaching and assessment, a coherent model is needed (Hirvela, 2004). Chan (2013) began to fill this gap by defining the target reading-into-writing construct EAP tests. A reading-into-writing process questionnaire was developed based on previous models of reading and writing (e.g. Hayes & Flower, 1980; Khalifa & Weir 2009; Shaw & Weir, 2007; Spivey, 2001) and trialled in a pilot with 99 students. The validated questionnaire was used to investigate 300 students’ processes on four reading-into-writing tasks under real-life and test conditions.

The results of explanatory factor analysis confirmed the underlying construct of different reading-into-writing sub-processes (see Table 1) (for the working definitions of each cognitive phase, see Additional file 1: Appendix A). Another contribution of this model is that it discerns which of these processes reflected differences between high- and low-scoring writers. The results show that high-scoring students reported more use of most of the processes (including task representation, careful reading, search reading, connecting and generating ideas, organising intertextual relationship between ideas, organising ideas in a textual structure, and monitoring and revising during text production at both low and high levels) than low-scoring students. However, due to the limitation of research tools, there was no further qualitative evidence

Table 1 Cognitive validity parameters in writing from sources (adapted from Chan, 2013)

Cognitive phases	Key cognitive processes
Conceptualisation	Task representation
	Macro-planning
Meaning construction	Careful reading
	Scanning, Skimming and Search reading
	Connecting ideas and generating new representations
Generating texts (execution)	Translating ideas into linguistic forms
	Micro-planning
Organising	Organising intertextual relationship between ideas
	Organising ideas in a textual structure
Monitoring and revising	While writing monitoring and revising (at low/high level)
	After writing monitoring and revising (at low/high level)

to reveal how the high- and low- scoring writers employed these processes differently. Chan (2013) recommended the use of online research tools in future studies to collect such evidence.

All in all, it appears that research conducted to date indicate that high- and low-scoring students compose reading-into-writing tasks differently. However, evidence has been far from comprehensive in demonstrating the differences in their processes and how such differences might influence the development of the final products. The next section reviews literature in relation to methodological issues of process studies.

Methods of previous process studies

One obvious obstacle to investigating cognitive processes is that they cannot be directly observed. Previous research has investigated reading and/or writing processes through two broad approaches: self-reporting or observation.

The most commonly used method is to ask participants to report their cognitive processes either concurrently or retrospectively. As it is time costly to collect and analyse think aloud or interview protocols, studies using these methods usually involve a small number of participants (e.g. Plakans, 2009; Spivey, 1990). Questionnaires and checklists, on the other hand, are usually used in large-scale studies (e.g. Chan, 2013; Chan et al., 2014; Weir et al., 2007). A major issue of using concurrent techniques is the extent of reactivity and disruption imposed on the actual writing processes (Stratman & Hamp-Lyons, 1994). Whilst retrospective techniques do not interfere with the writing processes, the tendency of over-reporting has raised concerns (Harwood, 2009). The self-reporting approach largely relies on participants' perception of their processes, as well as their ability to recall and report them (Smagorinsky, 1994). Other researchers have investigated writing processes by means of direct observation, video recording (Bosher, 1998) and screen capture software (Chan, 2011). This approach allows researchers to collect data with minimum interruption to the writing event. However, more systematic studies have been rare, especially with respect to how writers compose on a reading-into-writing test. This may be partly because the methods used for data collection and analysis were time-consuming and difficult when applied to a larger number of participants.

In the past two decades, the shift of writing on paper to computer in various education contexts has encouraged researchers to gather writers' keystroke data. A few studies have investigated revisions in L2 writing (e.g. Severinson Eklundh & Kollberg 2003). Others used keystroke logging to investigate writing processes in real-life professional contexts. For example, Leijten, Van Waes, Schriver & Hayes (2014) investigated the processes of proposal writing of one designer.

Keystroke logging not only has the advantage of being relatively unobtrusive but it also allows researchers to collect a complete and accurate record of writers' external text production process with accuracy to seconds or even to milliseconds (Leijten & Van Waes, 2013). A range of measures such as writing fluency, pauses and document switches is now made available to researchers. Some of these measurements might be used to predict the performances produced by test takers. For example, it is found that frequency of pauses discriminated writers at different proficiency levels (Spelman Miller, 2000; Wengelin & Strömquist 2004). Fluency is also reported to be a useful

measure to discriminate skilled writers from L2 writers (Spelman Miller, 2000). However, these data points are largely quantitative and might not directly reflect writers’ writing processes. An important question that still needs addressing is how to interpret reading-into-writing processes from keystroke logging. Furthermore, little has been done to investigate how test takers’ reading-into-writing processes may impact their final products.

Barkaoui’s (2016), which was reviewed above, employed keystroke logging to compare high and low proficient L2 test takers’ revision patterns on the TOEFL iBT independent and integrated tasks in relation to what (i.e. the type and focus of the revisions) and when (i.e. the occurrence of revisions during the first, second and third segments of the entire text production). Following this strand of investigation in language testing, the present study aimed to use keystroke logging to investigate test takers’ reading-into-writing processes (including, but not limited to, revisions). We were particularly interested in comparing how test takers at different score levels transformed ideas from the sources. In contrast to Barkaoui’s (2016) study, we analysed the various keystroke logging measures (e.g. document switches, linear textual logs and revision logs) in conjunction with the final texts and interview protocols (see Section 3.4 for details of data analysis).

The research question is as follows: What similarities and differences are observed in the reading-into-writing process among two L2 learners (at different score levels) in terms of *overall text production patterns, writing fluency, pause patterns, transformation from sources* and *revision patterns* obtained from keystroke logging?

Methods

Participants

The aim of this study was to explore the potential of the methodology on a small scale. Two Chinese participants, Ben and Mary, at the postgraduate level were recruited (their names have been changed) (Table 2). They were both international students studying at a British University.

Ben was a second year PhD student. Ben did not have a qualified IELTS, or equivalent, score to indicate his general English proficiency at the time of study. For reference, he took the IELTS test 5 years ago and had an overall band of 6.5 and writing of 6.5. But it should be noted that this might not reflect his current proficiency. Regarding academic writing experience, Ben had published a few articles in peer-reviewed journals and conference proceedings as a second author. He was confident in his academic writing skills.

Mary had an overall IELTS of 6.5 and writing of 6 (which was taken within 2 years at the time of the study). Mary held a first degree in Linguistics. Mary arrived in the UK a few months to study a MA in TESOL. Although she regarded herself as a proficient

Table 2 Background of the participants

Name	Discipline	Gender	Educational level
Ben	Engineering	Male	Postgraduate
Mary	Linguistics	Female	Postgraduate

English writer, she commented that she was trying to adapt to the ‘western way’ of academic writing. She had no experience in publishing academic journal articles.

While neither of them had taken any integrated reading-into-writing tests, they had worked on reading-into-writing tasks during their studies. Having completed undergraduate and MSc programmes in Engineering in the UK, Ben was familiar with tasks such as technical reports, summary and literature review. Mary worked more on essays which require synthesis from multiple sources.

They were both proficient in spoken English and had no difficulty communicating with the researcher in English.

Task

General English Proficiency Test (GEPT) Advanced Writing Task 1 was used in this study. The test, which was developed and administered by the Language Training and Testing Centre (LTTC) in Taiwan, targets English learners at the CEFR C1 level (effective operational proficiency). In the testing literature, while some research has been done on tasks which involve non-verbal materials (e.g. Bridges, 2010; Yu & Lin, 2014) and a single verbal text (e.g. Chan, 2011), comparatively little has been done with tasks involving multiple verbal materials at tertiary level (i.e. B2 or above). In addition, this task type is believed to be most demanding as compared to summary from a single text or summary from non-verbal materials (Severinson Eklundh & Kollberg, 2003). At the time of study, GEPT was the only standardised writing test which provides a reading-into-writing task involving multiple written materials.

The task used in the study requires students to write an essay entitled ‘Should endangered languages be saved from extinction?’ based on two articles for a national essay contest. Each source text is about 400 words long. They have to summarise the main ideas of both articles and come to a conclusion with their own viewpoint on the topic. Table 3 summarises the basic features of the task. The suitability of the task was trialled with a similar population in a previous study funded by LTTC (Chan et al., 2014) reviewed above.

For this investigation, an electronic version of the task was developed using Adobe Creator. It consists of four pages—Task Instruction, Source text 1, Source Text 2 and Writing Sheet. The Writing Sheet was in Microsoft Word document format whereas the others were in pdf format. Participants were allowed to switch between these pages during the test. All editing functions including spell and grammar check in Microsoft Word were disabled.

The scripts were independently marked by two raters who are lecturers from the Department of Language and Communication at a British University. Both had more than 5 years of marking reading-into-writing tasks. They were trained with benchmarked scripts of the same test task collected from a previous study (Chan et al., 2014). Each script was scored from 1 to 5. LTTC requires a band 3, which indicates a performance at the C1 level, to pass the test¹.

Table 3 Basic features of the GEPT Advanced Writing Task 1

Brief task instruction	Time allowance	Input	Output
Write a comparative essay summarising the main ideas from the sources and stating own viewpoint	60 min	2 articles	At least 250 words

Writing process data collection—keystroke logging and retrospective interview

A keystroke logging programme called Inputlog (v.5.2.0.1) (Leijten & Van Waes, 2013) was used to collect time-based data of the test takers' text production. Three modules of Inputlog were used:

- (1) The data collection module to record the fine grain of text production processes;
- (2) The data analysis module to perform quantitative analyses (e.g. pauses, document switch, revisions);
- (3) The 'play' module to play the recording of the writing session (see Additional file 1: Appendix B for a screen shot).

The following steps were followed in the one-to-one data collection session:

- (1) The researcher briefed the participant about the purpose of the research and the participant signed the consent form;
- (2) The researcher demonstrated how to manage the task on computer;
- (3) The participant completed the task under test conditions i.e. timed and supervised (they were not given access to any type of support while completing the task);
- (4) Immediately after (3), the participant participated in a retrospective interview where the recording of the writing session was used as a stimulus. Participants could watch their entire composing process on screen. They and the researcher could pause, fast forward and backward the recording as well as move it to a time of interest. Participants were asked to describe their composing processes while watching the recording. A set of questions about different reading-into-writing processes was asked to prompt the participants when necessary (for sample questions, see Additional file 1: Appendix C).

Data preparation and analysis

Based upon the recommendations outlined in Leijten et al.'s study (2014), the raw data (for a sample, see Table 4) generated by Inputlog was filtered for later analysis. The following steps were used to prepare the keystroke data:

- (1) Activities irrelevant to the actual test task, such as familiarisation and entering candidate's information, were removed;
- (2) All the revision logs were highlighted for further manual analysis;
- (3) All occurrences when the writer switched from one document to another were highlighted for further analysis.

The keystroke data were then analysed at three levels. First, a time-based analysis of each of the test taker's writing was generated (see Figs. 2 and 3 in the next section) to interpret the overall pattern of their writing processes. The second level focused on quantitative data which could be automatically generated by the software including *total number of words*, *fluency* and *pauses*.

Table 4 Sample of Inputlog's general analysis output

Type	Output	Start time	Start clock	End time	End clock	Action time	Pause time	Pause location	Pause location full	Position	Doc length
keyboard	t	662614	00:11:02.614	662723	00:11:02.723	109	141	2	BETWEEN WORDS	20	20
keyboard	h	662739	00:11:02.739	662817	00:11:02.817	78	125	1	WITHIN WORDS	21	21
keyboard	e	662832	00:11:02.832	662926	00:11:02.926	94	93	1	WITHIN WORDS	22	22
keyboard	SPACE	662926	00:11:02.926	663051	00:11:03.051	125	94	2	BETWEEN WORDS	23	23
keyboard	a	663238	00:11:03.238	663331	00:11:03.331	93	312	2	BETWEEN WORDS	24	24
keyboard	r	664548	00:11:04.548	664657	00:11:04.657	109	1310	1	WITHIN WORDS	25	25

Fluency

Eight measures were used to indicate writers' fluency including total task time, total number of characters (process and product), product/process ratio (based on the number of characters), total number of words (process and product) and number of words per minute (process and product) (see Table 6). The total number of words produced during the entire text production (as compared to the total number of words in the final product) was used to indicate fluency during other processes.

Pauses

Following the practice used in similar studies (e.g. Leijten & Van Waes, 2013; Leijten et al., 2014), a threshold of 2 s was used to indicate a pause from typing characters or moving the cursor. The patterns of pauses (i.e. when, where and how long each pause took place) were analysed.

The final level of analysis involved tabulation and manual coding. To reveal how writers in this study produced their text in relation to the source texts, the researcher produced chronological tables (see Tables 8, 9, 10 and 11) combining information generated by *document switch analysis* and *linear textual analysis*.

Document switch analysis

The pattern of how the writers switched between the four documents (i.e. Task Instructions, Source Text 1, Source Text 2 and Writing Sheet) of the task was analysed in terms of total number, time and duration.

Linear textual analysis

Inputlog provides an XML file with a basic log file to show all keystroke inputs of the whole or particular sections of the writing. It is possible to conduct the analysis in five options including fixed number of intervals or fixed duration. The option of focus-based intervals (document switch being set as the focus) was chosen in this study.

Revisions

Inputlog indicates when and where a change was made. However, it was felt that such data gives limited indication of the processes involved. Adapting the revision framework developed by Severinson Eklundh and Kollberg (2003), Stevenson et al. (2006) and Barkaoui (2016), all revisions logs were manually coded by the researcher according to four dimensions: location, domain, orientation, and action. Location refers to where a change took place in terms of the production of text. Domain concerns the level of linguistic unit at which a change took place. Orientation addresses whether a change is related to the linguistic, content or mechanic aspects of writing quality. The linguistic aspect is further subcategorised into spelling-, grammar-, punctuation-, and phrasing-related. Action identifies if a change is an addition, deletion or substitution. The working definition of each category used in this study is provided in Table 5. The data was also coded by a research assistant to ensure reliability. The agreement rate of the location, domain and action dimensions was above 95% and orientation 89%.

The interview was transcribed by the researcher. The transcript was then coded based on the reading-into-writing model presented in Table 1 (for samples of coding, see Additional file 1: Appendix D). The data was second coded by the same research assistant. The agreement rate was 92%. The purpose of the interview was to understand

Table 5 Manual analysis of revision logs (adapted from Stevenson et al. 2006)

Dimensions	Working definition
Location	Indicating the place in the text where the change was made, either at the point of inscription or previous text.
Level	Indicating whether the change took place at the levels of within a word, within a clause, within a sentence, across sentences or across paragraph.
Aspects	
Linguistic	A change relating to a linguistic feature. Grammar - a change to correct a grammatical mistake, e.g. tense, agreement, part of speech. Spelling - A change to correct a spelling mistake. This paper does not distinguish between spelling mistakes and typing errors. However, most spelling occurrences below word level seem to be typing errors. Punctuation - A change relating to elements such as hyphens, apostrophes, (de)capitalization, commas, semi-colons, full stop, question marks and exclamation marks. Phrasing - A non-error change to substitute a word/phase/clause/sentence with an alternative relating to considerations such as style, tone, and cohesion without changing the meaning.
Content	A change which affects the meaning of the text.
Mechanical	A change which cannot be categorised as Linguistic or Content. It usually concerns the format of the text.
Action	Indicating whether the change was an addition, deletion or substitution

the writers’ composing from their perspectives. While most keystroke data agreed with writers’ self-reporting, there were some discrepancies between the two.

Findings

This section first presents the two students’ performance on the test, followed by a comparison of their reading-into-writing processes in relation to overall text production patterns, writing fluency, pause patterns, transformation from sources, and revision patterns.

Performance

Ben’s performance was scored an overall 4 of 5 and Mary’s 2 of 5. The raters assigned the same rating to both performances. An additional commentary about the performance below was provided by the first rater.

Ben’s script

The writer addresses all parts of the task. S/he covers most of the key ideas from both sources. Personal opinions on the topic are clearly addressed. The script is well-structured and easy to follow, with appropriate paragraphing. There is evidence of the use of a wide range of structures, vocabulary and linking devices to complete the task. The writing is generally accurate and appropriate.

Mary’s script

Although the writer attempts to provide a summary from the two sources, some key points are not clearly stated. The personal opinions part is sufficiently addressed. There is no paragraphing. The development of the ideas is not always coherent. The range of vocabulary and structures is sufficient but there are many low-level errors, e.g. subject-agreement, in the writing.

Overall text production patterns

A time-based graphic representation was produced for each participant (see Figs. 3 and 4) to investigate the degree of linearity/recursion of their overall text production patterns. In each graph, the entire writing process on the task was plotted in a timeline (*x*-axis) against the number of characters produced (*y*-axis). There are five elements in the graphs *Process line*, *Product line*, *Cursor position line*, *Pause dots* and *Focus analysis line*. For example, if a writer put down ‘This is a trial’, deleted ‘trial’ and then replaced it with ‘demonstration’, the graph would appear as in Fig. 1. The upper solid line (Process line) shows the number of characters produced in the Writing Sheet. The bottom solid line (Product line) shows the actual number of characters in the final Writing Sheet. A gap between the first line (Process line) and second line (Product line) indicates a text deletion. The dotted line (Cursor Position line) shows the position of the cursor during the text production.

When a writer places the cursor at the end of the text, the Cursor Position line overlaps with the Product line. But when the writer moves the cursor to somewhere in the text previously produced, the Cursor Position line would be at a lower position of the Product line. In the example of Fig. 1, the Cursor Position line overlaps with the Product line from the beginning until 01:35 when the writer moved to the previous text to make the change. The small dot indicates the occurrence of a pause, based on a 2-s threshold, from the text production process. Finally, there is a Focus Line (see Fig. 2 for an example). The horizontal line represents the entire text production whereas each vertical stroke indicates when the test taker switched from one document to another, e.g. from Task Instructions to Source Text 1.

The graph representing Ben’s writing processes (Fig. 3) was relatively simple and linear as compared to Mary’s (Fig. 4). Ben spent just slightly more than half of the allocated hour, i.e. 32 min, to complete the task. He spent about 11 min on conceptualising the task and reading the source texts before starting to type his response on the Writing Sheet (the processes will be discussed in detail later). According to the Focus Analysis line at the bottom of Fig. 3, Ben only switched from one event to the other occasionally. Additional file 1: Appendix E provides a full account of his 25 switches during the task. The steady and narrow gap between the Process Line and Product line

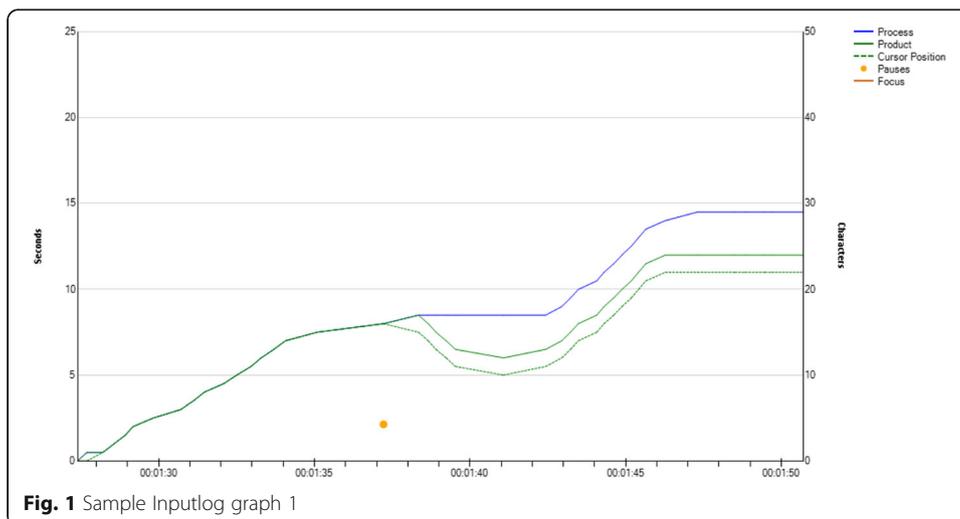
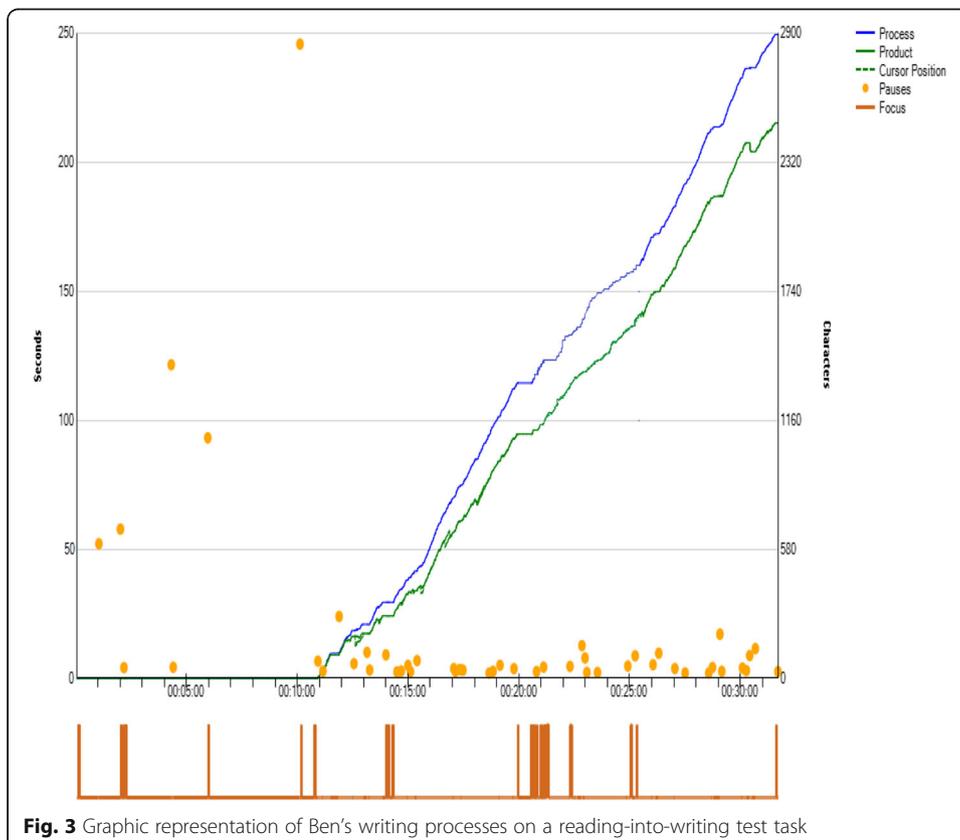


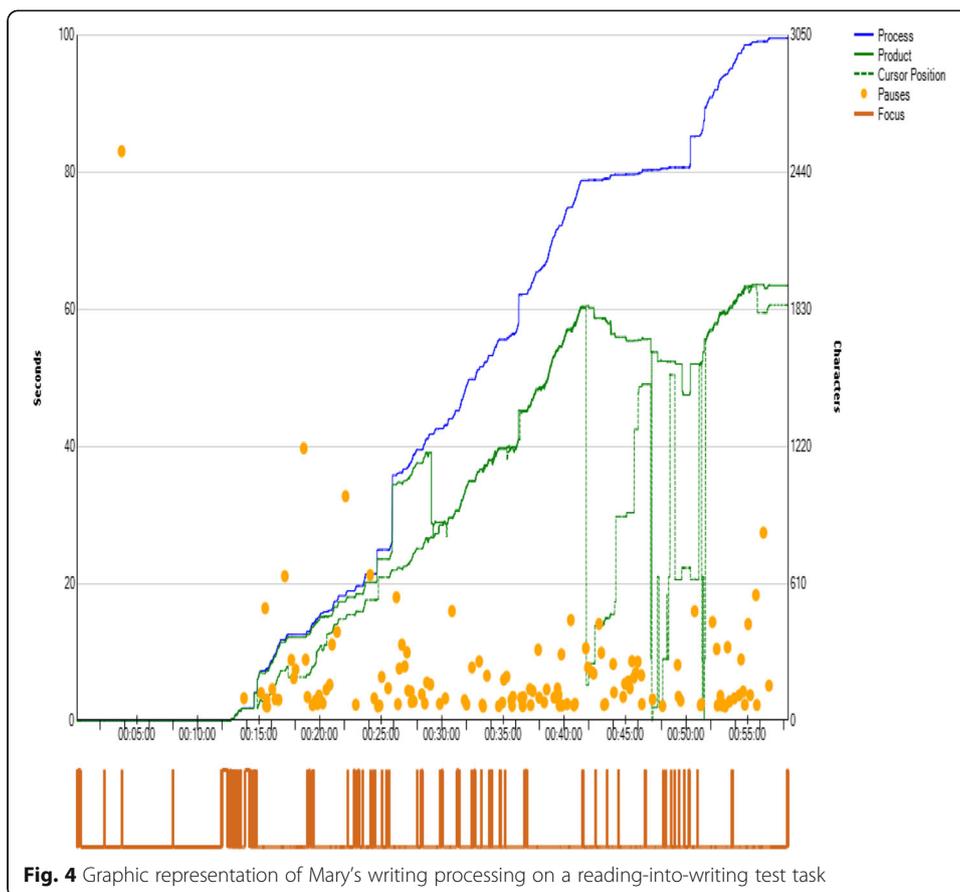
Fig. 1 Sample Inputlog graph 1



indicates that extensive deletion of texts was rare. The Cursor line is hardly observable (as it follows the Product line closely) which implies that Ben did not do much regressive revision. Indicated by density and location of the small dots, Ben's pauses (from typing) were brief (details see Pause patterns section).

According to Fig. 4, Mary used most of the hour (58.13 min) to respond to the task. Like Ben, Mary spent about 12 min on task conceptualization and reading before starting to write, but the graph indicates that Mary's writing processes were more recursive than Ben's. According to the Focus Analysis line, she switched between different documents intensively throughout the text production. A full account of the 47 switches she made is provided in Additional file 1: Appendix E. The gap between the Process Line and Product line is fairly narrow from 15 to 25 min of the task time, which indicates some occurrences of text deletion at this stage of Mary's writing. The gap then enlarges around 25 min of the task time which indicates a major text deletion occurrence. The gap further enlarges from 42 min till the end of the task time. As indicated by the Cursor Position Line, Mary was working on previously produced text extensively during this period. This suggests that she may have been reviewing and revising her text regressively at the





global level². Indicated by high density of the small dots in Fig. 4, her pauses were more frequent throughout than Ben's.

Writing fluency

As mentioned above, Ben spent about half an hour of the task time to complete the test whereas Mary used up the hour (see Table 6). Ben's total number of characters produced was 2918 and the total number of characters in his final text was 2509. Mary's total number of characters produced was 3049 but the total number of characters in her final text was only 1951. The number of characters includes all complete and incomplete words as well as spaces. Ben's product/process ratio based on the number of characters (0.86) was higher than that of Mary's (0.64). In other words, agreeing with Figs. 3 and 4, Mary deleted a higher proportion of her text than Ben did during text production.

For this study, it is more appropriate to consider test takers' writing fluency in relation to the total number of words they produced, as compared to writing fluency in producing individual words. The total number of words Ben produced was 487 and Mary 349. The final product of Ben contained 421 words and Mary 330 words. Ben had a noticeably higher text producing rate than Mary with respect to both number of words produced and number of words submitted. On average, Ben produced 15.50 words (process) and 13.41 words (final product) per minute. Mary produced 6.01 words (process) and 5.68 words (final product) per minute.

Table 6 Measures of writing fluency

Measure	Ben	Mary
Total Task Time (in minutes)	31.40	58.13
Total number of characters (Process)	2918	3049
Total number of characters (Product)	2509	1951
Product/Process ratio	0.86	0.64
Total number of words (Process)	487	349
Total number of words (Product)	421	330
Words per minute (Process)	15.50	6.01
Words per minute (Product)	13.41	5.68

Pause patterns

Table 7 shows a summary of the two writers’ pauses from typing characters or moving the cursor. The analysis excluded the pauses during the initial conceptualisation and reading phase. The total pause time of Mary was four times longer than Ben’s (858.22 vs 204.95 s). The mean pause length of Mary (5.65 s) was longer than Ben’s (4.1 s). In addition, Mary had almost three times more occurrences of pauses (i.e. a total of 152) than Ben (i.e. a total of 50).

In terms of the location of the pauses, about 65% of Mary’s pauses happened between words or clauses (for an average of 5.56 s each), 24% within words (for an average of 7.71 s) and 10% between sentences (for an average of 3.93 s). In contrast, a clear majority of Ben’s pauses (82%) happened between words or clauses (for an average of 4.31 s), 15% within words (for an average of 3.04 s). However, he did not pause often at the level of between sentences or paragraphs.

Interpretation of reading-into-writing sub-processes

Based on keystroke logging and interview data, we present more evidence of writers’ reading-into-writing processes. Segments of the writers’ keystroke logs, excerpts from final text and interview transcripts are provided as examples³ where appropriate.

Task representation

Almost all writers start a task by reading the task prompt to understand task instructions and plan their writing (Shaw & Weir, 2007). Ben and Mary in this study were no exception. Both students spent about 2 min at the beginning of the task time reading

Table 7 Measures of pauses

Measures	Ben				Mary			
	N	%	Mean length (in seconds)	SD	N	%	Mean length (in seconds)	SD
Total pauses (excluding the initial conceptualisation and reading phase)	50	100	4.10		152	100	5.65	
Within words	8	16	3.04	0.49	37	24.34	7.71	5.10
Between words or clauses	41	82	4.31	2.86	99	65.13	5.56	2.97
Between sentences	1	2	3.12	–	15	9.87	3.93	0.57
Between paragraphs	–	–	–	–	1	0.66	3.28	–
	Total		204.95		Total		858.22	

the task prompt. In the Interview, they reported macro-planning processes as they read the task prompt (see Examples 1 and 2).

'I was trying to understand the task instructions and follow what they require. I made plan in my mind' (Ben – Example 1).

'I read the instructions to find out what the task is about. I know I have to read the two passages about disappearing languages and summarise their points and then add my own conclusion...opinions' (Mary – Example 2).

Considering these protocols, one may expect Ben and Mary's processes of task representation to be quite similar. However, some interesting differences between the two students are observed from the keystroke logging data. According to the *document switch analysis* (Additional file 1: Appendix E), Ben consulted the task prompt twice, at the beginning for about 2 min (1 min 56 s)⁴ and very briefly (0 min 02 s) after reading both source texts. By contrast, Mary consulted the task prompt 11 times. Like Ben, she first read the task prompt for about 2 min (02 min 09 s) but she returned to the task prompt eight times as she composed the introduction of her essay.

By manually tabulating the *document switch analysis* and *linear textual analysis*, Tables 8 and 9 reveal how Ben and Mary composed their introduction.

As shown in Table 8, Ben composed the introduction without interruption for about 3 min (03 min 16 s) (Excerpt 1 shows the first paragraph from his final text). In his introduction, Ben stated the topic and summarised the common idea (i.e. the role of language) from both sources. He also indicated his stand on the issue.

This essay is about the argument of whether Endangered Languages should be saved from extinction or not. From both articles read, Language is a way of communicating between members of the human race as a result, Language in itself is important to human progression as a specie (Excerpt 1).

In contrast, as shown in Table 9, Mary consulted the task prompt eight times and each of the source texts once during the production of her introduction. She relied heavily on the sources for content and language by lifting phrases from the task prompt. As a result, her introduction has much resemblance to the task prompt (see Excerpt 2).

The aim of this article is to establish a comparison between the point of view of two renamed author: Chester Monrce and Gretchen Werner and state our own point of view about the topic discussed (Excerpt 2).

Table 8 Segments of Ben's text production (0:10:46–0:10:49)

Documents	Starting time	Duration	Keystroke logs
Task prompt	0:10:46	0:00:02	
Writing sheet	0:10:49	0:03:16	This eass###ssa #y is about the arguement of whether Eng#dangered Lanu#guages should be sav#f#ved from extinction or not. From both articles read, Language is a form of communicate#ion e#either #####between # #####within the human race. way #####ing #####bw#etween members of ### as ares### result, Language in itself is important to teh ###he progra#ession of ##### o##### human # as a specie.

Note: # indicates a deletion of the previous character

Table 9 Segments of Mary’s text production (0:11:56–0:19:29)

Documents	Starting time	Duration	Keystroke logs
Writing sheet	0:11:56	0:00:30	Should
Task prompt	0:12:26	0:00:03	
Writing sheet	0:12:30	0:00:09	Endangered Language
Task prompt	0:12:39	0:00:03	
Writing sheet	0:12:42	0:00:11	be saves ##d from ex
Task prompt	0:12:53	0:00:04	
Writing sheet	0:12:57	0:00:14	tinction?
Task prompt	0:13:11	0:00:03	
Writing sheet	0:13:13	0:00:15	<Save Endangered Languages
Task prompt	0:13:29	0:00:03	
Writing sheet	0:13:32	0:00:43	Before It’s Too Late
Task prompt	0:14:15	0:00:01	
Writing sheet	0:14:16	0:00:08	(n/a)
Source text 1	0:14:23	0:00:11	
Writing sheet	0:14:34	0:00:05	Chester Monrce > <Languages Don’t Need Saving: People Do
Source text 2	0:14:39	0:00:10	
Writing sheet	0:14:49	0:04:16	Gretchen Werner > The aim of this article is stablich #####establish a comparison between two renamed ##the point of view of author Chester Monrce and Gretchen Werner. The two ones discuss about en##Endangered Language. and to show
Task prompt	0:19:05	0:00:05	
Writing sheet	0:19:10	0:00:17	state m#our own point of t#view about
Task prompt	0:19:27	0:00:03	
Writing sheet	0:19:29	0:03:17	the topic## discussed#####

Note: # indicates a deletion of the previous character

Note 2: (n/a) indicates that the writer did not type anything within the duration

In this brief introduction, she used the same content words ‘establish’, ‘state’, ‘article’, ‘author’ as in the task prompt. She paraphrased other phrases/clauses from the task prompt; for example, by changing ‘comparative’ to ‘comparison (comparison)’ and ‘state your own viewpoint of this topic’ to ‘state our point of view about the topic’. In the interview, Mary believed that it is good to ‘stay close’ to the sources. She commented that ‘I see it is important to use exact words from the passages to show you read it (them) carefully. It (the task prompt) reminds me to put the words in quotation marks’. The part of the task instructions she was referring to was a warning of plagiarism. It states that students should use quotation marks if they use more than three consecutive words from the passages. Mary seemed to have mistaken the warning as an encouragement to lift words from the original texts. However, it is also possible that Mary believed in general that her response should closely reflect the original text for this type of task. She mentioned in the interview that use of quotation marks was ‘one of the good tips’ she learnt from EAP writing courses, and she applied these strategies frequently when writing academic essays. This, to some extent, reflects a misconception of

requirements for reading-into-writing tasks in general. In either case, her task representation of staying close to the sources influences how she completed the task, as illustrated in more details below.

Transforming ideas from sources (discourse synthesis)

Both writers reported that they read each of the source text carefully to understand the main ideas of the text before starting to write. During writing, they reread the source texts multiple times to select relevant information. Nevertheless, there were noticeable differences in how the two writers transformed ideas from the sources.

Ben’s processes of selecting, organising and connecting ideas appeared to happen co-incidently as he read (comprehend) and wrote (generating text). Table 10 illustrates how he composed the second and third paragraphs based on Source Text 1. At this point (0:14:05), he just finished writing the introduction (see Table 8). He returned to Source Text 1 twice each for 2 s. In the interview, he explained he ‘was looking for the spelling of the author’s name’ and reminding himself of the ideas he was going to write. He did not write anything the first time he returned to the Writing Sheet because he ‘was organising ideas’ in his mind. He then had a writing phase of about 6 min (06 min 14 s) and composed the two paragraphs.

Source Text 1 contains three main ideas of why endangered languages should be saved, i.e. destruction of identity, loss of linguistic resources and loss of specific knowledge. Ben selected two of them to be included in his essay. As previously mentioned, Ben did not refer to the source text when he was composing these paragraphs. He transformed the ideas largely in his own language (for example compare Excerpt 3 from the source and Excerpt 4 from Ben’s writing).

Table 10 Segments of Ben’s text production (0:14:05–0:20:34)

Documents	Starting time	Duration	Keystroke logs (final text at the time)
Source Text 1	0:14:05	0:00:02	
Writing sheet	0:14:07	0:00:11	(n/a)
Source Text 1	0:14:18	0:00:02	
Writing sheet	0:14:20	0:06:14	In the work of Chester Mource, he discussed that Endangered Languages should be preserved because they are part of the identity of a group of people. Without an identity, people do not know who they are and as a result, do not have a foundation upon which they can rely upon to make decisions. In addition, their languages sort of enable the group to know what is acceptable within the group and what is not. Without the language, a breakdown in self esteem, law and order can result. Furthermore, Chester Mource argues that if a Language is not saved from extinction, knowledge captured in the structure of the language could be lost resulting in a more improvised human race as a result of destruction of such precious life-saving knowledge such as what plants contain active ingredients for life-saving drugs.
Source Text 1	0:20:34	0:00:01	

Note: for clarity, the final text as it appears at the end of the duration is shown
 Note 2: (n/a) indicates that the writer did not type anything within the duration

He argues that the loss of a cultural leads to the extermination of self-worth in a society, intensifying problems of poverty, school drop-out rates, drug and alcohol abuse, and even suicide (Excerpt 3).

Without the language, a breakdown in self esteem, law and order can result (Excerpt 4).

There was also evidence of the process of connecting ideas. For example, Source Text 1 states that ‘the destruction of a language is the destruction of a rooted identity’. Ben illustrated this point by adding his own interpretation (see Excerpt 5).

Without an identity, people do not know who they are and as a result, do not have a foundation upon which they can rely upon to make decisions. In addition, their languages sort of enable the group to know what is acceptable within the group and what is not (Excerpt 5).

In contrast, Mary’s processes of selecting, organising and connecting ideas were less automatic. According to the interview, Mary attempted to anticipate the relationships between ideas from the source texts. However, her text seemed to show limited evidence of successful implementation of these high-level processes. Table 11 shows the period when Mary was composing a section of her text based on Source Text 1. She included all three main ideas by lifting chunks of texts from the source. These ideas appear in her text in the same order as in the original text. She connected these lifted texts with the use of formulaic expressions like ‘the first one is’ and ‘finally’.

She explained in the interview that her goal was ‘to include all the relevant ideas,’ and she would ‘tidy them up’ later. There was hardly any evidence of summarisation nor interpretation of the content which she lifted from the sources. It seems that her strategy was to delay part of the discourse synthesis process until when she revised her text. The writers’ revisions processes are discussed next.

Monitoring and revising

Ben mainly revised the immediate text as he composed. Ben submitted immediately after he had finished composing the essay because, according to him, the task was not

Table 11 Segments of Mary’s text production (0:22:50–0:28:23)

Documents	Starting time	Duration	Keystroke logs (final text at the time)
Writing sheet	0:23:08	0:00:21	advocates three result
Writing sheet	0:23:30	0:00:52	reason for it. Teh first one, according to the author is
Source text 1	0:24:22	0:00:11	
Writing sheet	0:24:32	0:00:55	apart from its practical use as a tool of communication, is also a means of cultural transmission.
Source text 1	0:25:27	0:00:12	
Writing sheet	0:25:39	0:02:36	The second reason for him is because languages are a valuable resource for linguists in their search for the relationship between languages and the development of mental processes its lost to the academic lost Finally, he argues
Source text 1	0:28:15	0:00:07	
Writing sheet	0:28:23	0:01:36	the lack of writing form, for this he gives a

Note: for clarity, the final text as it appears at the end of the duration is shown

Note 2: lifted text is bolded for reference

too challenging. He was confident that his essay ‘was good enough’. However, he commented that he would have revised the essay carefully if he were writing for assignments or tests⁵ in real-life. In contrast, Mary revised her previously produced texts extensively after two-thirds of her task time. She gave an account of her revision processes (see Example 3).

I made many changes to my essay at this stage. I changed all those I previously copied from the passages. I either put them in direct quotes or replaced them with my own words. I didn't worry too much about what I copied as long as they were relevant. But at this point, I needed to make my essay precise and coherent (Example 3).

Unlike Ben who could transform ideas from sources as he wrote, Mary apparently separated her discourse synthesis process into manageable steps. We would expect the difference in their strategies to be reflected in the various analyses of their revisions in relation to location, level, aspects and action (see Table 12).

It is somewhat surprising that although Ben thought he did not make many changes to his text, he actually made a total of 145 revisions (i.e. deletion, addition or substitution) when Mary made 238 revisions. Agreeing with their own account, almost all (91.72%) of Ben's change was made at the position of inscription as he composed his

Table 12 Revisions analysis

Measures	Ben	Mary
Total no of revisions	145	238
Location	%	%
Point of inscription	91.72	45.8
Previous text	8.28	54.2
	100	100
Level		
Below word	69.0	35.2
Lexical/Below clause	23.6	50.0
Clause and above	7.4	14.8
	100	100
Aspects		
Content	4.5	6.5
Linguistic		
Grammar	6.5	3.5
Spelling	49.6	13.5
Punctuation	1.4	13.5
Phrasing	17.9	34.2
Mechanical	20.1	28.8
	100	100
Action		
Addition	3.6	26.5
Deletion	91.4	64.1
Substitution	5.0	9.4
	100	100

text, whereas more than half of Mary's revisions (54.2%) was made to the previously produced texts.

The reasons why Ben did not realise he made that many changes could be because most of these changes were deletion (91.4%), within a single word (69%) and related to spelling or typo (49.6%). Only 4.5% of his revisions was related to content. Below is a rare example of a content revision where he made an addition (indicated by {}) to his text.

Furthermore, {Chester Mource argues that if a Language is not saved from extinction,} captured in the structure of the language could be lost resulting in a more improvised human race

On the other hand, Mary had more revisions above the level of within a single word than Ben. Half of her revisions were at the level of below a clause and 14.8% at the level of a clause or above. 34.2% of the revisions was phrasing but most of them were to replace a single word with a synonym. For example, 'globalisation' was substituted by 'international'. As previously discussed, it was Mary's plan to 'make many changes' at this stage to complete transforming ideas from the sources. However, a closer examination of the keystroke logs shows that the transformation was limited, as illustrated below:

Before revision

Teh first one, according to the author language is, apart from its practical use as a tool of communication, is also a means of cultural transmission. The second reason for him is because languages are a valuable resource for linguists in their search for the relationship between languages and the development of mental processes its lost to the academic lost.

After revision

The first one is that language 'is also a means of cultural transmission'. The second reason for him is academic lost, because 'languages are a valuable resource for linguists'.

Skilled writers eliminate text to diminish the amount of repetition (Severinson Eklundh & Kollberg, 2003) but Mary's deletion apparently serves to avoid excessive lifting from the sources. Instead of paraphrasing, summarising or elaborating the main ideas from the sources, she presented the ideas with, arguably, inappropriate use of quotation. As a result, as commented by the first rater, the development of her text was impeded.

Discussion

The aim of the paper is to explore writers' processes on a reading-into-writing test by employing keystroke logging. Towards this, we conducted a small-scale experimental study with two L2 students at different scores levels. We demonstrated a multi-level analysis of combing the various keystroke logging measures (e.g. document switches, linear textual logs and revision logs) with the students' final texts and interview protocols. This was done to explicate the two writers' reading-into-writing processes in relation to graphic time-based analysis of the overall text production patterns, quantitative measures of writing fluency and pauses, and reading-into-writing sub-processes (i.e. task representation, transforming ideas from sources and revisions). We now further discuss how the differences observed in their processes might, to some extent, account for their performances on the task.

The time-based analysis graphs are useful in visualising the pattern and degree of recursion of the two writers' processes of developing the final text within the task time. They show that there were differences in the two writers' reading-into-writing processes. For example, Ben's writing was rather linear whereas Mary's writing shows a pattern of alternating sources, writing and regressive revisions. However, given that Ben scored higher than Mary on the task, the patterns observed are somewhat different from expectation that less skilled writers usually adopt a more linear approach to writing because they are predominantly occupied by text generation (Flower et al., 1986). It seems difficult to interpret from the graphs how the observed differences might have contributed to their final performance.

On the other hand, the measures of fluency and pauses appear to be useful in indicating differences in writers' processes. The findings show that Ben who scored higher on the task had a higher writing fluency and paused less during text production than Mary. It is reported in the literature that skilled writers have a higher writing fluency than unskilled writers as most of the lower-level writing processes have become automatic for skilled writers (Field, 2004; Kellogg, 1996). Excessive pauses usually indicate writers' lack of ability to produce chunks of language when writers translate ideas into linguistic forms (Matsuhashi, 1981). But in the context of this study, it seems that the differences observed in the two writers' fluency and pausing patterns may also be because of their reading-into-writing strategies (or inappropriate use of these strategies). For example, Mary's strategy to rely heavily on the sources during writing resulted in more frequent and longer pauses. This is in line with previous research that suggests that the high demand of integrated tasks results in lower fluency and longer pauses in most writers' text production, especially those who do not have a good schema for constructing discourse from sources (Severinson Eklundh & Kollberg 2003).

Actual and full evidence of how the writers develop their texts on the test was made available by combining the document analysis and linear keystroke log analysis. As illustrated in the results section, this was found useful in understanding how Ben and Mary constructed their writing particularly in relation to their source use.

Some interesting insights into the differences between the two writers' processes which might have been overlooked or even misinterpreted by means of quantitative measures were revealed. For example, measures of Ben's processes in terms of frequencies and degree of recursion seem to suggest a simple and linear knowledge telling approach to writing (Scardamalia & Bereiter, 1987). However, the qualitative analysis of keystroke logging data shows that his processes of transforming ideas from sources into his writing were well embedded in his reading and writing. As he was able to execute these processes with high automaticity (Field, 2004), these high-level knowledge transformation processes might not be observable by other means.

The evidence was also useful in indicating any seemingly discrepancies between writers' perceived processes and their actual (or successful) execution of the text development processes. Both students reported similar use of source text in the interview, as also found in Plakans and Gebril's study (2012). For example, Mary might have planned to summarise ideas from the two sources and interpret relations between these ideas. However, evidence showed that she actually 'transformed' these ideas by lifting chunks of texts from sources and 'connected' them by using formulaic expressions, which is a typical feature found in low-scoring performances (Plakans & Gebril, 2013).

Finally, the revision analysis based on quantitative measures such as total number of revisions, location, level, aspects and action was useful to some extent. As reported in previous studies, differences were observed between the high-scoring and low-scoring writers' revision patterns in terms of these measures. However, as argued by Steverinsson Eklundh & Kollberg (2003), Ben's local elementary revisions on accuracy and form during writing did not necessarily reflect the writer's overall composing strategies. In other words, analysis of revisions relying merely on these measures might not tell us a lot about the writers' reading-into-writing processes. In this study, we further analysed writers' revisions with a close examination of the keystroke logs. This provides a new insight that some of the writers' discourse synthesis processes (i.e. selecting, organising and connecting ideas from sources) might be combined with their revising processes. Given its virtual role in the reading-into-writing literature, discourse synthesis was mostly examined separately from other processes. Nevertheless, the results indicate the importance of interpreting how writers transform ideas from sources in relation to other sub-processes.

Conclusion and recommendations

The study has confirmed that use of keystroke logging and interview provides a promising approach to the study of writers' reading-into-writing processes, which opens a window to cognitive validation of the task type. The results suggest that the nature of the writers' reading-into-writing processes might have a major influence on the writer's final performance.

Nevertheless, several limitations in the methodology should be noted. The number of participants of this study was small and hence the generalisability of the results is limited. Also, only one type of integrated reading-into-writing task was used. However, we did not intend to characterise the writers' proficiency based on this single task, especially when many intervening variables could have influenced their performance. In future studies, test takers' variables such as gender, computer familiarity, familiarity with integrated tasks, topic familiarity should be controlled. Although we appreciate the peril in including only two participants, the small sample size enabled us to provide a detailed account of how the various measures obtained from keystroke logging can be combined with other types of data (e.g. final product, interview protocol) to explicate the test takers' reading-into-writing processes.

As illustrated in the study, keystroke logging helps to generate actual evidence of test takers' text production processes, of which is arguably not accessible by other methods. However, one of the concerns of using keystroke logging data in process studies is the tendency to rely heavily on the numeric and mechanic measures (Leijten & Van Waes, 2013). The multi-level analysis proposed in this study is useful to reveal test takers' reading-into-writing processes, particularly in relation to how they (or fail to) transform ideas from sources.

The findings indicate that writers' reading-into-writing processes have an impact of the development of the text and their final performance on the test. This is in line with previous findings (e.g. Plakans, 2010; Wolfersberger, 2013). Ben, who showed evidence of transforming ideas from sources in the final text, scored higher on the test than Mary who mainly lifted contents from sources and revised by deletions. Confirming previous studies, low-scoring students may find it difficult to move away from the

source texts due to misinterpreted task representation (Wolfersberger, 2013) and inadequate discourse synthesis skills (Plakans, 2009). It is not our intention to draw conclusions regarding how higher- and lower- scoring writers composed differently. However, the study has demonstrated the strength of keystroke logging in providing actual visible evidence of test takers' reading-into-writing processes and identifying differences between their processes for further research in the future. In addition, as we see in Mary's case, the writer was aware of the importance of knowledge transformation in integrated tasks. However, due to inappropriate and ineffective reading-into-writing strategies, her final text does not necessarily reflect the writer's attempt to transform ideas from sources. This indicates that Mary, and perhaps other L2 writers, would benefit from teaching of reading-into-writing skills rather than training of mechanic strategies to avoid plagiarism.

On a final note, keystroke logging helps generate the quantitative data reported in this study. However, to achieve the level of details necessary for process research, extensive manual coding and tabulation is required to generate qualitative data in relation to text development, source use and revisions. However, as shown in this study, the approach, though intensively laborious, provides a new dimension of capturing and partially identifying differences in writers' reading-into-writing processes.

Endnotes

¹For more details, see https://www.lttc.ntu.edu.tw/GEPT1/Advanced/06.test_score.htm.

²The level refers to the distance of the revisions from the point of text production. A global level revision is a revision is made at a text-level and paragraph-level (Severinson Eklundh & Kollberg, 2003). For example, when the writer goes through a previously written paragraph and makes changes in it.

³Any grammatical mistakes are kept in these examples.

⁴The number indicates the time spent in minutes and seconds.

⁵Although the task was administered to the participants under testing conditions, the results would not have any impact to their studies.

Additional file

Additional file 1: Appendix A. Working definition of the cognitive phases. Appendix B. A screenshot of the stimulus video used in the stimulated recall interview. Appendix C. Sample guided questions in stimulated recall interview. Appendix D. Sample of transcript coding. Appendix E. Frequency of switches between documents. (DOCX 78 kb)

Funding

This is not a funded research.

Authors' contributions

The author read and approved the final manuscript.

Competing interests

The author declares no competing interests.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received: 22 February 2017 Accepted: 19 May 2017

Published online: 28 June 2017

References

- Ascención-Delaney, Y. (2008). Investigating the reading-to-write construct. *Journal of English for Academic Purposes*, 7, 140–150.
- Barkaoui, K. (2016). What and when second-language learners revise when responding to timed writing tasks on the computer: The roles of task type, second language proficiency, and keyboarding skills. *The Modern Language Journal*, 100(1), 320–340.
- 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.
- Bridges, G. (2010). Demonstrating cognitive validity of IELTS Academic Writing Task 1. *University of Cambridge ESOL Examinations Research Notes*, 42, 24–33.
- 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. University of Bedfordshire.
- 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 Reports, RG-03*, 1–89.
- Chan, S. H. C., Inoue, C., & Taylor, L. (2015). Developing rubrics to assess the reading-into-writing skills: A case study. *Assessing Writing*, 26, 20–37.
- Chan, S. H. C., Bax, S., & Weir, C. J. (in press). Researching participants taking IELTS Academic Writing Task 2 (AWT2) in paper mode and in computer mode in terms of score equivalence, cognitive validity and other factors. *IELTS Research Report*.
- Connor, U., & Carrell, P. (1993). The interpretation of tasks by writers and readers in holistically rated direct assessment of writing. In J. Carson & I. Leki (Eds.), *Reading in the composition classroom: Second language perspectives* (pp. 141–160). Boston: Heinle & Heinle.
- Cumming, A. (2013). Assessing integrated writing tasks for academic purposes: Promises and perils. *Language Assessment Quarterly*, 10(1), 1–8.
- Field, J. (2004). *Psycholinguistics: the Key Concepts*. London: Routledge.
- Flower, L. (1990). The role of task representaiton in 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 and Social Process* (pp. 35–73). New York: Oxford University Press.
- Flower, L., Hayes, J. R., Carey, L., Schriver, K., & Stratman, J. (1986). Detection, diagnosis, and the strategies of revision. *College Composition and Communication*, 37, 16–55.
- Grabe, W. (2003). Reading and writing relations: Second language perspectives on research and practice. In B. Kroll (Ed.), *Exploring the dynamics of second language writing* (pp. 243–259). Cambridge: Cambridge University Press.
- 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., & Flower, L. S. (1980). Identifying the organisation of writing processes. In L. W. Gregg & E. R. Steinberg (Eds.), *Cognitive processes in writing* (pp. 3–30). Hillsdale: Lawrence Erlbaum Associates.
- Hirvela, A. (2004). *Connecting Reading and Writing in Second Language Writing Instruction*. Ann Arbor: The University of Michigan 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: Lawrence Erlbaum Associates, Publishers.
- 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.
- Leijten, M., & Van Waes, L. (2013). Keystroke-logging in writing research: Using Inputlog to analyze and visualize writing processes. *Written Communication*, 30(3), 358–392.
- Leijten, M., Van Waes, L., Schriver, K., & Hayes, J. R. (2014). Writing in the workplace: Constructing documents using multiple digital sources. *Journal of Writing Research*, 5(3), 285–337.
- Matsuhashi, A. (1981). Pausing and planning: The tempo of written discourse production. *Research in the Teaching of English*, 15(2), 113–134.
- Plakans, L. (2009). Discourse synthesis in integrated second language writing assessment. *Language Testing*, 26, 561–585.
- Plakans, L. (2010). Independent vs. Integrated Writing Tasks: A Comparison of Task Representation. *Tesol Quarterly*, 44(1), 185–194.
- Plakans, L., & Gebril, A. (2012). A close investigation into source use in integrated second language writing tasks. *Assessing Writing*, 17(1), 18–34.
- Plakans, L., & Gebril, A. (2013). Using multiple texts in an integrated writing assessment: Source text use as a predictor of score. *Journal of Second Language Writing*, 22(3), 217–230.
- Scardamalia, M., & Bereiter, C. (1987). Knowledge telling and knowledge transforming in written composition. In S. Rosenberg (Ed.), *Reading, writing and language learning* (Vol. 2, pp. 142–175). Cambridge: Cambridge University Press.
- Severinson Eklundh, K., & Kollberg, P. (2003). Emerging discourse structure: computer-assisted episode analysis as a window to global revision in university students' writing. *Journal of Pragmatics*, 35(6), 869–891.
- Shaw, S., & Weir, C. J. (2007). *Examining Writing: Research and Practice in Assessing Second Language Writing, Studies in Language Testing 26*. Cambridge: Cambridge University Press and Cambridge ESOL.
- Smagorinsky, P. (1994). *Speaking about writing: Reflections on research methodology*. USA: Sage Publications, Inc.
- Spelman Miller, K. (2000). Academic writers on-line: investigating pausing in the production of text. *Language Teaching Research*, 4(2), 123–148.
- Spivey, N. N. (1990). Transforming texts: Constructive processes in reading and writing. *Written Communication*, 7(2), 256–287.

- Spivey, N. N. (2001). Discourse synthesis: Process and product. In R. G. McInnis (Ed.), *Discourse synthesis: Studies in historical and contemporary social epistemology* (pp. 379–396). Westport: Praeger.
- Spivey, N. N., & King, J. R. (1989). Readers as writers composing from sources. *Reading Research Quarterly*, 24(1), 7–26.
- Stevenson, M., Schoonen, R., & Deglopper, K. (2006). Revising in two languages: A multi-dimensional comparison of online writing revisions in L1 and FL. *Journal of Second Language Writing*, 15(3), 201–233.
- 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.
- Weigle, S. C. (2004). Integrating reading and writing in a competency test for non-native speakers of English. *Assessing Writing*, 9(9), 27–55.
- Weir, C. J., O'Sullivan, B., Jin, Y., & Bax, S. (2007). Does the computer make a difference? The reaction of candidates to a computer-based versus a traditional hand-written form of the IELTS writing component: effects and impact. In P. McGovern & S. Walsh (Eds.), *IELTS research reports* (Vol. 7, pp. 311–347). Canberra: British Council & IDP Australia.
- Weir, C. J., Vidakovic, I., & Galaczi, E. (2013). *Measured Constructs: A History of the Constructs Underlying Cambridge English Language (ESOL) Examinations 1913–2012*. Cambridge: Cambridge University Press.
- Wengelin, Å., & Stromqvist, S. (2004). Text-Writing Development Viewed Through On-line Pausing. In R. I. Berman (Ed.), *Language development across childhood and adolescence: Psycholinguistic and crosslinguistic perspectives. TILAR (Trends in Language Acquisition) series* (Vol. 3, pp. 70–190). Amsterdam: John Benjamins.
- Wolfersberger, M. (2013). Refining the Construct of Classroom- Based Writing-From-Readings Assessment: The Role of Task Representation. *Language Assessment Quarterly*, 10(1), 37–41.
- Yu, G., & Lin, S. (2014). A Comparability Study on the Cognitive Processes of Taking Graph-based GEPT-Advanced and IELTS-Academic Writing Tasks. *LTTC-GEPT Research Reports*, 02, 1–70.

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► springeropen.com
