

# L2 Writing in Test and Non-test Situations: Process and Product

Baraa Khuder\* & Nigel Harwood°

\* Aleppo University | Syria

° University of Sheffield | UK

**Abstract:** Test writers sometimes complain they cannot perform to their true abilities because of time constraints. We therefore examined differences in terms of process and product between texts produced under test and non-test conditions. Ten L2 postgraduates wrote two argumentative essays, one under test conditions, with only forty minutes being allowed and without recourse to resources, and one under non-test conditions, with unlimited time as well as access to the Internet. Keystroke logging, screen capture software, and stimulated recall protocols were used, participants explaining and commenting on their writing processes. Sixteen writing process types were identified. Higher proportions of the processes of translation and surface revision were recorded in the test situation, while meaningful revision and evaluation were both higher in the non-test situation. There was a statistically significant difference between time allocation for different processes at different stages. Experienced teachers awarded the non-test texts a mean score of almost one point (0.8) higher. A correlational analysis examining the relationship between writing process and product quality showed that while the distribution of writing processes can have an impact on text quality in the test situation, it had no effect on the product in the non-test situation.

**Keywords:** test writing, non-test writing, keystroke logging



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

Contact: Nigel Harwood, School of English, University of Sheffield, 1 Upper Hanover Street, Sheffield S3 7RA, | UK - n.harwood@sheffield.ac.uk

Copyright: Earli | This article is published under Creative Commons Attribution-Noncommercial-No Derivative Works 3.0 Unported license.

## 1. Introduction

L2 writers have long expressed their disappointment with writing proficiency tests; at least some test-takers believe that their true writing ability is not displayed due to the constraints of test conditions (Polio & Glew, 1996). Indeed, various researchers have raised concerns that writers might not go through the different stages of the writing process during the tests, omitting or curtailing processes such as planning and revision (e.g. Hamp-Lyons & Condon, 2000; Sanders & Littlefield, 1975; Weigle, 2002; Wolcott, 1987). However, most research that investigates the product of test and non-test writing, or investigates writing which is produced under conditions in which writers are afforded greater or lesser amounts of time, concludes that there is no statistically significant difference between the texts produced in each situation in terms of the product and final mark awarded (e.g. Caudery, 1990; Kenworthy, 2006; Kroll, 1990; Polio et al., 1998). Nonetheless, the area remains under-researched, and we have inadequate knowledge of exactly how the writing process differs between the two conditions, and whether and to what extent timed and untimed texts differ as far as raters' assessments are concerned. This paper pursues both lines of enquiry.

## 2. Review of Relevant Literature

### 2.1 Writing in test and non-test situations

One of the few studies which compares the writing process in test and non-test situations was conducted by Hall (1991). Participants were permitted unlimited time in the non-test condition to finish the writing task. Hall investigated the effect of writing apprehension on the composing processes of 6 ESL writers from various cultural and educational backgrounds and with different levels of English proficiency (Hall's participants had in fact failed an ESL essay test and were taking a writing course to help them pass it). This study focused on the time allocated to each process, and the frequency, location and duration of pauses while composing. Prior to conducting each writing session, participants completed a questionnaire that measured their anxiety about the timed/untimed situation in order to determine the correlation between anxiety and performance. Participants' writing behaviour was videotaped and post-hoc interviews were conducted. Hall (1991) found that the texts produced in test situations tended to be more syntactically complex than those produced in non-test conditions. There were no significant differences between the composing processes in each situation: participants displayed the same processes of pre-writing, inscribing and post-drafting, i.e., planning, writing and revising in Hayes and Flower's (1980) terms. Individual differences, like test apprehension and anxiety, affected the writing process more than timed/untimed conditions. The more anxious the writers were, the more they tended to pause in both situations. Hall's (1991) findings concerning the quality of the writing produced in the two conditions were inconclusive, although in general writers

produced superior texts in terms of complex structures in the test situation. However, the higher the subjects' anxiety level, the higher the marks awarded to their texts in the non-test situation.

Reviewing Hall's (1991) study, Curry (2004) indicated that, in the test situation, writers might focus on the word level (grammar and vocabulary choice) rather than the substance of their writing (generating ideas and considering the coherence of their arguments) because of the lack of resources; one can assume that, in the non-test situation, the availability of materials enables writers to read about the topic which in turn may encourage them to focus on the substance of their texts. Curry also pointed out the effect limited time might have on writers' priorities: In the test situations, writers most probably wrote one draft only, and thus focused on the word level at the expense of substance, attention to which being likely to require more time (cf. Uzawa, 1996). Time can also affect the writing process as it can motivate writers to engage in more extensive planning (Smith, 1994).

Not only can time affect the writing process, but also writers' perception of a writing task; as reflected, for instance, in their fear of it or their interest in doing it. Apprehensive writers may perceive writing as unrewarding and may appropriate other people's ideas rather than expressing their own (Faigley, Daly, & Witte, 1981). Anxious writers have low self-confidence in their ability to compose; they often believe that writing is an inner gift that people either have or do not have (Wachholz & Etheridge, 1996). Writers' views also should be understood in the wider socio-cultural context; how writers perceive a given task and the goals they set for themselves can be influenced by social and cultural factors associated with their learning context (Kormos, 2012). To what extent the culture and educational system value writing will impact upon the place of writing in the curriculum; and writing goals will be similarly affected (Durgunoglu & Verhoeven, 1998). These goals are also affected by writers' interest in composing and their confidence in their writing ability (Manchón, 2009). Thus, writers might differ in the way they perceive a writing task: some might consider it as an opportunity to express their ideas, while others simply perceive it as a tool to reflect their linguistic ability (Hayes, 1996).

Moving from our concern with the writing process and what might affect it, we now come to the writing product. Here we review studies that have compared the quality of writing in test and non-test situations, or under conditions in which writers are afforded a lesser or greater amount of time to compose and revise (e.g., Caudery, 1990; Kenworthy, 2006; Kroll, 1990; Polio et al., 1998). These studies indicate that the main advantage of having extra time and resources is a slight improvement on the lexical and grammatical levels of the text, but these improvements are relatively small and are insufficient to affect the total score awarded (e.g., Caudery, 1990; Polio et al., 1998). For instance, Kroll (1990) asked twenty five advanced ESL undergraduates to write four essays: two at home, in fourteen days, and two in class, in sixty minutes, finding no statistical difference in the quality of texts written at home and in the class. And in Caudery's (1990) study of twenty four EFL writers who wrote one in-class essay in forty

minutes and started another essay in class but then finished it at home in two days, the difference between the total scores awarded was not statistically significant, even though writers scored statistically significantly higher under the non-timed condition in grammar and syntax. Polio et al. (1998) also found non-statistically significant differences when comparing at-home and in-class writing when holistically rating the writing of sixty five ESL graduates and undergraduates. However, one deficiency in the studies above is that the tests the participants were exposed to were not part of a high stakes test or even a preparation for it; most writing done in these studies was part of a writing course, and thus it is difficult for us to infer the apprehension/anxiety level and the extent to which these writers perceived the task as an authentic one where they were obliged to reflect their true writing ability.

Relevant to our study, in which writers used extra resources (Internet resources and dictionaries), we now discuss studies which attempted to investigate the differences in the writing product of test and non-test situations by adding a non-test element to the test situation: for example, by providing test takers with extra time, dictionaries, or online materials. Researchers like Crone, Wright, and Baron (1993), Hale (1992), Livingston (1987), and Powers and Fowles (1996) investigated the effect of slightly increasing the time limits on writing performance in test conditions. Contrary to studies that compared writing at home and in class, these studies retained the test situation elements (limited time and no extra resources). This was done by slightly increasing the time limits from, for instance, thirty to forty minutes. Overall, this research indicated that giving examinees more time resulted in slightly higher scores, although the improvement in scores was statistically non-significant and did not affect the students' ranking relative to each other .

The use of dictionaries during writing tests has been found to affect the L2 writing process (Skibniewski & Skibniewska, 1986), if not the quality of the product (Kobayashi & Rinnert, 1992; East, 2008). Skibniewski and Skibniewska (1986) concluded that L2 writers who used dictionaries paused more and conducted more planning than writers who did not. Similarly, East (2008) found bilingual dictionary use can have a negative influence on the L2 writing process as expressed in participants' subsequent interviews. Most participants held negative opinions about using dictionaries mainly because it interrupted the writing flow. Some participants suggested that without consulting bilingual dictionaries they thought in L2; however, they tended to think in their L1 after using them.

Online materials are another element that can be added to test situations, as in the OBOW (Open Book Open Web) approach. OBOW is a form of assessment that attempts to present tests as real life performances in the sense of allowing test-takers to search web pages and books the same way they do in real life. The purpose of doing so, then, is to ensure that what is being tested is not test takers' memory but their information-seeking and problem-solving ability (Heijne-Penninga et al., 2008; Williams, 2011). Opinions as to the efficacy of OBOW are divided, however. Fulcher (2010) suggested that by allowing access to online materials, we might end up testing

the ability to read, search, and use online resources rather than testing writing ability. However, he conceded that traditional essay writing tests are inauthentic because in real-life performances we write in response to reading or listening.

Even though the above-cited studies demonstrated the effect of extra time and materials on the writing process and product, these studies only tended to focus on one or the other—process or product in test and non-test situations. We may therefore ask: Are there differences in the writing process and product across test and non-test situations, and, if so, what kind of differences? And is there a relationship between writing process and product across both test and non-test situations? Hence the present study aimed to compare writers' performances, processes and products in test and non-test situations, i.e., timed vs. untimed conditions, with writers being allowed access to online resources in the non-test condition.

Since this study explored differences in the writing process, it is relevant to briefly review work on writing process models.

## **2.2 Writing process models and studies**

Although this research investigates the L2 writing process, the literature on the L1 writing process is also of interest to our study: while L2 writing is strategically, rhetorically, and linguistically different from L1 writing (Silva, 1993), L1 research has helped inform work on the L2 (Shaw & Weir, 2007). Furthermore, research on the L2 composing process has only attracted attention more recently (Plakans, 2008). We will therefore briefly present research on L1 and L2 composing processes as background. This body of work also provides insights into the constructs that are measured in writing assessment (Bachman & Palmer, 1996).

### **2.2.1 Writing process models**

Hayes and Flower (1980) proposed that the writing process consisted of multiple recursions, involving three main stages: planning, translating and reviewing. Writers moved from one stage to the other when necessary and according to the "job demands" (p.209). Hayes (1996) later reformulated the model to show how individual writing behaviour can be influenced by several factors which will vary from writer to writer: these include goal setting, predispositions, beliefs and attitudes, and cost/benefit estimates.

Bereiter and Scardamalia (1987) proposed a distinction between novice and proficient writers by providing two writing models, one for knowledge telling and the other for knowledge transforming. They argue that novice writers follow the knowledge-telling model, which is semi-linear and involves recalling previous experiences and writing these down, while more proficient writers follow the knowledge-transforming model, which involves problem solving, careful planning and extensive revising. The knowledge-telling model includes a mental representation of the

task (planning), then writers move to recall their writing and content knowledge, that is, their understanding of the writing process and the topic respectively. Ideas are generated in a simple natural way; one idea follows the other, and their compositions are poorer in quality and briefer (Graham & Harris, 2000, 2003). On the other hand, more proficient writers tend to 'problematize' more and to move freely between both knowledge-telling and knowledge-transforming stages. This results in a more recursive composing process.

The above-mentioned models are concerned with the L1 writing process only. One of the rare models which attempts to describe the L2 writing process is that of Grabe and Kaplan (1996). They divided language knowledge into three types: linguistic knowledge, discourse knowledge and meta-cognitive knowledge. This model builds on the Hayes and Flower and Bereiter and Scardamalia L1 models, while taking account of the differences in proficiency level among L2 writers. Grabe and Kaplan's model stresses the importance of gathering and organizing information, and the mediation between the memory and what is chosen to be written and organized is determined by *goal setting*. Goal setting is affected by the context as represented by the reader, the purpose of writing, the writing genre, and the organizational plan.

### 2.2.2 Writing process studies

Most L2 cognitive writing studies have focused their analysis on certain types of writing processes only, such as analysing revision types or planning stages. For instance, planning behaviour has been studied in relation to writers' background knowledge, experience and skill, as in Cumming (1989). Cumming studied the writing behaviour of twenty three ESL writers who were classified into three groups according to their writing expertise in their mother tongue, French. Writers were asked to think aloud while responding to three writing tasks: a letter, an expository piece, and a summary of a booklet. Cumming (1989, pp.114-115) found that the writers' level of expertise affected their attention to gist and discourse organization. Differences in writing processes were identified across writing tasks, and across writers possessing different levels of expertise and with varying writing background knowledge. For example, the two writers described as advance planners who 'carefully thought out the content of each of their compositions before they began to write their texts' were experienced in the areas of technical writing and preparation of technical reports, whereas the emergent planners, who 'planned each composition as it was emerging from the page' had literary/creative writing backgrounds.

The effect of writers' proficiency level and writing experience on planning behaviour has been studied by Sasaki (2000), who compared three groups of writers: 1) expert vs. novice writers, i.e., professors of applied linguistics who write research papers vs. college freshmen who had received little L2 writing instruction; 2) more- vs. less-skilled novice writers, categorized according to marks awarded on an argumentative writing assignment; and 3) novices before and after six months of

instruction on process writing. Sasaki (2000) found that the expert writers did more global planning before starting to write, in contrast to the novices, who planned less before starting to write and stopped while writing to plan what to write next. The fact that writers differ in their planning behaviour according to their skill level can be linked with how revision behaviour is skill-dependent too (Zamel, 1982, 1983). Zamel (1982, 1983) found that skilled writers focus their revision process on the meaning level while writing and postpone text-level revisions (grammar and vocabulary) until the end of the writing session, whereas less-skilled writers keep interrupting their writing to make lexical and grammatical changes.

In terms of the effect of writing process on writing quality, by studying the writing behaviour of fifty one adolescent writers while thinking aloud, De Milliano et al. (2012) found that writers who planned more produced texts of higher quality. These writers were asked to write an untimed persuasive piece similar to the ones they write for their school assignments. Pre-writing planning affected writing quality statistically significantly. Also, paying more attention to formulation, as in translating ideas into linguistic form, resulted in statistically significantly better texts. However, monitoring, evaluation and revision did not have a statistically significant effect on text quality. Another researcher looking at the relationship between writing process and product in a timed context and as a part of a portfolio process is Worden (2009), who coded L1 juniors' eight hundred and ninety texts for planning and revision activities. Worden (2009) investigated the relationship between pre-writing and revision and text quality through having students write two essays and documenting their pre-writing and revision behaviours in their portfolio. This documentation included: 1) pre-writing as it appeared in outlines, mind mapping, and notes written on exam papers, and 2) revision as it appeared in the visible changes made, divided into global revision, which included revisions crossing sentence boundaries, and local revision, where writers revised within the sentence. In general, Worden's (2009) findings showed that a high amount of pre-writing planning resulted in longer, higher-quality essays. A large amount of revision, on the other hand, resulted, in most cases, in lower-quality essays; essays which were not revised at all were of higher quality. Generally speaking, both studies (De Milliano et al., 2012 and Worden, 2009) showed that pre-writing can affect writing quality positively, while revision can have either no effect (de Milliano et al., 2012) or a negative effect (Worden, 2009).

Time allocation for writing processes as a whole has rarely been studied, although Roca de Larios et al. (2008) is an exception, providing insights into how three different proficiency levels of EFL writers spent their composing time during a think-aloud study. Roca de Larios et al. found that most time was spent on formulation (i.e. converting thoughts/ideas into language), while the least time was spent on evaluation. Also noteworthy is that the higher the writers' proficiency level, the more balanced writers' time allocation for different writing processes: higher proficiency writers dedicated less time to formulation compared to the less skilled writers (around 60% of their

composing time rather than around 80%). Higher proficiency writers also dedicated more time to evaluation, revision and planning.

Other studies examined the occurrences of different writing processes and distinguished between writers according to their writing profiles (or *writing signatures* in Levy and Ransell's (1996) terms). Hayes and Flower (1980) distinguished between Planners and Revisers: Planners plan their writing extensively before putting it in its final form and revise their writing sentence by sentence, whereas Revisers write their ideas on paper very quickly and then revise in a later stage. Following Hayes and Flower's definitions, Tillema et al. (2011) also divided writers into Planners and Revisers. They aimed to determine to what extent writers are aware of their own writing profile by asking them to complete a writing-profile questionnaire before they started writing four argumentative essays in their L1, Dutch. The questionnaire consisted of questions that determine whether the writers believe themselves to spend more time on planning or revision during their writing. Writers were then asked to think aloud while responding to writing tasks and their writing processes were recorded using *Inputlog* (see Leijten & Van Waes, 2013), a keystroke logging software package. By dividing each writing session into three 'episodes', and by studying how Planners and Revisers behaved during each writing episode, Tillema et al. (2011) were able to describe the writers' processes. As expected, writers associated with different profiles distributed their writing processes differently: Planners planned and read the assignment more during the first episode while Revisers re-read the assignment more towards the end of the writing session. However, Planners and Revisers did not differ in the overall distribution of time dedicated to text production, reading the assignment, and revision.

Breetvelt et al. (1994) examined the relationship between process occurrences (i.e. how many times a process occurs in a writing stage) and writing product, analysing the writing processes and products of twenty ninth-grade students thinking aloud while responding to two essay tasks. Breetvelt et al. (1994) divided each writing session into three writing stages by dividing the time spent on each writing session into three equal parts. The relationships between text quality and cognitive activities seemed to depend on when a writing activity was employed; a cognitive activity might be positive when employed at a certain stage while having a negative effect when employed in another episode. For example, reading the assignment was positively related to the text quality when occurring in the first episode, but not when occurring in other episodes. A positive effect was identified from goal setting, generating of ideas, and structuring during the second episode, whereas a negative effect resulted from revision, pausing and giving comments during the same episode. Text quality benefited from self-instruction, goal setting, writing, and re-reading in the final episode.

The effect of the writing process as a whole on text quality was studied by Van Weijen et al. (2009) who investigated writers' use of their L1 (Dutch) when thinking aloud while writing in their L2 (English) by analysing their writing processes while writing four essays in their L1 and four other essays in their L2 in response to 8 tasks of varying levels of difficulty. Results supported the proposition that when writers face a



difficult task they tend to change their writing processes and use L1 more, and therefore what affects writing quality is not the use of L1 but rather the task difficulty itself. What interests us in this study was the finding that task difficulty might change the writing process; the more difficult the task is, the more meta-comments are made in writers' L1.

The foregoing discussion highlights that research to date supports the contention that writers usually retain the same writing profile across different writing situations. This led us to the following research questions and hypotheses:

RQ1. Is there a difference in the time allocation of different writing processes in test and non-test situations?

H1. There will be a different distribution of writing processes in test and non-test situations.

RQ2. Is there a difference in the quality of the writing product in test and non-test situations?

H2. There will be no statistically significant difference in the writing product in terms of quality between test and non-test situations.

RQ3. Do writers allocate different amounts of time to different composing activities at different stages of the composition process and does this allocation differ between test and non-test situations?

H3. Writers will allocate different amounts of time to different composing activities at different stages of the composition process.

H4. The amounts of time allocated to different composing activities at different stages of the composition process will differ between test and non-test situations.

RQ4. Is there a relationship between writing process and product quality?

H5. There will be a relationship between distribution of writing processes and resulting text quality.

### 3. Methodology

In order to address the above research questions, a mixed-methods approach was adopted. Ten L2 master's students at a UK university wrote essays in test and non-test situations, making the total dataset twenty essays. The writing sessions were recorded using keystroke logging and screen capture software. The keystroke logging program, *Inputlog* (<http://www.inputlog.net/>), was used to measure the time allotted to each writing process type. The screen capture program, *Screen Movie Studio* (<http://www.mandsoft.com/>), was used to inform the stimulated recall data. The first

author conducted individual stimulated recalls immediately after each writing session. The participants were asked to view and reflect on the screen-capture video and to report verbally on their writing process.

### 3.1 Participants

The ten participants in this study were international master's students at a UK university from different L1 backgrounds studying various linguistics-related programmes. Participants' proficiency level was determined by their IELTS score (which they had taken as a pre-entry requirement). For basic information on participants, see Table 1, which shows their experience of academic writing as a part of their undergraduate courses, and their test training prior to taking their IELTS exams and prior to their arrival in the UK.

### 3.2 Writing tasks

The writing tasks chosen were intended to avoid the topic-familiarity effect, since topic familiarity can have a considerable impact on writers' texts and ratings (see Cumming, Kantor, and Powers, 2002; Tedick, 1990). Moreover, writers may use different composing processes according to their level of familiarity with the assigned topic (Cohen, personal communication, 2012). Therefore, since our aim was to study the effect of test and non-test situations rather than topic effects, we chose non-academic tasks to lessen the chances of some participants being more knowledgeable regarding these topics than others. Since participants were studying in a foreign country having temporarily left their families to do so, the topics chosen were studying abroad and living alone (see Appendix A for prompts 1 and 2). The tasks are available online for IELTS exam practice<sup>1</sup>.

The prompts were of the typical IELTS task 2 type, being argumentative and asking test-takers in the test condition to write at least two hundred and fifty words in forty minutes with no access to online materials. In contrast, in the non-test situation, participants had unlimited time and access to online resources. To avoid order effects, participants did the tasks over the two sessions as described in Table 2. The time the participants spent on each session is also included.

*Table 1.* Participant information

Name <sup>a</sup>	Nationality	IELTS Writing Score	Academic Writing Experience	Test Taking Training	Previous Major before Arrival in the UK
Bella	Russian	6.5	Four courses, each three months	Two-month IELTS preparation course	BA Linguistics
Caterina	Italian	6.5	Three courses, each three months	Two-month IELTS preparation course	BA English literature
Eduardo	Spanish	6	Eight courses, each three months	Two-month IELTS preparation course	BA Linguistics
Gamze	Turkish	6	Five courses, each three months	Four-month IELTS preparation course	BA Linguistics
Kenan	Syrian	6	Six courses, each five months	Self-preparation using IELTS preparation book for almost two months	MA English literature
Lina	Syrian	6.5	Five courses, each three months. Plus intensive one-year course on academic writing.	Nine-month IELTS preparation course	MA language and linguistics
Miho	Japanese	6 <sup>b</sup>	One course, five months	Three-month TOEFL preparation course	BA Linguistics
Rami	Syrian	6 <sup>b</sup>	Five courses, each three months	Self-preparation using TOEFL published materials for a month	BA English literature
Reem	Syrian	6.5	Five courses, each three months. Plus intensive One-year course on academic writing.	Nine-month IELTS preparation course	MA language and linguistics
Tala	Iranian	6.5	Two courses, each three months	Two-month IELTS preparation course	BA Linguistics

**Note:** IELTS= International English Language Testing System.

<sup>a</sup> All participants' names are pseudonyms.

<sup>b</sup> IELTS scores approximately converted from TOEFL iBT writing component for comparison. The original score for both Miho and Kenan on the TOEFL iBT writing was 23<sup>2</sup>. The institute accepts both IELTS and TOEFL scores. The research was granted formal ethical approval by our institution, and participants signed a consent form allowing them the right to withdraw from the study at any time with no obligation to state their reasons for doing so.

**Table 2.** Writing sessions

	Task Condition and Prompt		Time Spent in each Session in Minutes	
	1 <sup>st</sup> session	2 <sup>nd</sup> session	Test	non-test
Bella	test (prompt 1)	non-test (prompt 2)	40	50
Caterina	non-test (prompt 1)	test (prompt 2)	35	30
Eduardo	test (prompt 2)	non-test (prompt 1)	35	40
Gamze	test (prompt 2)	non-test (prompt 1)	30	45
Kenan	test (prompt 2)	non-test (prompt 1)	40	60
Lina	test (prompt 1)	non-test (prompt 2)	40	54
Miho	non-test (prompt 2)	test (prompt 1)	30	45
Rami	non-test (prompt 2)	test (prompt 1)	40	120
Reem	non-test (prompt 1)	test (prompt 2)	40	75
Tala	non-test (prompt 1)	test (prompt 2)	30	30

A dependent *t*-test shows that the time spent on each writing session is statistically significantly higher in the non-test situation ( $M= 54.9$ ,  $SD= 26.5$ ) than it is in the test situation ( $M= 36$ ,  $SD= 4.5$ ),  $t(9)= -2.4$ ,  $p=.03$ .

### 3.3 Scoring

Since the tasks were adopted from IELTS preparation materials, it made sense to use the IELTS analytic scoring rubric, which is available online on the official IELTS website ([http://www.ielts.org/pdf/UOBDS\\_WritingT2.pdf](http://www.ielts.org/pdf/UOBDS_WritingT2.pdf)), to rate the writing. Three raters who work in a university were asked to blindly evaluate the texts, and were not informed whether the text to be rated had been written under test or non-test conditions. All are PhD holders teaching IELTS preparation courses and have experience grading writing proficiency tests. As previous research demonstrated the importance of making test-takers aware of the scoring procedure (Morozov, 2011), we made it clear to writers that the rubric to be used for grading their texts was the one used in IELTS tests, and we sent them a link to these criteria before data collection began.

### 3.4 Procedure

To compare the writing process and product in test and non-test situations required the use of relatively unobtrusive and non-reactive methods, i.e. keystroke logging and screen capture, as well as stimulated recall interviews and analysis of the texts produced.

#### 3.4.1 Keystroke logging

The Keystroke logging program Inputlog was the main data elicitation tool. This program opens a Word file for the participants to write in and the data was then examined with the help of the program's analysis facility. Although various types of

data can be generated from the analysis facility, our needs were satisfied by the program's general analysis data. We required a breakdown of the time allocated to various parts of the writing process (i.e. including planning, revision, and writing), and the general analysis data provided information on the length of time spent pausing, deleting, writing and inserting, in addition to time spent browsing the Internet and details of the sites visited. Using Inputlog made the breakdown procedure an easy one as it is clear from the general analysis the program provides where, for example, a planning segment or a pause begins. However, to classify all writing processes we also examined the interview data (as can be seen from the coding scheme in Appendix E). It should be noted that although Inputlog provides a playback facility, this specific facility did not suit the present study because it does not show Internet browsing and sites visited; the Inputlog playback facility shows only the Word file writers compose.

#### **3.4.2 Screen capture video**

While the writing sessions took place, writers' screens were captured using *Screen Movie Studio*, which is a screen-capturing tool. This tool is interview-friendly in that it stimulates participants' recall through 'visualization' (Park & Kinginger 2010, p.34).

#### **3.4.3 Observation**

While the writers were composing their texts, the first author was in the same room, observing their behaviour, and sharing their screens using *Skype*. By opening a Word document next to the shared screen, she was able to take notes on what she saw which informed the stimulated recall interview questions and allowed her to conduct the interview immediately after the writing sessions ended.

#### **3.4.4 Stimulated recall interviews**

Immediately after each writing session, we conducted a stimulated recall interview which was audiorecorded and transcribed. Previous research (e.g., Boshier, 1998) has used pre-selected parts of the composing video to show to participants, but this was avoided here for a number of reasons. First, we felt watching the whole video would better facilitate writers' recall; second, participants' responses were less likely to be as heavily influenced by the researchers' understanding of the writing process as they would have had the researcher alone been able to pause the video; and third, playing the video in its entirety meant the writers were able to pause the video wherever they felt the need to elaborate on a point. Hence, the interviews were able to focus on excerpts which interested both the writer and the researchers. We readily concede, of course, that Boshier's (1998) selective approach saves time and helps minimize participant/interviewer fatigue when the writing sessions are long and is apt when the researcher wants to investigate a specific issue. However, the exploratory nature<sup>3</sup> of this

study and the fact that writing sessions were of a manageable length for our purposes, lasting around 40-60 minutes (with the exception of one writer who took two hours to complete the untimed task) made watching the whole video a reasonable choice<sup>4</sup>. While watching the video, writers were asked questions such as:

Q-You paused for quite a while at the beginning of the session. What were you thinking of?

Q-You deleted this word more than once. What was the reason for that?

A list of non-specific stimulated recall questions that were asked in addition to writer-specific questions which varied from writer to writer can be found in Appendix B<sup>5</sup>. Gass and Mackey (2000) suggest that conducting stimulated recall

"has an advantage over a simple interview in that the [interview] relies heavily on memory without any prompts and it has an advantage over think-aloud protocols in that for think-alouds, the researcher needs to train participants, and even after training, not all participants are capable of carrying out a task and simultaneously talking about doing the task" (p.18)

Additionally, while thinking aloud provides information on how writers write stimulated recall allows us to probe into the reasons behind writing behaviour. Although the think aloud procedure can be combined with keystroke logging to explore writing profiles by looking at the differences in the distribution of writing processes over time (Tillema et al., 2011), stimulated recall interviews can allow us to investigate the same issues less intrusively, an essential consideration where timed writing is involved, as we will discuss further below.

#### **3.4.5 Coding the writing sessions**

To answer the first research question, "Is there a difference in the time allocation of different writing processes in test and non-test situations?", data gathered from the keystroke logging and screen capture programs along with observation notes were analysed. The relevant data from the stimulated recall interviews were also used to clarify the reason for pausing (such as planning wording and planning ideas) and revision (such as revising on the word or sentence level). The coding scheme identifies sixteen writing processes in the non-test situation and fourteen in the test situation (see Appendix E for a detailed description of the coding scheme). All twenty writing sessions, ten in test and ten in non-test situations, were analysed. As previous research using keystroke logging uses a threshold of two seconds for investigating pausing behaviour (Baaijen et al., 2012), all pauses that exceeded two seconds were considered meaningful and were investigated in the stimulated recall. To clarify the coding scheme, we present an example of how we analysed part of a writing session.

**Example of a coded writing session**

In the following, we provide a sample of a coded writing session written in the non-test situation. It starts at the beginning of the writing session (at the moment the writer opens the writing task on his PC) and ends when the writer finishes writing the introduction to his text.

- Superscript numbers refer to the relevant category of the coding scheme.
- Underlined words were not revised or deleted later.
- Words in capitals are our explanation of the process and words in lower case not underlined were revised either immediately or later.

[OPENS THE WRITING TASK 30 SECONDS PAUSING <sup>1</sup>] [OPENS WEB PAGES AND TYPES THE PROMPT IN IT <sup>2</sup>] [GOES BACK TO THE WORD FILE AND STARTS WRITING <sup>16</sup>] Studying abroad is an issue that <sup>16</sup> have <sup>8</sup> [DELETES have <sup>8</sup>] has been important for a while <sup>16</sup>. However, opinions <sup>15</sup> [PAUSES FOR 10 SECONDS <sup>7</sup>] [MOVES CURSOR AND DELETES issue AND REPLACES IT WITH case AND THEN PROCEEDS TO WRITE AT THE LOCATION OF OPINIONS <sup>11</sup>] on this issue are <sup>16</sup> [PAUSES FOR 6 SECONDS <sup>3</sup>] contradictory. [OPENS ONLINE DICTIONARY AND CHECKS FOR SYNONYM FOR TRAVELERS FOR 40 SECONDS AND GOES BACK TO THE WORD FILE TO PROCEED TO WRITE <sup>4</sup>] Immigrants are <sup>16</sup> [PAUSES FOR 20 SECONDS <sup>5</sup>] not usually treated in an equal way to natives. [READS FROM THE BEGINNING TO EVALUATE WHETHER HIS WRITING MAKES SENSE, PAUSES FOR 20 SECONDS <sup>6</sup>]. [DELETES However opinions on this issue are contradictory AND REPLACES IT WITH I personally regret studying abroad as it is not worth it <sup>15</sup>] ...

It will be noted from the above sample that after reading the task, the writer paused for thirty seconds and then checked online resources before writing anything. After that the writer started writing then revised on the surface level (grammatical mistake). The writer proceeds to write what will later be changed on the meaning level. The writer evaluates his text on the surface level for 6 seconds to make a distant change on the word-meaning level. The writer continues writing and stops shortly afterwards to plan how to express his ideas but then opens an online dictionary to check for a better wording. The writer writes a few more words then stops to generate ideas and then continues writing, before stopping once again to generate ideas. At the end of the introduction the writer makes changes on the meaning level.

After all the writing sessions had been analysed by the first author, 10% of the data were co-rated by a linguistics PhD student. The initial agreement rate was 90% (Cohen's Kappa 0.85). Most of the disagreement concerned revising, mainly word 1, and therefore the definition for this category was expanded for clarity. At first, we were

not clear whether to consider spacing, indenting, and line formatting changes and changes due to initial typos, but after discussion we decided that they could be considered as word 1 revision. Moreover, the first draft of the coding scheme included separate categories for online resources, divided into dictionary and general information resources. However, after discussion we found that writers used online resources as a part of their planning stage, and we therefore added online processes into the appropriate planning categories. The final agreement rate was 96% (Cohen's Kappa 0.95).

In order to compare the differences/similarities in time allocation in test and non-test situations, the quantitative measures from these data were analysed statistically by SPSS using a dependent *t*-test which compares two means that have come from the same participant. The third research question, "Do writers allocate different amounts of time to different composing activities at different stages of the composition process and does this allocation differ between test and non-test situations?", meant that we needed to divide the time spent on each writing task. Following Tillema et al. (2011), we divided each writing session into three stages: stage 1, stage 2, and stage 3, with time beginning to elapse from the point at which the task starts. We used the five major writing process activities we identified in the data (Appendix E): planning, meaningful revision, surface revision, evaluation and translation. A sixth preparation-to-write stage was added to the planning stage because in all cases a preparation-to-write stage included reading the prompt and making a general plan. This was done because preparation to write occurred only once (as its definition in Appendix E suggests) during the first writing stage. In order to avoid inter-writer variability of time spent on each task, we calculated the time spent on each writing activity during each writing stage as a percentage of the total time spent on each writing stage: for example a session that lasts for 55:32 minutes would be divided into three parts, each consisting of 18:44 minutes; we then calculated how much time the writer spent planning, for example, during stage 1, and found it was 10 minutes, expressed as 54.22% of the total writing activities during this part. For the purpose of our analysis we used the generalized linear model which analyses the relationship of several factors per person (writing stages, writing processes, and writing situations).

#### **3.4.6 Coding the stimulated recall data**

The data gathered from the stimulated recall interviews were used to examine the differences between the writing processes during test and non-test situations. Table 3 presents the coding scheme we developed to enable this. Since the aim of the interviews was to investigate the reason behind pausing and changes made, we focused the coding scheme on planning and revision, as can be seen below. The revision behaviour was then coded according to the same scheme used to analyse keystroke logging data, i.e. immediate and distant words 1 and 2, immediate and distant sentences, meaningful and surface revisions, as defined previously.



*Table 3.* Coding scheme for stimulated recall data

<b>Planning Wording</b>	Writers pausing because they are thinking of how to say what they want to say.
<b>Planning Ideas</b>	Writers pausing because they are thinking of what to say next.
<b>Revision</b>	Writers talking about their revision behaviour and their reasons for revising.

Ten percent of the data were coded independently by a second rater, a PhD student in linguistics, and the inter-rater agreement was 95%.

### 3.4.7 Scores analysis

Three trained raters scored the writing using the IELTS task 2 analytic rubric. As suggested by Hamp-Lyons (1990), two raters scored all the writing, and only when there was a mismatch was the third rater used. Raters were asked to comment on each essay in order to provide data explaining the reasons for awarding texts a certain mark. In addition, raters were asked to keep a record of their disagreements and how they solved them. One major disagreement between raters came about when marking an unfinished essay (written under test conditions). The fact that the piece was incomplete left a poor impression on one rater, who marked the writing down for all sub-scores, unlike the second rater's more favourable mark. The third rater suggested a mark similar to the second rater and the third marker's recommended mark was adopted after discussion. After resolving disagreements, the rate of agreement regarding the essay ratings was high, being .92.

The overall scores and the scores for each section of the analytic rubric were tabulated and analysed using a dependent *t*-test, which compared pairs of scores that come from each participant; which in our case are the IELTS writing component scores in test and non-test situations. To answer the fourth research question, "Is there a relationship between writing process and product quality?", we performed a multiple regression analysis for the major six writing processes as predictors of each writing score in turn for test and non-test situations separately.

## 4. Findings and Discussion

We here report and then discuss the results of quantitative analyses of the data collected from the keystroke logs, screen capture videos, observations, stimulated recalls, and raters' evaluation of the texts. We begin by presenting results related to the writing process. We then present analysis of the data related to the grading of the writing.

#### 4.1 Writing process

This section is divided into three sub-sections. We first provide an overall picture of the comparison between the writing processes in test and non-test situations. This analysis covers preparation to write, planning, evaluation, meaningful revision, surface revision, and translation. The sub-categories related to pausing and revision behaviour, which are grouped together in the overall picture, are discussed at greater length in the second and third sub-sections. The fourth sub-section focuses on the differences between the three writing stages across test and non-test situations.

##### 4.1.1 Overall picture of the writing process

This sub-section will first give an overall description of how test takers constructed their texts in test and non-test situations. As mentioned earlier, sixteen process types were identified in the twenty writing sessions, grouped into five types, as outlined in the framework previously described. The descriptive statistics for the use of each process type across test and non-test situations are shown in Figure 1 below. The data in this section are standardized according to the total task time.

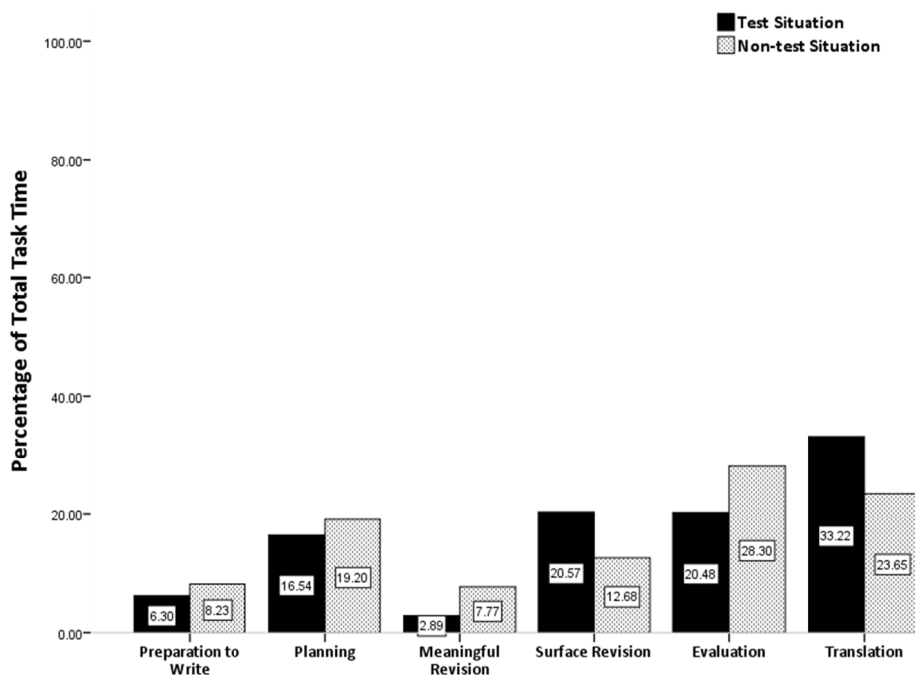


Figure 1. Overall writing process in test and non-test situations.

Figure 1 illustrates how the six writing processes occurred to a different extent in test and non-test situations. The activity which accounted for the largest percentage of total task time in test situations is translation, as opposed to evaluation in non-test situations. Meaningful revision in non-test situations accounted for more than double the time proportionally it was deployed in test situations. In contrast, in non-test situations, surface revision features far less than in test situations. In general, we can say that writers focus more on the level of meaning while revising in non-test situations and on the surface level in test situations.

The results of the dependent *t*-test revealed four statistically significantly different groups of data: surface revision, meaningful revision, evaluation and translation, as can be seen from Table 4.

**Table 4.** Results of the dependent *t*-test for the overall writing process in test and non-test situations

	Test situation		Non-test situation		<i>t</i> (9)	<i>sig.</i>
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>		
Preparation to write	6.3	4.4	8.2	7.5	-1.1	.29
Planning	16.5	9.6	19.2	6.9	-.7	.47
Meaningful revision	2.8	3.3	7.7	4.9	-3.5	.006*
Surface revision	20.5	4	12.6	2.3	5.5	.000*
Evaluation	20.4	6.2	28.3	7.4	-2.5	.02*
Translation	33.2	7.8	23.6	12.4	3.6	.005*

**Note:** \* Statistically significant at  $p < .05$ .

#### 4.1.2 Pausing

Pausing behaviour may feature in the following process categories<sup>6</sup>: offline preparation to write, online preparation to write, offline planning wording, online planning wording, planning ideas, meaningful evaluation, and surface evaluation. Investigating pausing behaviour provides further insights into the results in the previous section and into the use of online resources. Figures 2 and 3 show the difference in the time allocated for pausing in test and non-test situations. The data are again standardized according to the total pausing time.

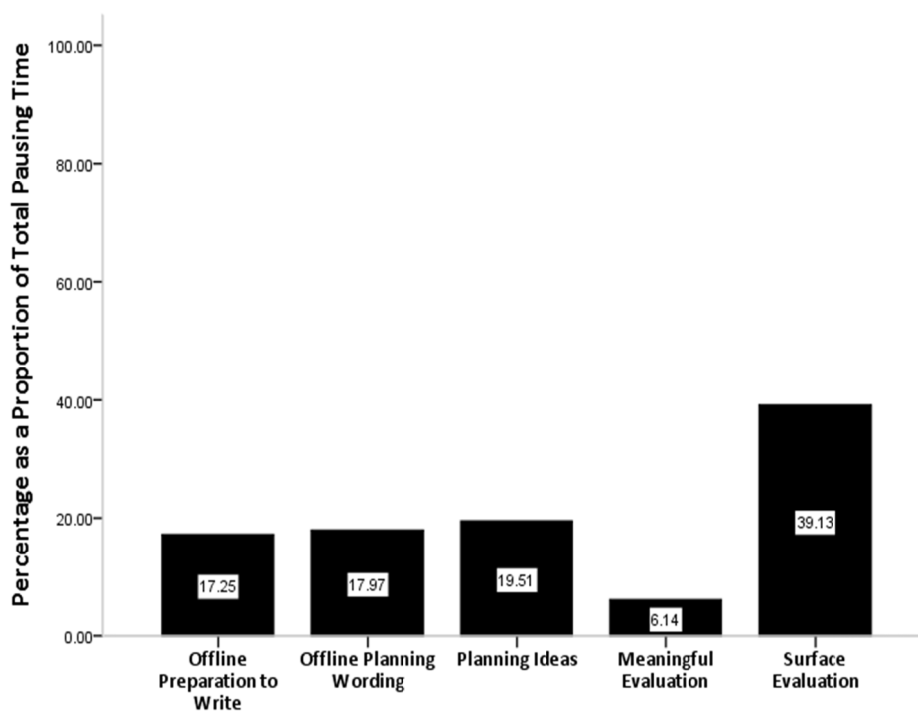


Figure 2. Pausing behaviour in the test situation.

Figures 2 and 3 show that the major reason for pausing was surface evaluation in both situations, although meaningful evaluation accounted for more than double the proportion of pausing time in non-test situations than in test situations. In the non-test situation, writers allocated about twice as much time to offline preparation to write as they did to online preparation to write (around 4% and 8% respectively). Surprisingly, planning ideas was allocated more time in test than in non-test situations with a difference of more than 8% of the total pausing time.

Looking at overall pausing behaviour, we can conclude that the general trend for the writers in test situations is to plan their ideas and wording before they write, while in non-test situations they plan less but spend more time on meaningful evaluation.

The results of the dependent *t*-test revealed four statistically significantly different groups of data: planning ideas, meaningful evaluation, online and offline preparation to write when compared to offline preparation to write in the test situation, as shown in Table 5.

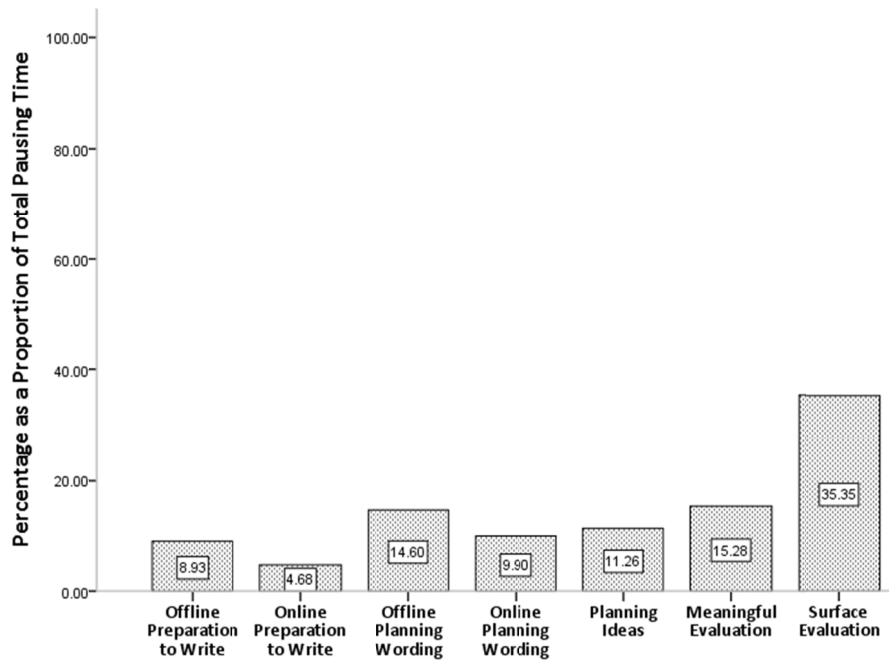


Figure 3. Pausing behaviour in the non-test situation.

Table 5. Results of the dependent t-test for pausing behaviour in test and non-test situations

	Test situation		Non-test situation				t(9)	Sig.
	M	SD	Offline		Online			
			M	SD	M	SD		
Preparation to write	17.2	8.1	8.9	6.3	4.6	8.7	<sup>a</sup> 3.1 <sup>*</sup>	<sup>a</sup> .01 <sup>*</sup>
							<sup>b</sup> 3.2	<sup>b</sup> .01 <sup>*</sup>
Planning wording	17.9	10.5	14.6	9.8	9.9	9.7	<sup>a</sup> 1.4	<sup>a</sup> .17
							<sup>b</sup> 1.8	<sup>b</sup> .10
Planning ideas	19.5	5.6	11.2	3.08			3.9	.004 <sup>*</sup>
Meaningful evaluation	6.1	5.6	15.2	11.3			-2.2	.04 <sup>*</sup>
Surface evaluation	39.1	12.2	35.3	15.1			.68	.51

Notes: \* Statistically significant at  $p < .05$ .

<sup>a</sup> Relationship between the writing process (e.g. preparation to write, planning wording, etc.) in the offline test condition and the online non-test condition.<sup>7</sup>

<sup>b</sup> Relationship between the writing process (e.g. preparation to write, planning wording, etc.) in the offline test condition and the offline non-test condition.

#### 4.1.3 Revision

Investigating revision behaviour gives insights into the level of changes (word or sentence) and the location of changes (immediate or distant). Word 1 level is concerned with grammatical accuracy, spelling, capitalization, word form, gender, punctuation, abbreviation, typos, spacing, indents, line format, etc., while Word 2 level involves changing the word for another because it expresses the idea more clearly although preserving the same meaning. When the changes are applied to more than one word but the meaning of the sentence remains unchanged, i.e. paraphrasing, then this was classified as revision on the sentence level. Changes in the meaning at the sentence level are considered as meaningful revision; this includes changing, adding or deleting ideas.

Figure 4 gives an overview of the types of revisions in test and non-test situations. The data are standardized in proportion to the total time spent on revision.

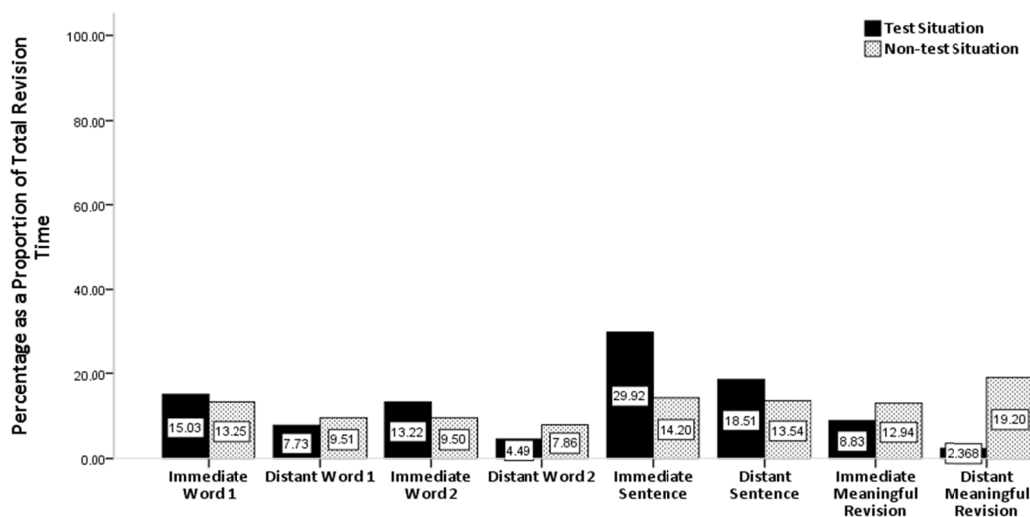


Figure 4. Revision behaviour in test and non-test situations.

The two most dramatic differences concern the time allocated to immediate sentence revisions, which is almost double the time in test as opposed to non-test situations, and the time allocated for distant meaningful revisions, which is dramatically higher in non-test situations. In general, it can be observed that writers spent a greater proportion of their time making immediate revisions at the sentence level in test situations while, in non-test situations, they spent more time revising at a higher level (i.e. at the meaning level). Writers also spent a greater proportion of time revising immediate words in test

rather than non-test situations, while the reverse was the case when it came to distant word revisions, non-test situations producing proportionally more revisions.

By looking at Table 6, we can see that the results of the dependent *t*-test revealed three statistically significantly different groups of data: distant word 2, immediate sentence and distant meaningful revision.

In short, writers employed various writing processes to a different extent in test and non-test situations. Writers conducted more surface, immediate revision in the test situation and more meaningful, distant revision in the non-test situation. More evaluation was also conducted in the non-test situation and a greater proportion of non-test evaluation was on the meaning level.

One interesting finding is that writers engage in more planning of ideas in the test situation. We conclude that writers write and then meaningfully revise in the non-test situation while they carefully plan before they write in the test situation, as they are worried that they do not have enough time to try out and then revise different ideas.

**Table 6.** Results of the dependent *t*-test for revision in test and non-test situations

	Test situation		Non-test situation		<i>t</i> (9)	Sig.
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>		
Immediate Word 1	15	4.2	13.2	5.9	.79	.44
Distant Word 1	7.7	6.1	9.5	6.5	-1.7	.11
Immediate Word 2	13.2	8.2	9.5	9.9	1.4	.17
Distant Word 2	4.4	2.8	7.8	4.5	-2.1	.05*
Immediate Sentence	29.9	7.9	14.2	4.5	5.9	.000*
Distant Sentence	18.5	11	13.5	13.7	1.2	.25
Immediate Meaningful Revisions	8.8	13.4	12.9	9.04	-.88	.40
Distant Meaningful Revision	2.3	5.7	19.2	16.3	-3.7	.004*

**Note:** \* Statistically significant at  $p < .05$ .

#### 4.1.4 Writing stages

Our third research question, "Do writers allocate different amounts of time to different composing activities at different stages of the composition process and does that

allocation differ between test and non-test situations?", can be answered by looking at Figure 5 below.

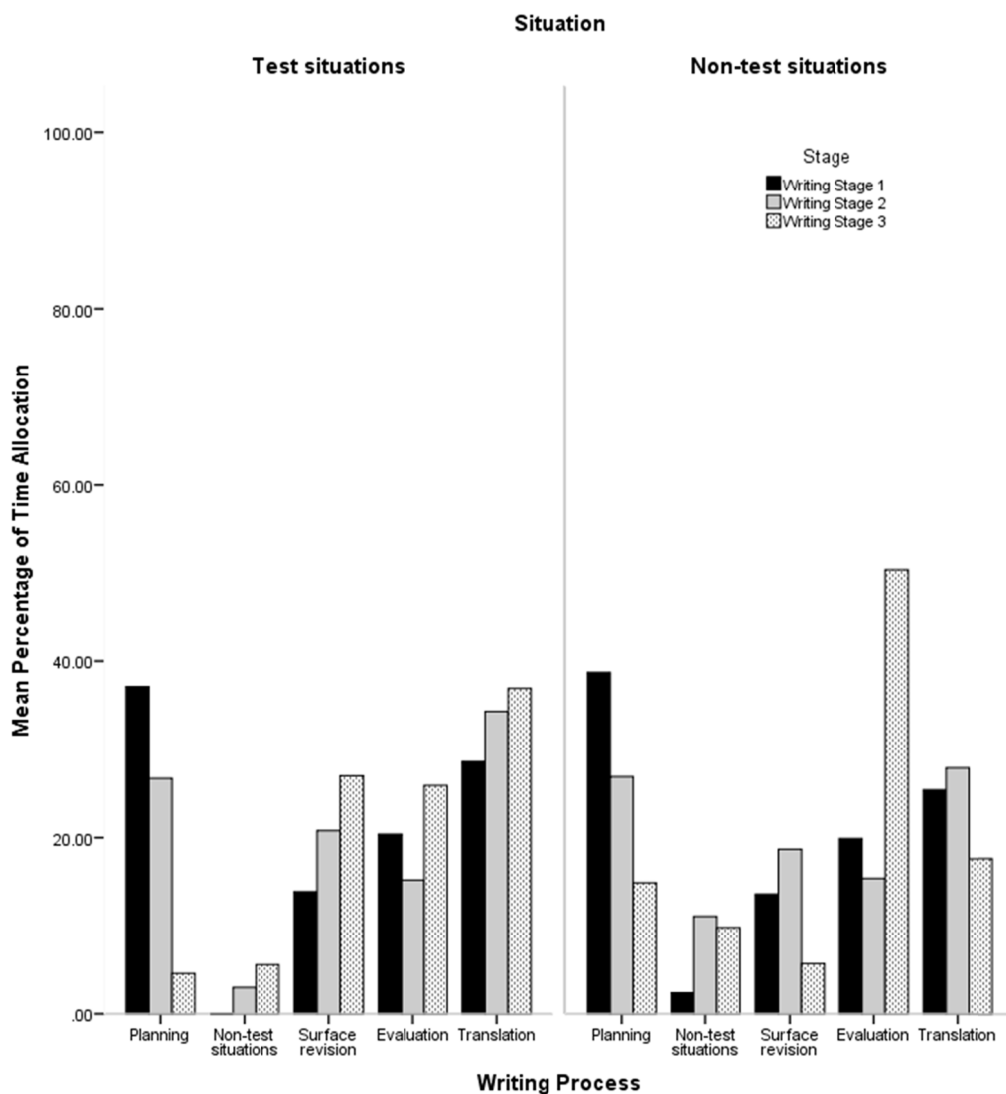


Figure 5. Writing processes in stage 1, stage 2, and stage 3 in test and non-test situations.

The figures reveal a similar writing pattern across both situations with some differences. In the first writing stage, meaningful revision occurred in the non-test situation, while it



did not occur at all in the test situation. Also, writers performed more surface revision in the third writing stage of the test situation than in the non-test situation. Translation occurred more in the test than the non-test situation during all writing stages. The results of the generalized linear model analysing the three way interaction between writing processes, writing stages, and writing situations were statistically significant ( $\chi^2=717.6$ ,  $df= 8$ ,  $p= .00$ ); the difference between processes related to situation is statistically significant depending on the writing stage.

Next we split the data we have by stage in order to see where the differences between the test and non-test situations are most marked. Figures 6, 7, and 8 show us the differences in time allocation for each writing process.

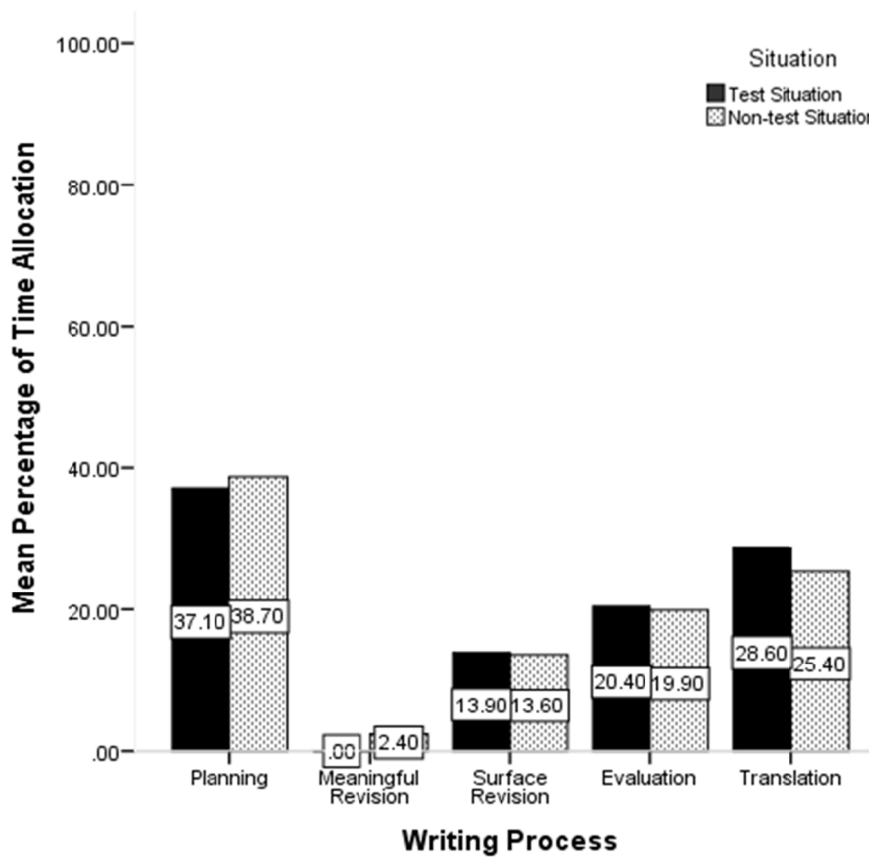


Figure 6. Time allocation for writing processes during the first writing stage.

Writers spent almost the same time on planning, surface revision, and evaluation across test and non-test situations. Similarly, the differences in time spent on meaningful revision and translation are slight, at between 2-3% across test and non-test environments. This suggests to us that writers started responding to the task in the same way in both test and non-test situations.

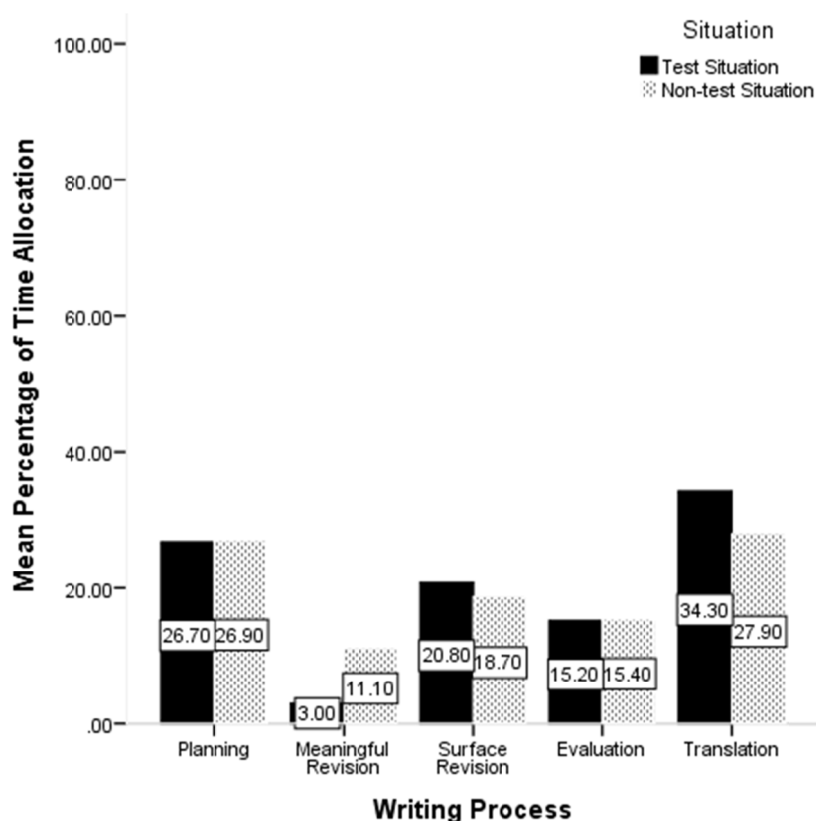


Figure 7. Time allocation for writing processes during the second writing stage.

During the second writing stage there is once again a similarity in the time spent on planning, surface revision, and evaluation in both situations. However, the figure for meaningful revision in the non-test situation is over 8% higher than in the test situation. And in the latter situation, writers spend more time focusing on producing their ideas in linguistic form (translation) than in the non-test situation.

Figure 8 vividly captures the marked differences in the application of writing processes under both conditions in the third stage. Writers now plan much less in the test situation, while spending more time engaging in meaningful revision than in the

previous stages, although they continue to spend more time on meaningful revision in the non-test environment. With reference to the non-test condition, despite falls in the time engaged in planning and meaningful revision compared to the second stage, the time writers devote to these processes is still appreciably larger than in the test situation. Writers focus their revision behaviour on the surface level in the test situation (27%) while devoting less time to surface revision in the non-test situation (5.7%). The highest percentage of time allocation in the non-test situation is devoted to evaluation (more than 50%). Therefore, it seems that writers spend the extra time available in evaluating their writing, while in the test situation they are still writing their ideas and making revisions related to grammar and vocabulary choice as seen in the higher figures for translation and surface revision in the test situation.

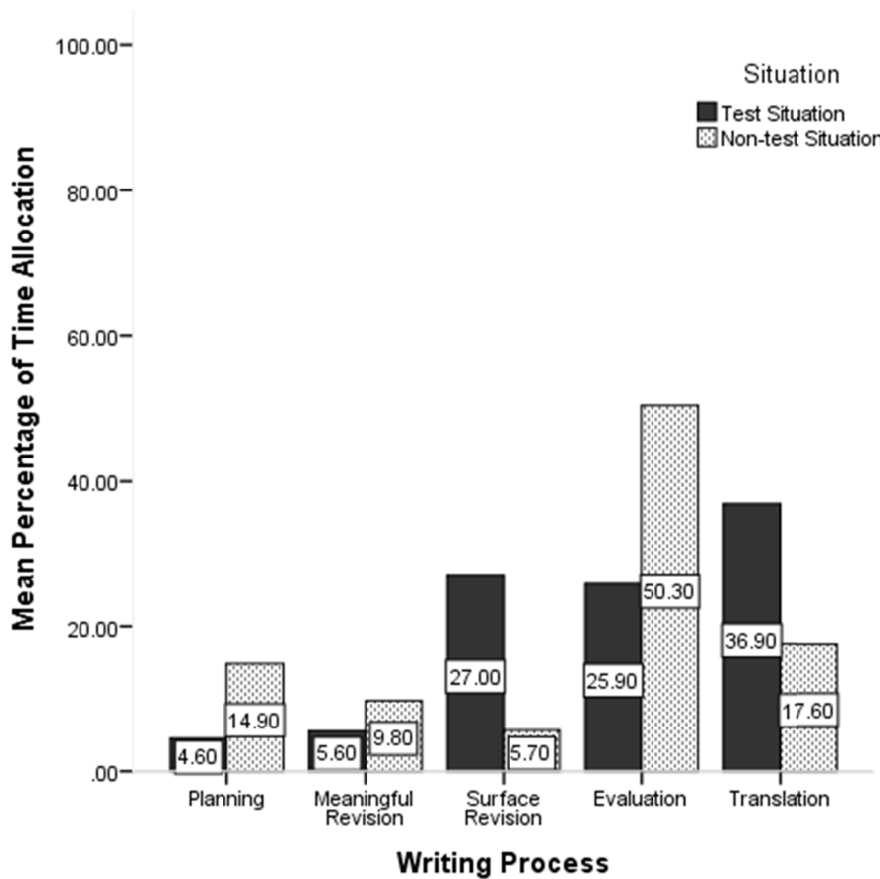


Figure 8. Time allocation for writing processes during the third writing stage.

The results of the generalized linear model show non-statistically significant differences between test and non-test writing in patterns of time allocation to processes at stage 1, large and highly statistically significant differences at stage 2 ( $\chi^2=20$ ), and very marked and highly statistically significant differences at stage 3 ( $\chi^2=190$ ).

We performed a follow up analysis using time allocation for each writing process as a percentage of the total task time to check which processes were allocated more time during each of the three stages. We split by stage and process and did dependent *t*-tests on situation only in generalized linear model with Bonferroni adjustment. The results do not show any statistically significant differences between any of the processes across test and non-test situations in the first and second writing stages. In the third stage, writers allocated a statistically significantly higher amount of time to evaluation ( $p=.005$ ); writers largely used the extra time they had in comparison to the timed condition for evaluation. Also, writers did statistically significantly more planning in the third writing stage in the non-test situation; writers had more time to think of including and planning new ideas at the final stage of their writing ( $p=.005$ ). On the other hand, they did more surface revision and translation in the test situation ( $p=.005$  for both).

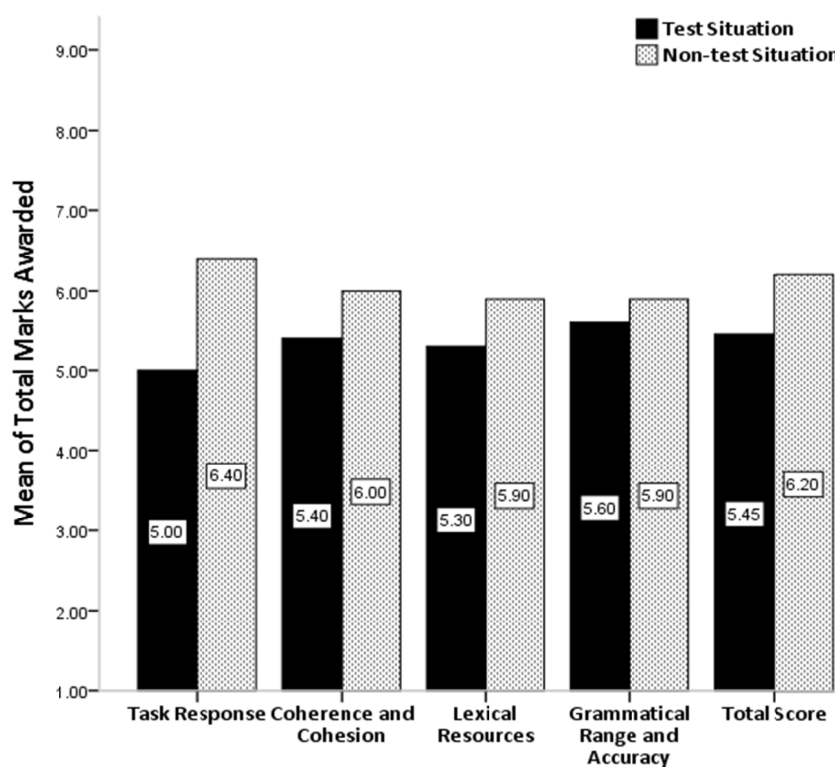


Figure 9. Marks awarded in test and non-test situations.

## 4.2 Writing product

This section focuses on the rating of the texts composed in the test and non-test situations. Using the IELTS analytic rubric allowed raters to give an overall mark for each text, as well as evaluations of the texts' task response, coherence and cohesion, lexical resources, and grammatical range and accuracy. Figure 9 provides a summary of the marks awarded to the texts. (See Appendix C for the text quality coding scheme, and Appendix D for a sample of a marked text.)

It can be seen that the mean score of the total marks awarded to the non-test writing is almost one point higher. When looking at the sub-categories measured by the analytic rubric, the major difference between the marks awarded concerns task response, which is almost one and a half point higher for non-test writing. The non-test writing was also scored more highly with regard to the other analytic criteria (coherence and cohesion, lexical resources and grammatical range and accuracy).

The results of the dependent *t*-test showed four statistically significantly different groups of data: Task response, coherence and cohesion, lexical resources and total score, as can be seen in Table 7.

**Table 7.** Results of the dependent *t*-test for marks awarded in test and non-test situations

	Test situation		Non-test situation		<i>t</i> (9)	Sig.
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>		
Task Response	5.0	1.7	6.4	.69	-2.5	.02*
Coherence and Cohesion	5.4	.96	6.0	1.05	-2.7	.02*
Lexical Resources	5.3	.94	5.9	.99	-2.2	.05*
Grammatical Range and Accuracy	5.6	.51	5.9	.56	-1.9	.08
<b>Total Score</b>	5.4	.77	6.2	.63	-3.5	.006*

**Note:** \* Statistically significant at  $p < .05$ .

In general, there were statistically significant differences in the total marks awarded. Statistically significant differences were found also when comparing marks awarded for cohesion and coherence, lexical resources, and task response. An approaching statistically significant difference was found when comparing the grammatical range and accuracy. All of this addresses our second research question, which asks whether there is a difference in the writing product in test and non-test situations. Writers in our study were able to make use of the additional resources and time afforded them in the non-test condition to produce writing which was superior lexically and grammatically. The non-test writing was also judged to be more coherent and to be superior in terms of task response.

### 4.3 Writing process and product

To answer the fourth research question, "Is there a relationship between writing process and product quality?", we performed a multiple regression analysis on the six major writing processes: preparation to write, planning, meaningful revision, surface revision, evaluation, and translation. We analysed the effect of each writing process on the total score and on each writing sub-score (task response, coherence and cohesion, lexical resource and grammatical range and accuracy) in test and non-test situations. There were no statistically significant relationships between total scores and any writing process, nor to any writing process in the non-test condition. Our results were not statistically significant in relation to the effect of any process in the non-test situation on the sub-scores. However, some of the processes had an impact on sub-scores in the test situation; translation affects both task response and cohesion and coherence scores with an almost statistically significant relationship with task response ( $t=2.691$ ,  $p=.055$ ) and a statistically significant relationship with cohesion and coherence ( $t=3.323$ ,  $p=.029$ ). Meaningful revision and surface revision also have near statistically significant correlations with coherence and cohesion ( $t=2.649$ ,  $p=.057$ ;  $t=2.496$ ,  $p=.067$ ).

A post hoc analysis involved calculating a difference score between each writer's test and non-test figures for all writing sub-scores and doing a regression analysis with the positive and negative scores (for example, if a writer's mark for task response in the test situations is 7 and 5 in the non-test situations then the difference score would be +2. Here a positive score shows where figures were higher for test than non-test; negative the reverse). This showed statistically significant or near statistically significant effects of preparation to write on task response ( $t=3.404$ ,  $p=.042$ ), grammar ( $t=2.851$ ,  $p=.065$ ) and total score ( $t=3.897$ ,  $p=.030$ ). This means that the greater the increase in the time devoted to preparation to write in the non-test situation over the test situation, the greater the improvement in score for the relevant feature. There were also two other near statistically significant, but negative, relationships with task response: planning ( $t=-3.002$ ,  $p=.058$ ) and surface revision ( $t=-2.584$ ,  $p=.082$ ). This means that writers who spent more extra time on planning and surface revision in non-test over test situations would only see relatively lesser improvements in task response scores.

In a word, in the test situation, cohesion and coherence is affected by translation, meaningful revision, and surface revision; task response is affected by translation. We do not have a clear effect of writing process on product quality in the non-test situation.

## 5. Discussion

This study investigates whether there is a difference in the writing process and product in test and non-test situations. Statistically significant differences were detected in the writing process across the two situations; namely in meaningful revision, surface revision, evaluation, translation, online and offline preparation to write, planning ideas,

meaningful evaluation, distant word 2 revision, immediate sentence revision, and distant meaningful revision. Differences in the time allocated to various writing processes across the two situations were found mainly in the third writing stage. A statistically significant difference was also found in the writing product across both situations, a result which contradicts the findings of previous research; statistically significant differences were found in task response, coherence and cohesion, lexical resources, and total score awarded. However, a correlation between the writing process and product quality showed that there is no relationship between time allocation for writing processes and product quality. These findings are discussed below with reference to each of our research questions and hypotheses.

RQ1. Is there a difference in the time allocation of different writing processes in test and non-test situations?

H1. There will be a different distribution of writing processes in test and non-test situations.

We expected that there would be a different distribution of writing processes over time in both situations in line with research by Curry (2004) and Sasaki (2000). The hypothesis was confirmed by looking at each writing process. This can be clarified by comparing the amount of time writers in both situations devoted to the various process categories; in the test situation translation was dominant (i.e., writers were translating their ideas into linguistic form). This finding parallels the findings of earlier L1 (e.g. Kellogg, 1987, 1988; Levy & Ransdell, 1995; Penningroth & Rosenberg, 1995) and L2 studies (e.g. Roca de Larios et al., 2008; Plakans, 2008; Wang & Wen, 2002). In the non-test situation, evaluation came to the fore, a finding that is not in line with previous research: in previous L2 studies, evaluation was allocated just 1% to 5% of the total task time (e.g. see Roca de Larios et al., 2008; & Sasaki, 2000, whose participants wrote in timed, test-like conditions, allocated one hour and thirty minutes respectively). Our writers used the extra time to evaluate their texts and this makes sense to us as the writers wrote almost the same text length in both situations (typically ranging between 300 and 400 words) and thus used the extra time for evaluation.

Looking at the pausing time in both situations, one can see that the participants' evaluation behaviour was mainly directed towards the surface level, i.e., checking for word level accuracy and appropriacy. Earlier studies of L2 writers' revision behaviour (e.g., Gaskill, 1986; Porte, 1996, 1997; Sengupta, 2000) also found that writers' major concern was at the word level. Porte (1996) argued that the reason is likely that writers have been taught to give priority to the word level, and that they have understood that they are marked only on word use rather than meaning and content.

A difference was found between time allocation for planning of words in both situations. Writers allocate more time to plan their words in non-test situations as can be shown by adding online to offline planning wording which constitutes more than

23% of the pausing time in the non-test situation compared to 17% in the test situation. Writers were more concerned with planning their words in the non-test situation, while in the test situation they were more concerned with planning their ideas. This can be related to the effect of using dictionaries as can be seen by the time spent on online planning wording by means of dictionary usage (13% of the total pausing time in the non-test situation). Dictionaries keep writers busy at the word level (Skibniewski & Skibniewska, 1986), but this relates to the often-cited drawback of dictionary use, namely that it “reinforce[s] the belief in a one-to-one relationship at word level between two languages” (Thompson 1987, p.282).

As for the level of revision, it is interesting to note that writers revised more on the meaning level in the non-test situation. Previous research (e.g. Hayes, 1996; Porte, 1996; Zamel, 1983) related revision behaviour to skill: more-skilled writers focus on the meaning level while less-skilled ones focus on the word level. This study shows how revision behaviour might also be linked to the writing situation: despite the fact that these writers' task schema includes attending to meaning-related problems, they did not activate this behaviour as much in timed situations.

Concerning the location of the revisions, writers did not differ in the time allocated for distant surface revisions across the two situations. However, they spent statistically significantly more time revising in the distant word 2 and distant meaningful categories in the non-test situation. Therefore, writers undertook a more recursive and non-linear writing process in the non-test situation. Severinson-Eklundh and Kollberg (2003) have described how a linear text is one that is produced more or less at the first time of asking, with few distant revisions, while producing a non-linear text involves a lot of distant revisions. We add to Eklundh (1994), who found that the writing process became more recursive as task difficulty increases, that recursiveness may also increase when extra time and materials are available.

RQ2. Is there a difference in the quality of the writing product in test and non-test situations?

H2. There will be no statistically significant difference in the writing product in terms of quality between test and non-test situations.

Contrary to our hypothesis, and to the findings of several previous studies which studied writing in test and non-test situations, or in situations where writers were afforded lesser or greater amounts of time to compose (Caudery, 1990; Kenworthy, 2006; Kroll, 1990; Polio et al., 1998), we found a statistically significant difference in the writing products of test and non-test situations as judged by raters. However, this finding corresponds to that of Hall (1991), who also found that there were some marked differences in the writing product in test and non-test situations.



A statistically significant difference was found in the raters' judgements on the coherence and cohesion of the texts produced under the two conditions. Writers wrote texts with a more sophisticated progression of ideas and a better use of cohesive devices in the non-test situation. Previous research (e.g. Polio et al. 1998) did not report on this finding. This might be due to the rubric they were using not assessing these aspects of authorship. However, it seems that the time spent on translation, meaningful revision, and surface revision under time pressure had an impact on the product quality. By focusing their composing time on translation in the test situation, the writers were able to express their ideas more coherently. Coherence seems to have also been improved by both types of revision in the test situation. Although previous research showed that revision does not inevitably improve the overall text quality or the grammatical and lexical sophistication levels (de Milliano et al., 2012; Worden, 2009), our research does not necessarily contradict this as examining the effect revision has on coherence and cohesion was neglected in previous research. Therefore, our study adds to previous research (e.g. de Milliano et al., 2012 & Worden, 2009) that time pressure might have resulted in writers using some processes (translation, meaningful revision, and surface revision) more effectively.

Also corresponding to previous research (e.g., Caudery, 1990; Kenworthy, 2006; Polio et al., 1998) is the finding of an approaching statistically significant difference in the quality of the product produced under the two conditions on the grammatical level. However, two other findings contradict previous work: (i) the finding that non-test texts were rated statistically significantly higher on the lexical resource level (cf. East, 2008; and Kobayashi & Rinnert, 1992, which found that using dictionaries did not have an effect on participants' scores); and (ii) the finding that some writers made use of the extra time available by writing statistically significantly higher-level texts (cf. Crone et al., 1993; Hale, 1992; Livingston, 1987; Powers & Fowles, 1996).

RQ3. Do writers allocate different amounts of time to different composing activities at different stages of the composition process and does this allocation differ between test and non-test situations?

H3. Writers will allocate different amounts of time to different composing activities at different stages of the composition process.

H4. The amounts of time allocated to different composing activities at different stages of the composition process will differ between test and non-test situations.

Our hypotheses were partially confirmed. Looking at writers' writing profiles by analysing their writing behaviour across three writing stages, we found that writers do not actually change their time allocation for different writing processes at the beginning of the writing session, but then start changing it gradually to make the most statistically significant change in the third stage, where writers pause more in the non-test situation

(as reflected in planning and evaluation) while they revise and translate more in the test situation. Previous L1 and L2 research showed that writers' writing profiles do not change across different writing sessions; this means that they do not change if the writing situation remains the same (Levy & Randsell, 1996). As we examined the writing processes across different situations, our finding that writers changed their writing behaviour only in the third stage does not necessarily contradict these other studies. It might be that writers started their writing session in the non-test situation the same way they did in the test situation but differed only in the final stage which, as we discuss below, can be seen as an attempt to spend more time on the task.

Although the writers changed their writing processes statistically significantly in the third stage by devoting about 50% of their time to evaluation, it is questionable how effective this evaluation stage was, as their evaluation behaviour rarely led to revision. This might indicate poorly-developed evaluation skills. Another explanation might be that writers were actually satisfied with what they had written and therefore no revision was actually needed. Therefore, in both cases, evaluation was not productive, and this suggests to us that writers were actually only trying to spend more time on a task as they felt they should make use of the extra time available but did not know how. After finishing their task, writers felt hesitant to hand in their writing as they were worried that they "might regret it later" (Gamze). However, they mostly did not make noticeable changes to it.

RQ4. Is there a relationship between writing process and product quality?

H5. There will be a relationship between distribution of writing processes and resulting text quality

Based on various studies (e.g., Breetveld et al, 1994; Braaksma et al, 2004; Van den Bergh & Rijlaarsdam, 1999, 2001) we expected that there would be a relationship between distribution of writing processes and resulting text quality. However, our analysis showed a rather mixed picture, with some processes appearing to have an impact on the writing product while others did not. The post hoc analysis showed producing texts with higher grammatical and lexical levels was not associated with revision, but rather surface revision had a negative effect on task response in the non-test situation. We might infer from this result that when writers focus on the word level, they forget about, or at any rate neglect, the overall requirements of a task. Higher scores for grammar and overall performance are associated with spending more time on the preparation to write stage; the pre-writing stage was found to affect the product quality in general and maybe the more time spent on the pre-writing stage can result in producing ideas featuring a more complex sentence structure. In fact the finding that the pre-writing stage can affect the overall score of the writing product is a finding that corresponds with previous research (de Milliano et al., 2012; Worden, 2009). However, planning seems to affect the task response negatively in the non-test situation. It seems

that writers tended to over problematize the writing prompt and this resulted in forgetting about or neglecting the main requirements of the task.

It seems that it is under time pressure with a lack of resources that extra time spent on specific processes has an impact. This means that spending more time on a writing process may not necessarily produce a higher-quality text; what matters is whether the extra time is used effectively. Limiting time tended to result in writers using it more effectively. However, we should note that, despite the fact that the writers did not make use of processes effectively, they nonetheless scored higher in the non-test situation. This suggests to us that maybe it is the concept of time and not time itself that affected writers' scores. Our thinking is as follows: When writers were given more time, they scored more highly even though their writing processes were more effective in the test situation. Therefore, what affected writers' marks was not time itself, but rather the idea that they had limited time in the test situation; having more time in the non-test situation was a relief for writers and perhaps resulted in better text quality even though time was not used effectively as far as the application of several writing processes was concerned.

## 6. Conclusion

This study showed that giving writers more time increases the quality of their texts in a meaningful way, with mean scores of almost one point higher awarded to the non-test condition texts. As we hypothesized, when it comes to writers' writing profile, writers did not differ in their planning and revision behaviour across both situations and during different writing stages. However, they differed in their evaluation as they evaluated more in the third stage of the non-test situation, a finding that contradicts what we expected. Nonetheless, it is not a surprising one as it makes sense that writers would use that extra time for re-reading what they have written. A surprising finding not in line with our predictions is the statistically significant difference in the quality of texts produced in both situations (probable reasons for this are discussed in the discussion section above). A finding that adheres to expectations is the statistically insignificant relationship between time spent on writing processes and writing product; spending more time on a writing process did not result in a better text. Writing processes were mostly used more effectively under limited time.

Finally, more research on the area of writing in test and non-test conditions is necessary with more participants included to test the generalizability of our findings. A deficiency in the current study is that the writers were not responding to an authentic test and this might have affected their motivation to respond to the prompts; therefore, a study where the tasks are part of a real test might remedy this. Also, replicating the study with a larger number of tasks is important as this will show us whether writers' products and processes remain consistent over more than two tasks. Time as a concept (i.e. the idea of having limited time) had an effect on writers and this suggests writing

teachers would do well not only to explicitly raise awareness of and teach composing strategies but also how to cope with different writing situations, as well as perhaps teaching some techniques that would help apprehensive writers to reflect their true writing ability under test conditions, such as outlining. Training writers to write in test, timed conditions might also lessen their writing apprehension by getting them used to such conditions. We can also conclude by pointing out how we need to not only teach writing as a process (stressing the idea of planning and revising) but also to teach our students how to use these strategies appropriately, since doing more planning did not necessarily produce a better text.

### Notes

1. The prompts are available online at:
  - First prompt: <http://www.bestieltsonline.com/a-sample-ielts-essay-question/>
  - Second prompt: [http://www.ielts-exam.net/ielts\\_writing\\_samples\\_task\\_2/634/](http://www.ielts-exam.net/ielts_writing_samples_task_2/634/)
2. The comparison was made using the comparison table provided by the TOEFL official site: [http://www.ets.org/s/toefl/pdf/supplementary\\_comparison\\_tables.pdf](http://www.ets.org/s/toefl/pdf/supplementary_comparison_tables.pdf)
3. The exploratory nature of this study lies in the interview data which is not in focus here (due to the limited scope of this paper) but is the focus of another paper (Khuder & Harwood, in preparation).
4. Watching the whole video was used with all participants except for one participant (Rami) who composed for two hours. In his case, we selected segments ourselves where the data seemed interesting in addition to asking him to nominate segments for discussion himself.
5. Some of the questions included in Appendix B focus on probing more deeply into writers' motivations for their writing behaviour, which is outside our area of focus in the present paper. Results on this aspect of the study are reported in Khuder & Harwood (in preparation).
6. It should be noted that online preparation to write and online planning wording do not occur in the test situation, as writers were not allowed to access Internet resources in the test situation, contrary to the non-test situation.
7. As we allowed for the use of online resources in the non-test condition, some of the writing processes included both offline and online data (online planning wording and offline planning wording); while in the test condition the test condition included data only for offline writing processes. This table shows the analysis of the relationship between test offline and non-test online activities as well as the relationship between test offline and non-test offline activities.

### Acknowledgements

We wish to thank Phil Scholfield for his help with the statistics in this article.

This project is also uploaded on [WritingPro](#) an open access repository on writing process research.

## References

- Baaijen, M.V., Galbraith, V. & de Glopper, K. (2012). Keystroke analysis: reflections on procedures and measures. *Written Communication*, 29, 246-277. doi: 10.1177/0741088312451108
- Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice*. Oxford: Oxford University Press.
- Bereiter, C., & Scardamalia, M. (1987). *The psychology of written composition*. Hillside, NJ: Lawrence Erlbaum.
- 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, 205-241. doi: 10.1016/S1060-3743(98)90013-3
- Breetvelt, I., Van den Bergh, H., & Rijlaarsdam, G. (1994). Relations between writing processes and text quality: when and how? *Cognition and Instruction*, 12(2), 103-123. [http://dx.doi.org/10.1207/s1532690xci1202\\_2](http://dx.doi.org/10.1207/s1532690xci1202_2)
- Caudery, T. (1990). The validity of timed essay tests in the assessment of writing skills. *ELT Journal*, 44, 222-231. doi: 10.1093/elt/44.2.122
- Crone, C., Wright, D., & Baron, P. (1993). Performance of examinees for whom English is their second language on the spring 1992 SAT II: Writing Test. *Unpublished manuscript prepared for ETS*. Princeton, NJ: Educational Testing Service.
- Cumming, A. (1989). Writing expertise and second language proficiency. *Language Learning*, 39, 81-141. doi: 10.1111/j.1467-1770.1989.tb00592.x
- Cumming, A., Kantor, R., & Powers, D. (2002). Decision making while rating ESL/EFL writing tasks: a descriptive framework. *The Modern Language Journal*, 86, 67-96. doi: 10.1111/1540-4781.00137
- Curry, M. J. (2004). UCLA Community College review: Academic literacy for English language learners. *Community College Review*, 32 (2), 51-68. <http://dx.doi.org/10.1177/009155210403200204>
- De Milliano, I., van Gelderen, A., & Slegers, P. (2012). Patterns of cognitive self-regulation of adolescent struggling writers. *Written Communication*, 29(3), 303-325. <http://dx.doi.org/10.1177/0741088312450275>
- Durgunoglu, A. Y. & Verhoeven, L. (1998) Epilogue: Multilingualism and literacy development across different cultures. In A. Y. Durgunoglu & L. Verhoeven (Eds.). *Literacy development in a multilingual context: Cross-cultural perspectives* (289-298). Mahwah, NJ: Lawrence Erlbaum Associates.
- East, M. (2008). *Dictionary use in foreign language writing exams: Impact and implications*. Amsterdam/Philadelphia, PA: John Benjamins. <http://dx.doi.org/10.1075/llt.22>
- Faigley, L., Daly, J.A., & Witte, S.P. (1981). The role of writing apprehension in writing performance and competence. *Journal of Educational Research*, 75(1),16-21. <http://dx.doi.org/10.1080/00220671.1981.10885348>
- Fulcher, G. (2010). 'Open book - open web tests'. [Languagetesting.info](http://languagetesting.info). Retrieved from: <http://languagetesting.info/features/open/book.html> access [17/July/2012]
- Gaskill, W.H. (1986). Revising in Spanish and English: a process-oriented study of composition. Unpublished PhD Thesis, University of California, Los Angeles.
- Gass, S. M., & Mackey, A. (2000). *Stimulated recall methodology in second language research*. Mahwah: Lawrence Erlbaum.
- Graham, S., & Harris, K. R. (2000). The role of self-regulation and transcription skills in writing and writing development. *Educational Psychologist*, 35, 3-12. doi: 10.1207/S15326985EP3501\_2
- Graham, S., & Harris, K. R. (2003). Students with learning disabilities and the process of writing: A meta-analysis of SRSD studies. In H. L. Swanson, K. R. Harris, & S. Graham, (Eds.), *Handbook of learning disabilities* (pp.323-344). New York: Guilford Press.
- Grabe, W., & Kaplan, F. L. (1996). *Theory and practice of writing: An applied linguistic perspective*. London: Longman.

- Hale, G.A. (1992). Effect of amount of time allowed on the Test of Written English. *TOEFL Research Report #RR-92-27*. Princeton, NJ: Educational Testing Service.
- Hall, E. (1991). Variations in composing behaviours of academic ESL writers in Test and non-test situations. *TESL Canada journal*, 8, 9-33.
- Hamp-Lyons, L. (1990). Second language writing: Assessment issues. In B. Kroll (Ed.), *Second language writing* (pp.69-86). Cambridge: Cambridge University Press. <http://dx.doi.org/10.1017/CBO9781139524551.009>
- Hamp-Lyons, L., & Condon, W. (2000). *Assessing the portfolio: Principles for practice, theory, and research*. Cresskill: Hampton Press.
- Hayes, J.R., & Flower, L.S. (1980). Identifying the organization of writing processes. In L. W. Gregg and E. R. Steinberg (Eds.), *Cognitive Processes in Writing* (pp.3-30). Hillsdale, NJ: Lawrence Erlbaum.
- Hayes, J. R. (1996). A new framework for understanding cognition and affect in writing. In C.M. Levy & S. Ransdell (Eds.), *The science of writing: Theories, methods, individual differences, and applications* (1-27). Hillsdale, NJ: Lawrence Erlbaum.
- Kellogg, R. T. (1987). Effects of topic knowledge on the allocation of processing time and cognitive effort to writing processes. *Memory and Cognition*, 15, 256–266. doi: 10.3758/BF03197724
- Kellogg, R. T. (1988). Attentional overload and writing performance: Effects of rough draft and outline strategies. *Journal of Experimental Psychology. Learning, Memory & Cognition*, 14, 355–365. <http://dx.doi.org/10.1037/0278-7393.14.2.355>
- Kenworthy, R. (2006). Timed versus at-home assessment tests: does time affect the quality of second language learners' written compositions?' *TESL-EJ*, 10, 1.
- Khuder, B. & Harwood, N. (in preparation). Individual differences in L2 writing: Six L2 writers' writing behaviour in test and non-test situations.
- Kobayashi, H. & Rinnert, C. (1992). Effects of first language on second language writing: translation versus direct composition. *Language Learning*, 42, 183-209. doi: 10.1111/j.1467-1770.1992.tb00707.x
- Kormos, J. (2012). The role of individual differences in L2 writing. *Journal of Second Language Research*, 21 (4), 390-403. <http://dx.doi.org/10.1016/j.jslw.2012.09.003>
- Kroll, B. (1990). What does time buy? ESL student performance on home versus class compositions. In B. Kroll (Ed.), *Second language writing: Research insights for the classroom* (140-154). Cambridge: Cambridge University Press.
- Heijne-Penninga M, Kuks JBM, Schonrock-Adema J, Snijders TAB, CohenSchotanus J. (2008). Open-book tests to complement assessment programmes: Analysis of open and closed-book tests. *Adv Health SciEduc*, 13, 263–273. <http://dx.doi.org/10.1007/s10459-006-9038-y>
- 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. doi: 10.1177/0741088313491692
- Levy, C. M., & Ransdell, S. (1995). Is writing as difficult as it seems? *Memory and Cognition*, 23, 767–779. doi: 10.3758/BF03200928
- Levy, C. M., & Ransdell, S. (1996). Writing signatures. In C.M. Levy & S. Ransdell (Eds.), *The science of writing: Theories, methods, individual differences, and applications* (149–16 1). Mahwah, NJ: Erlbaum.
- Livingston, S.A. (1987). The effects of time limits on the quality of student-written essays. *Paper presented at the Annual Meeting of the American Educational Research Association*, New York, NY.
- Manchón, R. M., (2009). Individual differences in foreign language learning: The dynamics of beliefs about L2 writing. *RESLA*, 22, 245-268.
- Morozov, A. (2011). Student attitudes toward the assessment criteria in writing-intensive college courses. *Assessing Writing*, 16(1), 6–31. <http://dx.doi.org/10.1016/j.asw.2010.09.001>
- Park, K. & Kinginger, C. (2010). Writing/thinking in real time: digital video and corpus query analysis. *Language Learning and Technology*, 14, 31-50.

- Penningroth, S. L., and Rosenberg, S. (1995). Effects of a high information-processing load on the writing process and the story written. *Applied Psycholinguistics*, 16, 189–210. doi: 10.1017/S0142716400007086
- Plakans, L. (2008). Comparing composing processes in writing-only and reading-to-write test tasks. *Assessing Writing*, 13, pp.111-129. doi: 10.1016/j.asw.2008.07.001
- Polio, C., Fleck, C. & Leder, N. (1998). "If I only had more time:" ESL learners' changes in linguistic accuracy on essay revisions. *Journal of Second Language Writing*, 7, 43-68. doi: 10.1016/S1060-3743(98)90005-4
- Polio, C. & Glew, M. (1996). ESL writing assessment prompts: How students choose. *Journal of Second Language Writing*, 5, 25-50. doi: 10.1016/S1060-3743(96)90014-4
- Porte, G. (1996). When writing fails: How academic context and past learning experiences shape revision. *System*, 24, 107-16. doi: 10.1016/0346-251X(95)00056-P
- Porte, G. (1997). The etiology of poor second language writing: The influence of perceived teacher preferences on second language revision strategies. *Journal of Second Language Writing*, 6, 61–78. doi: 10.1016/S1060-3743(97)90006-0
- Powers, D.E. and Fowles, M.E. (1996). Effects of applying different time limits to a proposed GRE writing test. *Journal of Educational Measurement*, 33, 433-452. doi: 10.1111/j.1745-3984.1996.tb00500.x
- Roca de Larios, J., Manchón, R., Murphy, & L., Marín, J. (2008). The foreign language writer's strategic behaviour in the allocation of time to writing processes. *Journal of Second Language Writing*, 17, 30-47. doi: 10.1016/j.jslw.2007.08.005,
- Sanders, S.E., & Littlefield, J.H. (1975). Perhaps test essays can reflect significant improvement in freshman composition: Report on a successful attempt. *Research in the Teaching of English*, 9, 145-153.
- Sasaki, M. (2000). Toward an empirical model of EFL writing processes: An exploratory study. *Journal of Second Language Writing*, 9, 259–291. doi: 10.1016/S1060-3743(00)00028-X
- Sengupta, S. (2000). An investigation into the effects of revision strategy instruction on L2 secondary school learners. *System* 28, 97-113. doi: 10.1016/S0346-251X(99)00063-9
- Severinson Eklundh, K. S., and Kollberg, P. (2003). Emerging discourse structure: Computer-assisted episode analysis as a window to global revision in university students' writing. *Journal of Pragmatics*, 35, 869-891. doi: 10.1016/S0378-2166(02)00123-6
- Severinson Eklundh, K. S. (1994). Linear and nonlinear strategies in computer-based writing. *Computers and Composition*, 11, 203-216. doi: 10.1016/8755-4615(94)90013-2
- Shaw, S.D. and Weir, C.J. (2007). *Examining writing: Research and practice in assessing second language writing*. Cambridge: Cambridge University Press.
- Silva, T. (1993). Toward an understanding of the distinct nature of L2 writing: The ESL research and its implications. *TESOL Quarterly*, 27, 657–677. doi: 10.2307/3587400
- Skibniewski, L., & Skibniewska, M. (1986) Experimental study: The writing process of intermediate/advanced foreign language learners in their foreign and native languages. *Studia Anglica Posnaniensia*, 19, 142-163.
- Tedick, D.J. (1990). ESL writing assessment: subject-matter knowledge and its impact on performance. *English for Specific Purposes*, 9, 123-143. doi: 10.1016/0889-4906(90)90003-U
- Tillema, M., Van den Bergh, H., Rijlaarsdam, G. & Sanders, T. (2011). Relating self reports of writing behaviour and online task execution using a temporal model. *Metacognition Learning*, 6, 229–253. <http://dx.doi.org/10.1007/s11409-011-9072-x>
- Thompson, G. (1987). Using bilingual dictionaries. *ELT Journal*, 41, 282-286. doi: 10.1093/elt/41.4.282
- Uzawa, K. (1996). Second language learners' processes of L1 writing, L2 writing and translation from L1 into L2. *Journal of Second Language Writing*, 5, 271–294. [http://dx.doi.org/10.1016/S1060-3743\(96\)90005-3](http://dx.doi.org/10.1016/S1060-3743(96)90005-3)

- Van Weijen, D., Van den Bergh, H., Rijlaarsdam, G., & Sanders, T. (2009). L1 use during L2 writing: An empirical study of a complex phenomenon. *Journal of Second Language Writing*, 18, 235–250. <http://dx.doi.org/10.1016/j.jslw.2009.06.003>
- Wang, W., & Wen, Q. (2002). L1 use in the L2 composing process: An exploratory study of 16 Chinese EFL writers. *Journal of Second Language Writing*, 11, 225–246. doi: 10.1016/S1060-3743(02)00084-X
- Wachholz, P. B. & Etheridge, C. P. (1996). Writing self-efficacy beliefs of high- and low apprehensive writers. *Journal of Developmental Education*, 19(3), 16-24.
- Weigle, S.C. (2002). *Assessing writing*. Cambridge: Cambridge University Press. <http://dx.doi.org/10.1017/CBO9780511732997>
- Williams, J.B. (2011). Open Book Open Web (OBOW) exams final examination. presentation transcript, Knowledge university, retrieved from <http://www.slideshare.net/jembwilliams/open-book-open-web-obow-exams>. [Accessed 15 June 2012]
- Wolcott, W. (1987). Writing instruction and assessment: The need for interplay between process and product. *College Composition and Communication*, 38, 40-46. <http://dx.doi.org/10.2307/357585>
- Worden, D. (2009). Finding process in product pre-writing and revision in timed essay responses. *Assessing Writing*, 14(3), 157-177. <http://dx.doi.org/10.1016/j.asw.2009.09.003>
- Zamel, V. (1982). Writing: the process of discovering meaning. *TESOL Quarterly* 16, 195-210. doi: 10.2307/3586792
- Zamel, V. (1983). The composing processes of advanced ESL students: Six case studies. *TESOL Quarterly*, 17, 165-188. doi: 10.2307/3586647



## Appendix A: Writing prompts

1- Present a written argument or case to an educated reader with no specialist knowledge of the following topic.

***Nowadays more people are choosing to live with friends or alone rather than with their families. This trend is likely to have a negative impact on communities .***

***To what extent do you agree or disagree with these opinions?***

You should use your ideas, knowledge and experience and support your arguments with examples and relevant evidence.

*Write at least 250 words.*

2- Present a written argument or case to an educated reader with no specialist knowledge of the following topic.

***Nowadays many students have the opportunity to study for part or all of their courses in foreign countries.***

***While studying abroad brings many benefits to individual students, it also has a number of disadvantages.***

***To what extent do you agree or disagree with these opinions?***

You should use your ideas, knowledge and experience and support your arguments with examples and relevant evidence.

*Write at least 250 words.*

## Appendix B: Stimulated recall questions

### Before writing

- 1- Can you tell me about your previous writing knowledge? What kind of writing courses did you attend?
- 2- What do you think of the writing task, easy/difficult?
- 3- What do you think the prompt is asking you to do?
- 4- How did you read the prompt?
- 5- Can you tell me about your knowledge of the topic and whether you found the task difficult?
- 6- How did you plan your ideas before writing?
- 7- The ideas you had before you started writing, did any of them change AS you were writing?
- 8- What did this plan include? How did you choose to divide your ideas?
- 9- What were thinking of when giving this example?

### General questions

- 10- What were you doing?
- 11- Why did you do that?
- 12- Why did you change that word/phase/sentence?
- 13- What was your reason for pausing?
- 14- In order to make changes, how much did you read from the sentence?
- 15- Why did you make these changes?

### After writing

- 16- On a scale of 1-10, where 1 is very dissatisfied and 10 is very satisfied, how happy were you with your performance? Why?
- 17- Were you satisfied with the words you had for your ideas?
- 18- How would you describe your writing experience now?
- 19- To what extent were you confident while writing?
- 20- Did you write the way you usually do in real life? What do you feel is the difference?
- 21- Are there any ideas which you wanted to include but could not because of the lack of vocabulary, structure knowledge? (Only after the test situation session)
- 22- If you could do the writing again, what would you do differently?
- 23- How did the external resources, like online dictionaries and other websites, affected your writing? (Only after the non-test situations session)
- 24- Finally, is there anything else you would like to say about this writing task?

### Appendix C: Text quality coding scheme

Sub-text types	Textual indicators of performance level	
	More basic	More developed
Task response	Misunderstanding of the prompt; no clear position is taken; repeating the same idea. e.g. a writer was unable to complete his text so his argument was unfinished; scoring 4.	Clear position and well-supported ideas. e.g. a writer clearly disagreeing with the task and providing two examples why they do not disagree—however they did not expand on it well; scoring 7.
Coherence and cohesion	Little evidence of logical textual organization (paragraphs, introduction). e.g. a writer's text was not divided into introduction and body paragraphs but rather she wrote the whole text as one block where she narrates her opinion of disagreeing and say in the text on how she is not totally against the idea but then conclude with disagreeing; scoring 4.	Logically better organized texts with effective use of cohesive devices. e.g. a writer wrote a clear introduction that states her position and each of her body paragraphs included a clear idea that was discussed logically but we had an overuse of cohesive devices; scoring 7.
Lexical resource	Use very limited range of vocabulary and make mistakes in spelling and word formation. e.g. part of a 4 score text: <i>"Living alone has many advantages and disadvantages. This phenoomina is widly spread in Europe and some other countries. Boys and girls tend to live out side their houses either with friends or alone."</i>	Uses uncommon vocabulary successfully. e.g. part of a 7 score text: <i>"However, I believe that the advantages far outweigh the disadvantages and these disadvantages could be kept to minimum if students are well-acquainted with the culture norms of the country they intend to travel"</i>
Grammatical range and accuracy	Use of very limited grammatical structure with errors in grammar and punctuation. e.g. part of a 4 score text: <i>'Nowadays lots of student desire to study in a foreign country, in my view this choice brings lots of advantages and some difficulties that are disatvantage for others.'</i>	Use of wide-range structures which are mostly error-free. e.g. part of a 7 score text: <i>"It cannot be denied that at first they will go through some errors and even take vital decisions that may affect their entire life; nevertheless, it is through these trails that they become decisive decision-makers."</i>

## Appendix D: Sample of a written text with marks awarded for each section

*Nowadays many students have the opportunity to study for part or all of their courses in foreign countries.*

*While studying abroad brings many benefits to individual students, it also has a number of disadvantages.*

*To what extent do you agree or disagree with these opinions?*

Nowadays, many students decide to pursue their studies in foreign countries either before graduation or after graduation. I think that studying abroad could be beneficial; however, several issues should be taken into consideration before taking such a decision.

One of the advantages of studying abroad is the fact that students will not only be exposed to other cultures but also the high value that is often associated with being a holder of a foreign certificate. For example, those who have completed their studies in foreign countries like the UK and USA have more opportunities of getting jobs than those who decided to study in their countries.

Another merit of studying in foreign countries could be that students who study abroad have to learn the language of the targeted country; hence, they become multilingual which also could be considered as an important qualification for their future careers.

However, it cannot be denied that there are some drawbacks of studying abroad. Firstly, students will be away from their families, and thus they have to depend on themselves emotionally, financially and socially. For instance, they will be responsible for their cooking, washing and house chores. In addition, studying abroad costs students too much money and, in some cases, students get loans to afford studying abroad. Therefore, they have to work to be able to pay back these loans.

Summing up, being an international student is valued at both the social and the educational level. Nevertheless, it is not an easy decision due to some personal as well as financial problems.

### Total and sub-marks awarded

Sub-sections of analytic rubric	Marks awarded
Task response	7
Coherence and Cohesion	6
Lexical resources	6
Grammatical Range and accuracy	6
<b>Total</b>	<b>6.5</b>

## Appendix E: Coding scheme for writing process data

### Preparation to write

#### 1. Offline preparation to write:

From the time the writer opens the task on computer until the writing session starts. This includes reading the prompt, and pausing. No Internet resources are used.

#### 2. Online preparation to write:

Time spent using Internet resources before the actual writing starts in the non-test situation.

### Planning

#### 3. Offline planning wording:

When writers pause before writing; planning the words that express their ideas.

#### 4. Online planning wording:

When writers pause before writing to use online resources, mainly dictionaries, to check for the appropriate word in the non-test situation.

#### 5. Planning ideas:

When writers pause before writing in order to plan their ideas..

### Evaluation

#### 6. Meaningful evaluation:

When writers read a previous part of the text to make sure the writing is sound in organization and meaning. This might result in revision.

#### 7. Surface evaluation:

When writers read a previous part of the text to make sure the text is grammatically correct. This might result in revision.

### Surface revision

In the revision categories (for both meaningful and surface) below, *immediate* revisions refer to those made in the same sentence as the writer's cursor location, and *distant* revisions refer to those made in any other sentences apart from the sentence the writer was working on.

#### 8. Immediate word 1:

#### 9. Distant word 1:

} At the word level, concerned with grammatical accuracy, spelling, capitalization, word form, gender, punctuation, abbreviation, typo, spacing, indent, line format. e.g. studies → studying

- 10. Immediate word 2:**
- 11. Distant word 2:**
- 12. Immediate sentence:**
- 13. Distant sentence:**
- Meaningful revision**
- 14. Immediate meaningful revision:**
- 15. Distant meaningful revision:**
- Translation**
- 16. Translation:** The writer produces text that is not deleted later on.
- Changing the word for another because it expresses the idea more clearly although preserving the same meaning, e.g. they → other countries.
- When the changes are applied to more than one word but the meaning of the sentence remains unchanged, i.e. paraphrasing, e.g. studying abroad → being a holder of a foreign certificate.
- Changes in the meaning at the sentence level. This will include changing, adding or deleting ideas. This includes starting an idea then deleting it after deciding not to continue with it. e.g. studying abroad has negative and positive effects → I, personally, cannot argue against the idea of living abroad as it has many advantages.