ETS TOEFL

Quality Beyond Measure.

TOEFL iBT® Research Report
TOEFL iBT–25
ETS Research Report No. RR–15-04

Test Takers' Writing Activities During the *TOEFL iBT*® Writing Tasks: A Stimulated Recall Study

Khaled Barkaoui

June 2015

The *TOEFL*® test was developed in 1963 by the National Council on the Testing of English as a Foreign Language. The Council was formed through the cooperative effort of more than 30 public and private organizations concerned with testing the English proficiency of nonnative speakers of the language applying for admission to institutions in the United States. In 1965, Educational Testing Service (ETS) and the College Board assumed joint responsibility for the program. In 1973, a cooperative arrangement for the operation of the program was entered into by ETS, the College Board, and the *Graduate Record Examinations*® (*GRE*®) Board. The membership of the College Board is composed of schools, colleges, school systems, and educational associations; GRE Board members are associated with graduate education. The test is now wholly owned and operated by ETS.

ETS administers the TOEFL program under the general direction of a policy board that was established by, and is affiliated with, the sponsoring organizations. Members of the TOEFL Board (previously the Policy Council) represent the College Board, the GRE Board, and such institutions and agencies as graduate schools of business, two-year colleges, and nonprofit educational exchange agencies.

❖ ❖ ❖

Since its inception in 1963, the TOEFL has evolved from a paper-based test to a computer-based test and, in 2005, to an Internet-based test, the *TOEFL iBT*® test. One constant throughout this evolution has been a continuing program of research related to the TOEFL test. From 1977 to 2005, nearly 100 research l reports on the early versions of TOEFL were published. In 1997, a monograph series that laid the groundwork for the development of TOEFL iBT was launched. With the release of TOEFL iBT, a TOEFL iBT report series has been introduced.

Currently this research is carried out in consultation with the TOEFL Committee of Examiners (COE). Its members include representatives of the TOEFL Board and distinguished English as a second language specialists from academia. The committee advises the TOEFL program about research needs and, through the research subcommittee, solicits, reviews, and approves proposals for funding and reports for publication. Members of the TOEFL COE serve 4-year terms at the invitation of the Board; the chair of the committee serves on the Board.

Current (2014-2015) members of the TOEFL COE are:

Sara Weigle - Chair	Georgia State University
Yuko Goto Butler	University of Pennsylvania
Sheila Embleson	York University
Luke Harding	Lancaster University
Eunice Eunhee Jang	University of Toronto
Marianne Nikolov	University of Pécs
Lia Plakans	University of Iowa
James Purpura	Teachers College, Columbia University
John Read	The University of Auckland
Carsten Roever	The University of Melbourne
Diane Schmitt	Nottingham Trent University
Paula Winke	Michigan State University

To obtain more information about the TOEFL programs and services, use one of the following:

E-mail: toefl@ets.org
Web site: www.ets.org/toefl

ETS is an Equal Opportunity/Affirmative Action Employer.

As part of its educational and social mission and in fulfilling the organization's nonprofit Charter and Bylaws, ETS has and continues to learn from and to lead research that furthers educational and measurement research to advance quality and equity in education and assessment for all users of the organization's products and services.

TOEFL iBT Research Report Series and ETS Research Report Series ISSN 2330-8516

RESEARCH REPORT

Test Takers' Writing Activities During the *TOEFL iBT*® Writing Tasks: A Stimulated Recall Study

Khaled Barkaoui

York University, Toronto, Ontario

This study aimed to describe the writing activities that test takers engage in when responding to the writing tasks in the *TOEFL iBT*® test and to examine the effects of task type and test-taker English language proficiency (ELP) and keyboarding skills on the frequency and distribution of these activities. Each of 22 test takers with different levels of ELP (low vs. high) and keyboarding skills (low vs. high) responded to 2 TOEFL iBT writing tasks (independent and integrated) on the computer. Each participant then provided stimulated recalls about the writing activities they used when performing each writing task. Stimulated recalls were coded and the results were compared across tasks and test-taker groups. The findings indicated that the participants engaged in various construct-relevant activities, such as interacting with the writing task and resources, planning, generating, evaluating, and revising. Additionally, test takers' writing activities varied significantly across tasks and to a lesser extent across test-taker groups. Participants' writing activities varied most across writing tasks and, to a lesser extent, across English proficiency groups. Low keyboarding skills seem to have affected mainly activities on the independent writing task. To better understand the role of keyboarding skills in performance on the TOEFL iBT writing tasks and to address the test's extrapolation inference, future studies need to compare the writing performance of test takers with different levels of second language (L2) proficiency and keyboarding skills in test and nontest settings.

Keywords Academic writing; second-language writing; computer-based writing tasks; stimulated recalls; independent writing tasks; integrated writing tasks; writing activities; TOEFL iBT; keyboarding skills; second-language proficiency

doi:10.1002/ets2.12050

The primary goal of this study was to describe the writing activities that test takers engage in when responding to the independent and integrated writing tasks in the *TOEFL iBT*® test. A secondary goal was to examine whether and how these activities are influenced by test-taker English language proficiency (ELP) and keyboarding skills. The examination of test-taking processes can contribute to a test's validity argument in two ways. First, as Shaw and Weir (2007) argued, examining test-taking processes can help establish whether a second-language (L2) test activates the types of mental processes that a theory of L2 knowledge and performance views as essential elements of L2 performance (cf. Chapelle, 2008; Cohen, 2012; Cohen & Upton, 2007; Cumming, Kantor, Powers, Santos, & Taylor, 2000; Weir, 2005). For example, a test claiming to evaluate academic L2 writing ability would be expected to include tasks calling for test takers to actually use academic L2 writing skills and processes, such as planning, monitoring, and revising, in responding to tasks, rather than to rely on other skills (Cohen & Upton, 2007). To the extent that this is not the case, the explanation inference in the test's validity argument is threatened (Chapelle, 2008). Second, such research can help establish the extent to which the test engages test takers in the same cognitive processes involved in writing in real-life contexts (Cohen, 2012; Shaw & Weir, 2007; Weir, 2005). To the extent that this is not the case, the extrapolation inference is weakened (Chapelle, 2008; Shaw & Weir, 2007).

Following Messick's (1989) distinction between construct-relevant and construct-irrelevant variance in test scores, Cohen (2012) distinguished between construct-relevant and construct-irrelevant test-taking processes and strategies (cf. Cumming et al., 2000).[1] Construct-relevant processes and strategies are those that theory indicates are relevant to the construct being measured. Cohen (2012) called them *test management strategies*, and they include, for example, planning, evaluating, and revising one's text. Test developers expect test takers to engage these construct-relevant processes when responding to test items and tasks. Construct-irrelevant processes and strategies, in contrast, refer to *test-wiseness strategies*. These strategies "involve using knowledge of testing formats and other peripheral information to answer test items without going through the expected cognitive processes" (p. 264). They represent shortcuts that allow the test taker to respond to the test items without actually engaging L2 knowledge and performance ability. As Chapelle (2008) explained,

Corresponding author: K. Barkaoui, E-mail: KBarkaoui@edu.yorku.ca

evidence that writing tasks elicit construct-relevant writing strategies and processes shows that "scores are influenced by relevant psycholinguistic knowledge, processes, and strategies" (p. 338).

Numerous studies have examined the cognitive processes and writing activities that L2 learners engage in and the sources of knowledge they draw upon when writing in L2, as well as the factors that influence these processes and activities (see Roca de Larios, Murphy, & Marin, 2002; Weigle, 2002). Most of this research has been built on theoretical models of first-language (L1) writing, particularly the Hayes and Flower (1980) model and the Scardamalia and Bereiter (1987) model. Hayes and Flower's model of writing consists of three components: the task situation (including the writing task, the writer's motivation, and the text written so far), the writer's long-term memory, and the writing process. Hayes and Flower proposed that writing consists of three major processes: (a) planning, which involves retrieving relevant information from long-term memory and converting the task situation into goals that shape and guide subsequent processes; (b) translating, which consists in converting plans and ideas in the writer's memory into linguistic text; and (c) reviewing, which involves evaluating and improving the quality of the text produced. The writing process is not linear, however. The three major processes can occur in different orders and combinations as the writer juggles various constraints, draws on several sources of knowledge, and attends to several cognitive demands simultaneously while writing.

Scardamalia and Bereiter (1987) proposed two models to describe how unskilled and skilled writers write: knowledge telling and knowledge transforming. Unskilled writers tend to adopt a knowledge-telling approach whereby they generate text by locating topic and genre identifiers in the writing task, which then results in memory probes, that is, the retrieval of relevant and associated content. These writers tend to be occupied by "the activity of finding a next thing to say" (p. 145) rather than with the elaboration of an overall plan and specific goals or the application of pertinent strategies to solve problems. They manage to produce coherent text by retrieving a number of associated ideas that are already available in memory. Skilled writers, in contrast, use a knowledge-transforming approach to writing. This approach is more cognitively complex and demanding and involves problem analysis and solving, goal setting, and consideration of content issues, rhetorical requirements, and the emerging text.

The Hayes and Flower (1980) and Scardamalia and Bereiter (1987) models of writing have been used extensively to frame research on L1 and L2 writing. They also provide a framework for conceptualizing and explaining performance on writing tests. For example, according to the Hayes and Flower model, it is expected that test takers will engage in planning, translating, and reviewing when writing during the test. Consequently, studies aiming to examine a writing test's validity argument need to provide evidence that a writing test does actually engage the linguistic knowledge, processes, and strategies (e.g., planning, translating, reviewing) expected by theory (Cohen, 2012). If it does, then this lends support to the validity of inferences based on test scores about test takers' L2 writing abilities. As a result, the first aim of this study was to describe the type, frequency, and distribution of writing activities that test takers engage in when responding to TOEFL iBT independent and integrated writing tasks. A second gaol was to examine whether and how these activities vary depending on test-taker ELP and keyboarding skills. The TOEFL iBT validity argument rests on an assumption that writing task scores are attributed to a construct of academic ELP (Chapelle, 2008). Since TOEFL iBT is delivered on the computer, test performance might be influenced by test-taker keyboarding skills as well. Keyboarding skills are not part of ELP as currently defined and operationalized by the TOEFL iBT. Consequently, differences in test performance that can be attributed to a test taker's lack of keyboarding skills, rather than lack of ELP, constitute a source of construct-irrelevant variance (Taylor, Jamieson, Eignor, & Kirsch, 1998; Wang, Eignor, & Enright, 2008; Wolfe & Manalo, 2005). This study thus aimed (a) to describe the writing activities that test takers engage in when responding to TOEFL iBT independent and integrated writing tasks and (b) to examine whether and how these activities are influenced by task type, test-taker ELP, and keyboarding skills. The following section reviews relevant literature on the effects of the three variables (L2 proficiency, task type, and keyboarding skills) on L2 writing performance before describing the context, methods, and findings of the current study.

Literature Review

The Effects of Second-Language Proficiency on L2 Writing Processes

According to the Scardamalia and Bereiter (1987) model, writers with different levels of writing expertise approach the writing task differently. Similarly, many studies have shown that writers with different levels of L2 proficiency differ considerably in terms of their L2 writing processes (Cohen, 1994; Cumming, 1989; Hall, 1991; Plakans, 2009; Raimes, 1987; Roca

de Larios, Manchón, Murphy, & Marín, 2008; Roca de Larios, Murphy, & Manchón, 1999; Sasaki, 2000, 2002; Whalen & Ménard, 1995). Raimes (1987), for example, found that high-proficiency students engaged in more planning, rehearsing, rescanning, revising, and editing than did less proficient students. Similarly, Sasaki (2000, 2002) found that L2 proficiency influences L2 learners' writing processes significantly. For example, more proficient writers tended to spend more time planning before starting to write, to plan overall organization in more detail, to write faster and more, to reread or refine their expressions more often, and to pause to think less frequently after making their global plans than did less proficient writers. Less proficient writers, in contrast, tended to make less detailed plans and to stop and plan what they were going to write every time they finished writing one semantically coherent chunk, suggesting that they were employing a "what next strategy." Additionally, as their L2 proficiency improved, students tended to do less local, online planning and more global, prewriting planning and rereading of their texts (Sasaki, 2002).

Roca de Larios et al. (2008) found that more proficient students devoted more time to planning, evaluating, and revising their texts and less time to formulation. Additionally, writing processes were differentially distributed across the writing session depending on the writer's L2 proficiency. Specifically, while low-proficiency students maintained the same pattern of time allocation throughout the writing process, high-proficiency students showed a more diversified time allocation to different processes at different stages of the writing session (e.g., more formulation in early stages and more revision in later stages). These findings suggest that "as proficiency increases, writers appear to be able to strategically decide *what* attentional resources to allocate to *which* writing activities *at which* stages of the writing process" (p. 43). As Whalen and Ménard (1995) explained, low L2 proficiency can lead to more attention to the linguistic aspects of the text, which can impede attention to other textual aspects, idea generation, and upper level processing (cf. Broekkamp & van den Bergh, 1996; Chenoweth & Hayes, 2001).

Another set of studies has investigated the question of the effects of L2 proficiency on L2 writing processes by examining the relationships between L2 learners' writing processes and the quality of their texts as measured by test scores. The findings of these studies are mixed for two main reasons (Alamargot & Chanquoy, 2001; Stevenson, Schoonen, & de Glopper, 2006). First, the relationship between writing processes and scores is indirect, with writing processes and strategies affecting text characteristics and the latter influencing test scores (van der Hoeven, 1999a, 1999b). Second, as Breetvelt, van den Bergh, and Rijlaarsdam (1994) have demonstrated, the relationship between cognitive activities and text quality seems to depend on when the activity is employed during the writing process (cf. Roca de Larios et al., 2008; van der Hoeven, 1999a; van Weijen, van den Bergh, Rijlaarsdam, & Sanders, 2008). For example, Breetvelt et al. (1994) found that generating a higher number of ideas early in the writing process and fewer in the end is positively associated with text quality (cf. van der Hoeven, 1999a). In contrast, the number of revisions is generally negatively correlated with text quality when revisions are made in the first or second third of the writing process; there was no relationship at all in the last third of the writing process. Van Weijen et al. (2008) found that the correlation between reading the assignment and text quality was positive at the beginning and subsequently becomes negative, while for formulation and planning it was negative in the beginning and then became positive from the middle of the writing process onward.

The findings of the studies reviewed above indicate that (a) the same cognitive activity serves different functions at different stages in writing and (b) the occurrence of writing activities varies significantly during the writing process (Breetvelt et al., 1994; Roca de Larios et al., 2008; van der Hoeven, 1999a, 1999b; van Weijen et al., 2008). As Roca de Larios et al. (2008) have shown, rather than being randomly activated and distributed, the cognitive activities that writers engage in depend on (a) the repertoire of strategies at the writer's disposal and (b) the task situation, which is influenced by the distance between the intended and produced texts at any point during the writing process (cf. van der Hoeven, 1999a; van Weijen et al., 2008). These studies indicate also that skilled writers not only use strategies quantitatively and qualitatively differently from less skilled writers, but also distribute their writing activities differently over the writing process (Roca de Larios et al., 2008; van der Hoeven, 1999a, 1999b; van Weijen et al., 2008). Van der Hoeven (1999a), for example, found that for less proficient writers, generating activities almost stayed the same or increased over time, suggesting that they adopted a knowledge-telling approach and were not able to adapt to the changing task situation. More proficient writers, in contrast, adapted their writing processes as the task situation changed. Specifically, they generated more frequently at the beginning of the writing process, but their generating activities decreased over time as other cognitive activities became more important. Generating early in the writing process may reflect the development of a conceptual representation of the task as a whole, which is an important ingredient in the production of high-quality texts. The decrease of generating activities and increase in other activities over time suggest that these writers use a knowledge-transforming strategy.

As discussed above, Roca de Larios et al. (2008) found that the amount of time devoted to different writing processes varied depending on writer L2 proficiency and that while low-proficiency writers maintained the same pattern of time allocation throughout the writing process, more proficient writers showed a more differentiated time allocation to different writing activities during the writing process. These studies show that it is important to examine not only what activities writers engage in, but also when they engage them during the writing process (Breetvelt et al., 1994; Roca de Larios et al., 2008; van der Hoeven, 1999a; van Weijen et al., 2008). Time during the writing process is a proxy variable for the changing task situation (Roca de Larios et al., 2008; van Weijen et al., 2008). Task situation refers to the writing context (Hayes & Flower, 1980). As the text evolves (e.g., through the addition of new content or the revision of already written text), the task situation changes. It is expected that writers will adapt their writing activities to those changes. This study will examine variability in test-taker writing activities across the writing process as well.

The Effects of Writing Task on Second-Language Writing Processes

While several studies have examined the effects of variation in task characteristics on L2 learners' writing processes (e.g., Clachar, 1999; Cumming, 1989; Krapels, 1990; Raimes, 1987), there is limited research on test takers' performance on independent and integrated writing tasks similar to those in the TOEFL iBT writing section. Independent writing tasks are tasks that require test takers to write about a topic based on their personal experience and/or general knowledge without referring to any other source, while integrated tasks require the test taker to read and/or listen to one or more texts (e.g., a reading passage, a lecture) and then create a written response (Cumming et al., 2000; Jamieson, Eignor, Grabe, & Kunnan, 2008). Integrated writing tasks entail the use of two or more language skills, while independent tasks are assumed to provide a measure of writing as an independent skill (Jamieson et al., 2008; Read, 1990). Examination of writing processes prompted by different tasks can help determine whether these tasks tap the same construct (Barkaoui, Brooks, Swain, & Lapkin, 2013; Y. Lee & Kantor, 2005).

A handful of studies have examined the effects of independent and integrated writing tasks on test scores (Y. Lee & Kantor, 2005), writing processes (Plakans, 2008), and text features (Cumming et al., 2005). Y. Lee and Kantor (2005), for instance, found high correlations among scores on independent, listening-based, and reading-based writing tasks, which suggest that these tasks may be measuring the same underlying construct. Cumming et al. (2005) found significant differences across integrated and independent writing tasks in terms of various linguistic and discourse features (e.g., text length, vocabulary, argument quality, grammatical complexity). While some studies have examined the writing processes of L2 learners when responding to integrated tasks (e.g., Cohen, 1994; Esmaeili, 2002; Plakans, 2009) or independent tasks (e.g., Clachar, 1999; Raimes, 1987), only one study has compared L2 writing processes with both task types (Plakans, 2008). Plakans (2008) found that the independent tasks led the students to engage in more initial and less online (i.e., during writing) planning, to reread their texts more frequently, to often orient themselves, to put considerably more effort into planning content before writing, and to make more negative evaluative comments. With the integrated tasks the students tended to reread the prompt more frequently, to engage in more thinking for task interpretation, and to engage in more online planning. Because the source texts in the integrated tasks provided students with both ideas and organization to apply in their writing, they did not need to spend as much time planning original content and organization at the beginning of the writing process. The higher frequency of online planning with the integrated tasks suggests that the students adopted a more recursive and less linear approach to meaning making during writing.

The Effects of Keyboarding Skills on Second-Language Writing Processes

Since the early 1980s, several studies have examined the impact of the computer on L1 and L2 learners' writing processes and text quality, mostly in nontest settings. Shaw (2005) identified three main patterns in the findings of this line of research (cf. Y. Lee, 2002; Pennington, 1996; Slattery & Kowalski, 1998). First, the findings are mixed, with some studies finding negative effects, others finding positive effects, and still others finding no effects of the computer on learners' writing processes or texts. Second, the computer seems to have different effects on L2 writers than on L1 writers. Finally, because most of this research has focused on the use of computers for teaching and learning purposes, the findings might have limited generalizability to assessment contexts. In particular, most of these studies aimed to improve the participants' writing performance and, as a result, allowed them more time and permission to use various writing and editing tools

(e.g., spelling and grammar checkers), which are likely to influence learners' writing processes and texts, but are not usually available to test takers in computer-based (CB) writing tests.

Most studies that have examined the effects of the computer on writing performance focused on comparing learners' writing processes, texts, and/or scores when writing on paper and on the computer (e.g., Breland, Lee, & Muraki, 2004; Burke & Cizek, 2006; Horkay, Bennett, Allen, Kaplan, & Yan, 2006; H. K. Lee, 2004; Russell & Haney, 1997; Wolfe & Manalo, 2005). Few studies have aimed to determine the extent to which the cognitive processing involved in responding to the two formats is similar (e.g., Baker & Kinzer, 1998; Haas, 1989; Y. Lee, 2002; Li, 2006; van Waes & Schellens, 2003; Weir, O'Sullivan, Jin, & Bax, 2007). Van Waes and Schellens (2003), for example, found that writing on the computer led to a more fragmented and recursive writing process than writing on paper, while Y. Lee (2002) found that some L2 writers employed different processes and focused on different aspects of writing across writing modes.

Previous studies on the effects of the computer on writing performance suffered from three main limitations. First, many of these studies did not consider writers' familiarity and experience with the computer, although there is evidence that test-taker computer familiarity and ability can moderate the effects of delivery mode on test performance. For instance, it seems that test takers with high levels of computer experience receive higher scores on word-processed essays, while test takers with lower levels of computer experience receive higher scores on handwritten essays (Wolfe & Manalo, 2005). Second, studies that considered writers' familiarity and experience with the computer relied on self-report measures of computer skills. However, perceived computer ability may be very different from actual ability (McCourt Larres, Ballantine, & Whittington, 2003). Several authors have emphasized that future studies need to specifically and directly assess test takers' keyboarding skills, that is, keyboarding speed and accuracy (Burke & Cizek, 2006; Connelly, Gee, & Walsh, 2007; Horkay et al., 2006). Finally, most of these studies focused on the effects of writing modes on test scores, but not test-taker writing processes. Findings from these studies are mixed, with some studies finding significant effects of computer experience on test scores (e.g., Horkay et al., 2006; Russell & Haney, 1997), others finding no significant effects (e.g., Maycock & Green, 2005), and still others finding that these effects vary depending on other factors such as task type (e.g., Burke & Cizek, 2006). Another possible explanation for the mixed results is that the rapid increase in the availability of computers and increased familiarity with technology over the past two decades may make early research findings less relevant to current students.

Theoretically, keyboarding skills can influence test takers' writing processes, texts, and scores. Cognitive models of writing (e.g., Fayol, 1999; Hayes & Flower, 1980; Kellogg, 1996; McCutchen, 1996, 2000; Scardamalia & Bereiter, 1987; Torrance & Galbraith, 2006) provide an explanation of how and why keyboarding skills can affect writing performance. According to these models, writing is a complex activity that requires the coordination of a variety of different cognitive processes that can compete for cognitive resources that are limited (Fayol, 1999; Hayes, 1996, 2006; Kellogg, 1996; McCutchen, 1996; Torrance & Galbraith, 2006). With increasing demand by some writing processes, performance based on other processes, which rely on the same cognitive resources, may suffer (Broekkamp & van den Bergh, 1996; Chenoweth & Hayes, 2001; Fayol, 1999; McCutchen, 1996, 2000; Olive & Kellogg, 2002). For example, writers with poor keyboarding skills may be forced to focus their attention and cognitive resources on motor activities (e.g., typing), and as a result, other processes and aspects of writing (e.g., planning, organization, reviewing) might be left unattended to, which can lead to poor text quality (Alves, Castro, de Sousa, & Stromqvist, 2007). From this perspective, if low-level skills such as keyboarding and spelling are automated,[2] they will not require any attentional resources and, consequently, will not constrain or influence the writing process and its outcomes (Fayol, 1999; Torrance & Galbraith, 2006). However, poor keyboarding skills may force writers to focus their attention and cognitive resources on motor activities (i.e., typing), and consequently, other higher order processes (e.g., planning, revising) might be left unattended to, which can lead to poorer text quality and lower scores (Alves et al., 2007). Additionally, there is evidence that when instructed to write using an unfamiliar method (e.g., typing, writing in capital letters), L1 writers tend to pause more frequently and to write more slowly, indicating a trade-off between the formulation and execution systems (Bourdin & Fayol, 1994; Olive & Kellogg, 2002). These effects might be magnified for L2 writers with low computer ability when writing on the computer under test conditions (Wolfe & Manalo, 2005).

A few studies have examined the relationships between keyboarding skills and writing processes. In a study of L1 writing, Wolfe, Bolton, Feltovich, and Niday (1996) found that, while the writing mode did not make a difference for students with mid to high levels of experience writing with computers, students with a low level of comfort and experience with computers scored almost one point lower ($M = 3.30$) and produced shorter essays and more simple sentences on the

CB version than on the paper-based (PB) version of a writing test ($M = 4.13$). Wolfe et al. explained that for students with low computer skills, writing on the computer seems to add "a physical and cognitive burden that interferes with [their] writing and cognitive processes" (p. 141). Alves et al. (2007) found that slow and fast typists employed different strategies when writing on the computer and that slow typists tended to produce shorter texts. Because they could not think and type at the same time, slow typists might be using a serial way of composing, whereby they devote pauses to high-level writing processes, such as planning and revising, and execution periods to typing.

In a study comparing the revision processes of four advanced L2 writers when writing on computer and on paper, Phinney and Khouri (1993) found that experience with the computer was a stronger factor than writing proficiency in determining students' writing strategies. The less experienced computer users spent less time revising, made more surface changes, and used the computer functions less frequently than did the experienced computer users. The experienced users showed a greater concern for the content than did the less experienced users, who indicated apprehension about using the computer and were concerned with correctness. The findings of these studies suggest that computer experience can affect writers' writing processes significantly when writing on the computer. From an information-processing perspective, lack of familiarity with writing on the computer can force writers to focus their attention and cognitive resources on motor activities (i.e., typing), which can inhibit attention to other processes and aspects of writing (e.g., planning, reviewing; Alves et al., 2007). To my knowledge, no previous studies have examined test-taker writing processes when responding to L2 writing tests on the computer, including TOEFL iBT, or the effects of L2 proficiency, keyboarding skills, and task type on those processes. This study aims to address this research gap.

Research Questions

This study was part of a larger project that aimed to examine the effects of ELP and keyboarding skills on test-taker performance on TOEFL iBT writing tasks. Barkaoui (2014) found that overall ELP and writing ability in English contributed substantially to variance in scores on both independent and integrated task scores, while keyboarding skills had a significant, but weak effect on scores on the independent task only. This study has two goals: (a) to describe the writing activities that a subsample of test takers studied by Barkaoui engaged in when responding to TOEFL iBT independent and integrated writing tasks on the computer and (b) to examine whether and how these activities are influenced by task type, test-taker ELP, and keyboarding skills. Specifically, the study addressed the following research questions:

1. What writing activities do test takers engage in when responding to TOEFL iBT integrated and independent writing tasks?
2. To what extent and how do the type and frequency of these activities vary across the writing process?
3. To what extent and how do these activities vary depending on test-taker ELP and keyboarding skills?
4. What are the relationships, if any, between test takers' writing activities and the quality of their texts?

Method

Participants

Ninety-seven students participated in the main research project (see Barkaoui, 2014). The students belonged to four groups: two ELP levels (high and low) by two keyboarding skill levels (high and low). All students were recruited at an English-medium university in Southern Ontario. The high ELP groups included postadmissions students in their first or second year of university (graduate or undergraduate) study. The low ELP groups included preadmission students who were enrolled in low- to high-intermediate (preacademic) English as a second language (ESL) classes. Keyboarding skill level was determined based on the results of two typing tests administered to participants at the beginning of the study. Both typing tests were similar in terms of length (one 200-word passage), time (2 minutes), and typing instructions and requirements. The passage was presented in the upper half of the computer screen, and participants then typed the text into a blank text box located at the lower half of the screen (www.assesstyping.com). Participants were instructed to type each text as quickly and as accurately as possible within 2 minutes (see Appendix A for a description of the typing tests). Participants with net typing speed (i.e., typing speed adjusted for typing accuracy) of 30 words per minute (WPM) or less were included in the low keyboarding skill groups, while students with net typing speed of 40 WPM or more were included in the high keyboarding skill groups (see Appendix A for more details).

Table 1 Descriptive Statistics for Test Scores and Net Typing Speed (WPM) by Group

ELP	Keyboarding skill	*n*		Typing speed	PB task score	CB integrated task score	CB independent task score
Low	Low	6	*M*	20.08	2.50	2.08	2.42
			SD	3.02	0.89	0.92	0.49
	High	5	*M*	44.10	2.80	1.90	2.70
			SD	3.52	0.27	1.14	0.45
High	Low	5	*M*	23.10	3.90	3.70	3.40
			SD	3.76	0.55	0.84	0.42
	High	6	*M*	61.00	3.83	4.58	4.08
			SD	5.66	0.75	0.49	0.92

Note. CB = computer based; ELP = English language proficiency; PB = paper based.

All 97 participants completed one PB writing task and two CB writing tasks (see below). In addition, a randomly selected sample of 22 students provided stimulated recalls for the CB writing tasks. This report focuses on the stimulated recall data only. Table 1 displays the number of participants in each group and descriptive statistics for their writing and typing skill test scores.

More than half the participants ($n = 14$) were males. Their ages ranged between 18 and 31 years ($M = 22.8$, $SD = 4$). They spoke 11 different first languages including Chinese ($n = 7$), Vietnamese ($n = 3$), Korean, Bengali, and Farsi ($n = 2$ each). The majority ($n = 19$) were in Canada for less than 1 year at the time of data collection; the remaining, all in the high ELP group, were in Canada between 1 and 2 years. The high ELP group included six graduate and five undergraduate students from various departments (e.g., law, economics, engineering, computer sciences, finance, marketing, management, and mathematics).

The majority reported that they were familiar with independent ($n = 18$) and integrated writing tasks ($n = 12$). Less than half ($n = 10$) reported that they had taken TOEFL at least once before, and a quarter ($n = 6$) reported that they had taken *TOEFL*® preparation classes before. Less than half the participants ($n = 10$) reported that they had taken a writing test on the computer before in the context of TOEFL ($n = 7$) and/or other CB tests ($n = 4$).

Writing Tasks

Three writing tasks were used in this study, two independent tasks and one integrated task. The independent writing tasks consisted of writing an essay about a general topic (30 minutes), while the integrated task consisted of listening to a lecture and reading a text about a topic (5 minutes) and then writing a summary of both the lecture and the reading (20 minutes). Test takers had access to the reading text but not the lecture during the writing segment of the integrated task. The three tasks, obtained from the TOEFL iBT Form Creator software, are representative of TOEFL iBT writing tasks. A PB version of one independent writing task was administered to the participants at the beginning of the study. The other independent and integrated tasks were administered to the participants on the computer. With both CB tasks, the participants had access to only three editing functions: cut, paste, and undo.

Stimulated Recalls

To collect data about the participants' writing activities while completing each writing task, each participant provided a stimulated recall of his or her writing sessions. This involved watching a playback of the writing session immediately after completing a writing task on the computer and describing what he or she was thinking before, while, and after completing the task. Stimulated recalls are based on the assumption that replaying the writing session will stimulate recall of mental processes that occurred during writing (Bosher, 1998). This method has been used in previous research (e.g., Bosher, 1998; Sasaki, 2000) to identify the aspects of writing that writers attend to, the problems they encounter, and the writing activities they engage in when composing.

Unlike think-aloud protocols, stimulated recalls ask participants to verbalize their internal thoughts after, rather than while, completing a writing task. L2 learners, in particular, may find it easier to think aloud after writing, particularly if they have to think aloud in L2 (Bosher, 1998; Sasaki, 2000). Additionally, because they take place after the writing task is completed, stimulated recalls do not interfere with the writing process and allow the researcher to ask questions about the

writing process (Bosher, 1998; Sasaki, 2000). However, unlike interviews that often provide generalized statements about writing processes, stimulated recalls allow the researcher to inspect specific occurrences of writing activities and strategies. Stimulated recalls have their limitations, however. In particular, they can elicit only what writers can recall or what writers think they were thinking about at the point of time in question. What writers recall however, may not be a faithful reproduction of what they were thinking about at that particular moment (Bosher, 1998; Sasaki, 2000). The following section describes the stimulated recall instructions and procedures used in this study to mitigate these limitations.

Data Collection Procedures

A pilot study was conducted to try the data collection tools and procedures. Based on this pilot study, some minor revisions were made to the data collection tools and procedures. At the beginning of the main study, recruitment e-mails and flyers were sent to international postadmission (undergraduate and graduate) and preadmission (ESL) students. Students who responded to the recruitment e-mails and flyers were instructed to complete the two online typing tests. Only students with net typing speed of 40 WPM or higher and those with net typing speed of 30 WPM or lower were invited to participate in the main study ($N = 97$). Twenty-two students were randomly selected to provide stimulated recalls.[3]

Each of the 22 students completed an informed consent form and then responded to the PB independent writing task (30 minutes) in a small group (of four to eight students) in a classroom. About 1 week later, I met with each student to train him or her on stimulated recalls. Each student was met individually in a quiet office to do the training, perform the writing tasks, and provide stimulated recalls. During the training session, the student first responded to the PB writing task again but on the computer. The screen-recording program Morae (http://www.techsmith.com/morae.html) was used to record all the student's writing activities on the computer as well as a video of the student while performing the task (using an external camera connected to the computer). Morae then combined the two recordings (on-screen and student's activities) into a picture-in-picture (PiP) video, which could be played back for the participant. Next, the student was provided with written and oral instructions and explanations on how to perform the stimulated recall task. The participant was instructed to watch the video and say aloud what he or she was thinking at the time of the writing and to talk freely about his or her thoughts and actions as his or her text appeared on the screen (see Appendix B for stimulated recall instructions). The student then watched the recording of the writing session on the computer and described what he or she was thinking while completing the writing task. The student was allowed to self-initiate replays, choose segments to comment on, and stop the replay if he or she needed time to talk about a specific writing event; I prompted talk when necessary. Following Lindgren (2005), only open prompts were used, such as "What are you doing now?" referring to a long pause or a revision in the text, or "Can you talk about that revision?" If the student could not recall the item at once, no further questions were asked (cf. Lindgren, 2005; see Appendix B). When it was felt that the student understood how to perform the stimulated recall task, the student took a short break.

Next, the student completed the CB integrated task (25 minutes), while Morae recorded both the on-screen writing activities and the student's activities (e.g., jotting down notes on scratch paper). Immediately after finishing the integrated task, the replay facility in Morae was used as a prompt for recall of the writing process. At the end of the first stimulated recall session, the student took a short break and then completed the CB independent task (30 minutes) while Morae recorded the writing session and the student's activities. Then, the student watched a playback of the second writing session (using Morae) and provided stimulated recall. Morae was used to record both (a) the playback of the video of the writing session itself and (b) what the student said while watching the video as a new PiP video file. After a short break, the student was asked about his or her thoughts about the stimulated recall process and if it affected his or her writing performance. Finally, each student completed a short online background questionnaire and was paid $80 for participating in the study. All the materials that students used or produced (e.g., notes, drafts) were collected.

The writing sessions varied in length between 16 and 30 minutes ($Mdn = 27$ minutes). The stimulated recalls varied in length between 20 and 51 minutes ($Mdn = 32$ minutes). The transcripts varied in length between 1,435 and 5,241 words ($Mdn = 2{,}612$ words). The correlation between length of writing session and stimulated recall length in minutes was $r = .65$. The correlation between length of stimulated recall in minutes and length of transcripts (i.e., number of words) was $r = .70$. These patterns indicate that longer writing sessions were associated with longer stimulated recall sessions and that longer stimulated recalls were associated with longer transcripts.

Each writing sample was rated by two independent, trained TOEFL iBT raters at Educational Testing Service (ETS) on the 5-point holistic rating scale for the TOEFL iBT writing section. Interrater reliability (Cronbach's alpha) was .88 for the

PB independent essays, .94 for the CB integrated essays, and .87 for the CB independent essays. The final score for each writing sample is the average of the scores from the two raters.

Data Coding and Analysis

Coding of Stimulated Recalls

The 44 stimulated recalls (22 test takers by two tasks) were transcribed, segmented, and then coded, after establishing intercoder agreement, in terms of various writing activities as described in Figure 1 and Appendix C. First, each writing session for each participant was divided into three equal segments. For example, if a writing session was 27 minutes long, it was divided into three 9-minute segments. Next, each stimulated recall was divided into three parts, that is, one part corresponding to each of the three segments of the writing session. Each stimulated recall was then segmented into idea units, with each unit being assigned one code according to the predominant writing activity reported (cf. Mateos, Martin, Villalon, & Luna, 2008; Sasaki, 2000). Some units, however, were assigned more than one code. The coding scheme (see Figure 1 and Appendix C) was built on those developed by Sasaki (2000); Mateos et al. (2008), and Plakans (2008, 2009)

A. Interacting With Task (84%)
- Reading test instructions
- Reading the writing task
- Reflecting on the writing task

B. Interacting With Sources (for integrated task only) (76%)
- Referring to sources
- Reflecting on sources
- Integrating sources
- Checking comprehension
- Detecting comprehension difficulty
- Using comprehension strategy
- Reacting to sources
- Mining

C. Planning and Organizing (90%)
- Global planning
- Local planning
- Introduction planning
- Conclusion planning

D. Generating and Retrieving (88%)
- Plan retrieving
- Self-based generating
- Text-based generating
- Source-based generating

E. Detecting Writing Difficulty (71%)
- Content
- Language
- Rhetoric

F. Using Writing Strategy (72%)
- Using writing strategy

G. Evaluating (85%)
- Local text
- Global text
- Content
- Language
- Rhetoric
- Text length
- Reacting to own writing

H. Revising (93%)
- Content
- Language
- Rhetoric

I. Procedural (81%)
- Describing actions
- Verbalizing a proposition
- Checking the time

Figure 1 Coding scheme for writing activities (percentage of intercoder agreement). Based on Mateos et al., 2008; Plakans, 2008, 2009; Sasaki, 2000.

as well as preliminary analyses of data from this study. I first compiled a list of writing activities reported in the literature. Next, based on preliminary examination of the current data, I added activities and dropped others that were not used by participants in this study. The final coding scheme consisted of 36 writing activities under nine main categories as listed in Figure 1 (see Appendix C for definitions and examples of codes).

A research assistant (RA) was trained before coding all the stimulated recalls. To achieve acceptable levels of intercoder agreement, several rounds of discussion, training, coding, and checking were conducted. First, the coding scheme was discussed and piloted on one stimulated recall by the RA and the author. This led to some minor modifications in terms of the number and descriptions of the codes. This process was performed twice before finalizing the coding scheme. Next, the RA independently coded all the stimulated recalls using NVivo. NVivo allowed the viewing and coding of both the transcript and the video recording of the stimulated recall simultaneously. The stimulated recall video provided a rich context for interpreting and coding the stimulated recalls, since the coder could read and hear what the student was saying and watch what the student was seeing during the stimulated recall session (i.e., playback of the student's writing session in PiP video). Finally, a randomly selected sample of six stimulated recalls (i.e., 14%) was coded by the present author to estimate intercoder agreement. The overall intercoder agreement was 82%, but the percentage of agreement varied between 71% (for detecting writing difficulty) and 93% (for revising). Figure 1 displays the percentage of intercoder agreement for each of the nine main categories of writing activities.

Statistical Analyses

The focus in this study is on describing and comparing the frequency and distribution of reported writing activities. Consequently, the coded data were tallied and percentages of reported writing activities were computed for each test taker for each task as follows: counts of coded writing activities (e.g., revising) were summed for each test taker for each writing task and then divided by the total number of instances of the writing activities reported by that particular test taker for that particular task to obtain a percentage of times that that code (i.e., revising) occurred. These percentages served as the data for comparison across groups and tasks. Percentages, rather than absolute frequencies, were used because of the variability in the number and type of writing activities reported across participants and tasks. This variability makes comparisons of reported writing activities across groups and tasks problematic. For example, if two participants report using a particular activity the same number of times (e.g., 10), but one reports twice as many activities overall than the other (e.g., 20 vs. 40), reporting that both participants mentioned the activity the same number of times is misleading, as this does not take into account the relative frequency of reported activity. The use of percentages, instead of absolute frequencies, solves this problem. In the example above, the percentages will be 50% and 25%, which better reflect the differences between the two participants in terms of reported activities. This applies to task and group comparisons as well.[4]

Statistical tests were then conducted on the main categories (shown in bold in Figure 1). Subcategories were used for descriptive purposes only and to explain significant differences in main categories. Because the writing activities reported differed across writing tasks, analyses were conducted for each task separately, except when tasks were compared. Because the distributions of the percentages of the reported activities were significantly different from normal and because of the presence of many zeroes in the data, nonparametric tests were used to address the research questions of the study. To compare test-taker groups, Kolmogorov-Smirnov two-sample tests[5] were conducted, with test-taker group (e.g., ELP group) as the independent variable and percentages of reported writing activities as the dependent variables. To compare tasks, Wilcoxon signed-rank tests[6] were computed with task as the independent variable and percentages of reported writing activities as the dependent variables. To compare writing activities across writing phases, Friedman tests[7] were conducted, with writing phase as the independent variable and percentages of reported strategies as the dependent variables. Where a significant difference was detected, the Friedman test was followed by pairwise comparisons across writing phases using Wilcoxon signed-rank tests. A Bonferroni correction was applied to pairwise comparisons across phases, so all effects could be reported at appropriate level of significance (i.e., $.05/3 = .017$). To address Research Question 4, correlational analyses using the Spearman rho (r_s) coefficient were computed between the percentages of reported writing activities and task scores.

Because nonparametric statistical tests rely on ranks, rather than the value of scores and percentages, the following descriptive statistics are reported below: mean (*M*), median (*Mdn*) and the minimum (or lowest, *Min*) and maximum (or highest, *Max*) values for each main category. Additionally, following Field (2009), *r* is used as a measure of effect size.[8] This coefficient is constrained to lie between 0 (*no effect*) and 1 (*maximum effect*). Following Cohen (1988); Field (2009,

p. 57) suggested the following guidelines for interpreting effect sizes: small effect, $r = .10$; medium effect, $r = .30$; and large effect, $r = .50$. As Field (2009, p. 57) clarified, a small effect ($r = .10$) explains 1% of the total variance, a medium effect ($r = .30$) explains 9%, and a large effect ($r = .50$) explains 25% of the total variance.

Results

Writing Activities by Task Type

The total number of writing activities reported in the 44 stimulated recalls was 4,763 activities (*Mdn* = 111, *Min* = 70, *Max* = 168). Of these, 2,302 were reported with the integrated task (*Mdn* = 104) and 2,461 with the independent task (*Mdn* = 117). Tables 2 and 3 display descriptive statistics for writing activities for the integrated and independent tasks, respectively. The column labeled "*n*" indicates the number of participants (out of 22 participants) who mentioned the activity at least once. The column labeled "Total raw frequency" lists the number of times each writing activity was reported across all participants. The columns labeled "*Min*" and "*Max*" under "Raw frequency" indicate the minimum and maximum numbers of activities reported by any participant. The columns labeled "% out of total number of activities" indicate the mean, median, minimum, and maximum for the percentage of each activity in relation to the total number of activities reported.

Table 2 shows that for the integrated task the most frequently reported category of activities is interacting with sources (*Mdn* = 19.09%), followed by evaluating (18.16%), procedural (12.92%), detecting writing difficulty (10.98%), generating and retrieving (9.07%), planning and organizing (8.48%), revising (7.53%), using writing strategy (5.11%), and interacting with task (4.51%). In terms of subcategories of writing activities, the 10 activities that were reported most frequently with the integrated task were revising language (*Mdn* = 6.28%), detecting writing difficulty with language (6.25%), checking time (5.54%), using writing strategy (5.11%), local planning (5.03%), referring to sources (5.00%), evaluating local text (3.88%), source-based generating (3.83%), evaluating language (3.70%), and verbalizing proposition (2.83%). All of these activities were mentioned at least once by 20 participants or more, except for verbalizing proposition, which was mentioned by 14 participants only. Figure 2 lists some excerpts that illustrate each of these subcategories of writing activities. The following activities were reported least frequently with the integrated task (*Mdn* = 0.00%): text-based generating, revising rhetoric, introduction planning, conclusion planning, and detecting writing difficulty with rhetoric. All of these activities were reported by fewer than half the participants. For example, text-based generating was reported by only three participants, while revising rhetoric was reported by only five participants.

For the independent task, Table 3 shows that the most frequently reported category of activities is evaluating (*Mdn* = 24.19%), followed by planning and organizing (14.89%), detecting writing difficulty (14.59%), procedural (14.19%), generating and retrieving (11.92%), revising (11.44%), using writing strategy (5.76%), and interacting with task (2.70%). In terms of subcategories of writing activities, the 10 activities that were reported most frequently with the independent task were revising language (*Mdn* = 9.45%), self-based generating (9.09%), local planning (8.93%), evaluating local text (8.40%), verbalizing proposition (8.31%), detecting writing difficulty with language (7.22%), evaluating language (6.15%), using writing strategy (5.76%), detecting writing difficulty with content (3.93%), and global planning (3.61%). Each of these activities was mentioned at least once by at least 21 participants (out of 22 participants in the study). Figure 3 lists some excerpts that illustrate each of these subcategories of writing activities. The following activities were reported least frequently with the independent task (*Mdn* = 0.00%): reading test instructions (mentioned by only five participants) and revising rhetoric (mentioned by only 10 participants).

Overall, with both tasks, the participants reported more evaluating than revising and more detecting writing difficulty than using writing strategy (to solve difficulties). In terms of subcategories of writing activities, of the 10 most frequently reported activities, seven were the same for both tasks. These were revising language, detecting writing difficulty with language, using writing strategy, local planning, evaluating local text, evaluating language, and verbalizing proposition. Referring to source, source-based generating, and checking time were among the 10 most frequently reported activities with the integrated task, while self-based generating, detecting writing difficulty with content, and global planning were among the 10 most frequently reported activities with the independent task.

Comparisons of the proportions of main categories of writing activities across writing tasks (i.e., Wilcoxon signed-ranks tests) indicated that the integrated task elicited significantly more activities related to interacting with task

Table 2 Descriptive Statistics for Writing Activities for the Integrated Writing Task

		Raw frequency			% out of total numbers of activities			
Writing activity	*n*	Total	*Min*	*Max*	*M*	*Mdn*	*Min*	*Max*
Interacting with task	22	111	2	10	4.97	4.51	1.55	8.11
Reading test instructions	19	30	0	4	1.37	1.21	0.00	3.77
Reading the writing task	16	33	0	5	1.55	1.25	0.00	5.43
Reflecting on writing task	20	48	0	6	2.05	1.78	0.00	5.41
Interacting with sources	22	496	8	43	21.59	19.09	7.55	44.33
Referring to sources	22	130	2	13	5.76	5.00	1.89	14.94
Reflecting on sources	15	42	0	7	2.05	1.78	0.00	5.41
Integrating sources	18	43	0	7	1.88	1.24	0.00	6.48
Checking comprehension	17	35	0	5	1.63	1.27	0.00	5.06
Comprehension difficulty	19	82	0	10	3.65	2.83	0.00	11.39
Comprehension strategy	20	56	0	9	2.44	1.68	0.00	9.28
Reacting to sources	14	29	0	6	1.28	1.18	0.00	6.19
Mining	22	79	1	11	3.24	2.68	0.78	8.27
Planning and organizing	22	188	2	17	8.67	8.48	1.84	16.88
Global planning	20	51	0	6	2.25	2.35	0.00	5.19
Local planning	22	106	1	9	4.93	5.03	0.78	8.86
Introduction planning	9	14	0	4	0.64	0.00	0.00	3.60
Conclusion planning	10	17	0	3	0.84	0.00	0.00	3.80
Generating and retrieving	21	206	0	26	9.16	9.07	0.00	19.55
Plan retrieving	16	59	0	8	2.55	2.27	0.00	6.90
Self-based generating	13	29	0	8	1.29	1.02	0.00	8.25
Text-based generating	3	4	0	2	0.19	0.00	0.00	2.53
Source-based generating	20	114	0	14	5.13	3.83	0.00	13.51
Detecting writing difficulty	22	255	2	23	11.15	10.98	2.06	25.29
Content	14	74	0	12	3.46	2.83	0.00	13.79
Language	22	165	2	22	6.98	6.25	2.06	16.30
Rhetoric	10	16	0	3	0.71	0.00	0.00	3.45
Using writing strategy	21	121	0	12	5.12	5.11	0.00	8.96
Evaluating	22	441	7	45	18.77	18.16	6.02	37.74
Local text	21	118	0	14	5.02	3.88	0.00	13.21
Global text	16	30	0	4	1.28	1.18	0.00	3.70
Content	17	45	0	6	2.02	1.55	0.00	7.41
Language	22	105	1	16	4.41	3.70	1.15	15.09
Rhetoric	16	29	0	6	1.29	1.05	0.00	4.65
Text length	16	66	0	19	2.63	1.09	0.00	14.73
Reacting to own writing	13	48	0	12	2.13	1.33	0.00	11.32
Revising	21	206	0	25	8.63	7.53	0.00	19.38
Content	15	36	0	7	1.53	1.18	0.00	5.43
Language	21	162	0	19	6.73	6.28	0.00	14.94
Rhetoric	5	8	0	2	0.37	0.00	0.00	2.44
Procedural	22	278	1	21	11.95	12.92	1.15	18.07
Describing actions	20	71	0	9	3.35	2.67	0.00	9.78
Verbalizing a proposition	14	71	0	10	2.81	2.83	0.00	8.49
Checking time	21	136	0	17	5.79	5.54	0.00	13.18
Grand total		2,302	74	163				

($Z = -2.16$, $p < .05$, $r = .46$) than did the independent task (see Table D1 in Appendix D). The independent task prompted significantly more activities related to planning and organizing ($Z = 3.52$, $p < .05$, $r = .75$), detecting writing difficulty ($Z = 2.19$, $p < .05$, $r = .47$), evaluating ($Z = 2.94$, $p < .05$, $r = .63$), and revising ($Z = 2.26$, $p < .05$, $r = .48$) than did the integrated task (see Tables 2 and 3). There were no significant differences across tasks in relation to generating and retrieving ($Z = 1.61$, $p > .05$), using writing strategy ($Z = .41$, $p > .05$), and procedural ($Z = 1.09$, $p > .05$), although the independent task resulted in a higher proportion of generating and retrieving and procedural activities (see Tables 2 and 3). In terms of subcategories of writing activities, the participants reported reading test instructions, describing actions, and checking the time more frequently with the integrated task. They reported global planning, local planning, self-based generating, difficulty with content, difficulty with language, difficulty with rhetoric, evaluating local

Table 3 Descriptive Statistics for Writing Activities for the Independent Writing Task

		Raw frequency			% out of total numbers of activities			
Writing activity	*n*	Total	*Min*	*Max*	*M*	*Mdn*	*Min*	*Max*
Interacting with task	21	74	0	7	3.18	2.70	0.00	7.95
Reading test instructions	5	6	0	2	0.25	0.00	0.00	2.27
Reading the writing task	17	26	0	4	1.14	0.95	0.00	5.48
Reflecting on writing task	18	42	0	6	1.79	1.29	0.00	4.80
Planning and organizing	22	358	6	28	14.92	14.89	5.36	24.56
Global planning	21	94	0	10	3.99	3.61	0.00	10.00
Local planning	22	216	2	17	8.94	8.93	2.27	18.28
Introduction planning	15	22	0	5	0.89	0.90	0.00	4.39
Conclusion planning	15	26	0	3	1.10	1.09	0.00	2.74
Generating and retrieving	22	259	5	19	11.42	11.92	2.98	21.43
Plan retrieving	12	28	0	5	1.15	0.86	0.00	4.27
Self-based generating	22	211	1	18	8.96	9.09	0.85	17.14
Text-based generating	15	20	0	3	0.93	0.88	0.00	4.29
Detecting writing difficulty	22	368	3	38	14.52	14.59	4.11	23.03
Content	22	117	1	14	4.85	3.93	0.70	10.53
Language	22	208	1	31	8.00	7.22	1.37	18.79
Rhetoric	17	43	0	7	1.66	1.19	0.00	5.60
Using writing strategy	22	146	2	16	5.59	5.76	2.41	9.70
Evaluating	22	601	10	62	24.29	24.19	10.92	37.29
Local text	21	200	0	20	7.97	8.40	0.00	13.89
Global text	12	31	0	7	1.31	1.09	0.00	6.14
Content	18	77	0	8	3.19	2.49	0.00	8.60
Language	21	160	0	18	6.35	6.15	0.00	13.89
Rhetoric	18	46	0	10	1.87	1.10	0.00	8.47
Text length	17	48	0	7	1.93	1.51	0.00	5.56
Reacting to own writing	12	39	0	10	1.66	0.75	0.00	10.84
Revising	22	298	3	30	11.84	11.44	2.63	22.03
Content	17	50	0	9	1.92	1.17	0.00	7.89
Language	22	231	1	23	9.28	9.45	0.88	17.81
Rhetoric	10	17	0	4	0.64	0.00	0.00	3.39
Procedural	22	347	2	32	14.23	14.19	2.54	25.60
Describing actions	11	19	0	3	0.82	0.30	0.00	3.61
Verbalizing a proposition	21	225	0	27	9.02	8.31	0.00	22.69
Checking time	20	103	0	13	4.39	3.43	0.00	13.64
Grand total		2, 461	70	168				

text, evaluating language, evaluating content, revising language, and verbalizing proposition more frequently with the independent task (see Tables 2 and 3).

Variability in Writing Activities Across the Writing Process

As noted above, each writing session for each participant and each task was divided into three equal segments, and the number and percentage of writing activities reported in relation to each segment were computed and compared across segments. Tables 4 and 5 report descriptive statistics for the main categories of writing activities across the three writing phases for the integrated and independent writing tasks, respectively. To examine whether the differences in the median percentages across the three writing phases for each main category of writing activities are statistically significant, Friedman tests were conducted for each main category for each task separately, with writing phase as the independent variable and percentage of main category of reported writing activities as the dependent variables. When Friedman tests detected a significant result, Wilcoxon signed-ranks tests were conducted to compare each pair of phases (see Tables D2 and D3 in Appendix D). A Bonferroni correction was applied to pairwise comparisons across phases, so all effects are reported at appropriate level of significance (i.e., $p < .017$).

For the integrated task, Friedman tests were significant for all main categories ($p < .05$) except procedural, while for the independent task, the test was significant for all main categories ($p < .05$) except generating and retrieving, detecting

Writing activity	Mdn %	Frequency (n participants)	Example (participant code, task)
Revising language	6.28	162 (21)	Still I'm, I'm, I know that the sentence is wrong but I can't think of another way of writing so I was like adding like words and trying to rearrange the sentence. (GL, Integrated)
Detecting writing difficulty with language	6.25	165 (22)	Uh, I'm just trying to look at I'm trying to look at the, uh, the reading part the first the this word (I think) I'm try—I'm not sure the spelling so I'm trying to look at the word so I stopped a lit—a little bit and then I start my first theory…(DZ, Integrated)
Checking the time	5.54	136 (21)	I just like, uh, just wrote down without thinking of anything I thought I was not be able to finish at that time so I was like in a hurry. (GL, Integrated)
Using writing strategy	5.11	121 (21)	I…yeah, I thought for a little while to explain further, but I think I was in doubt, uh, to know where, and I finally decided that it's enough to put it there THEY HAVE A BUILT-IN COMPASS everything else should be clear. (DF, Integrated)
Local planning	5.03	106 (22)	I was thinking what I should write on the second paragraph what I should write what I should write about second paragraph, so, I, I wrote something short. (EC, Integrated)
Referring to source	5.00	130 (22)	I wanted to, you know, make familiar to understand what the third paragraph are saying. So I just reread it again, I read the paragraph, and I noticed about the time what how time is, how long time is left that I have enough time to go for details or not, And I didn't get it again I didn't get the whole idea what it's about, I think that maybe because I remembered I heard something about crystals in the listening, and I hear there was something about magnetic crystal x as internal compasses I thought maybe that is what the paragraph is about, so I rewrite it here. (MA, Integrated)
Evaluating local text	3.88	118 (21)	Oh just, uh, oh the starting point, uh, is not good or not the starting point is more clearer or not is it not clear but I choose it, mmm. (SM, Integrated)
Source-based generating	3.83	114 (20)	I made some examples from the reading part and I, I read the, the refute part into the same examples which shows the, uh, the differences, I want to make the statement clear whether uh, uh, how it disagrees with the, the theory. (DZ, Integrated)
Evaluating language	3.70	105 (22)	I think I my sentence is not, it's not in the correct way I'm thinking making that into proper sentence. (GL, Integrated)
Verbalizing proposition	2.83	71 (14)	I think about because in the lecture I listen that, uh, the last one, it's the most complete theory about the mmm, about [TEXT] so, so I try to make the last one is the best one, this one is uh (R: Do you mean the first?) the first one is the more complete theory because (xx) so I try to argue in my essay THE LAST ONE IS THE BEST, THESE TWO IS FLAW, BECAUSE [TEXT]. Yeah. (DL, Integrated)

Figure 2 The 10 most frequent subcategories of writing activities reported with the integrated task. The passages are marked with the following transcription conventions: ( ) = uncertain transcription; x = incomprehensible word; comma = short pause; capital letters = words that the student has written, was thinking of writing, and/or read directly from his/her text; italics = text read from task; underlined = text read directly from the reading text or heard from the lecture; [ ] = procedural and other behaviors; ? = questioning intonation; [TEXT] = text read from task, reading text, and/or lecture (not included here for test security reasons).

Writing activity	Mdn %	Frequency (n participants)	Example (participant code, task)
Revising language	9.45	231 (22)	Yeah, I want to write that PEOPLE DIFFER and people are different other than telling people are different people differ and I show differ like differ, and there was and xxx and there and I deleted one (GL, Independent)
Self-based generating	9.09	211 (22)	How to, how to use [laughs] I'm trying to find some memories in my brain how to support the examples how to support my second points and now I'm thinking about the second one (FF, Independent)
Local planning	8.93	216 (22)	And uh, it's now I'm thinking about the specific, specific, uh, example, uh, here, uh, specific example reason about the x university so the this to (reason) the I could I could make a, mmm, in a short short, in a short time and, mmm, I'm trying to make three, trying to make three, uh, examples, but I had only two examples in this time, so, so I'm thinking about one more (HT, Independent)
Evaluating local text	8.40	200 (21)	I were thinking that GLAD is not a formal word, and VARIOUS is not good (WG, Independent)
Verbalizing proposition	8.31	225 (21)	at the end I I want to make a conclusion I want to uh (xx) is I want to write uh I I DISAGREE WITH THE STATEMENT THAT IN THE IN THE QUESTION HERE (DL, Independent)
Detecting writing difficulty with language	7.22	208 (22)	Yeah, there I was, like, looking for a word, maybe like SKILLS, I wrote the SKILL THERE again I'm thinking I should not be writing that word again but I couldn't find some other word there, and I think I wrote THIS SKILLS again, it's like I'm repeating the same thing but I couldn't find (GL, Independent)
Evaluating language	6.15	160 (21)	When I'm writing this, mmm, I, it's not, I was deleted it because I'm not sure whether it is academic or not because I'm using EVEN CHILDREN UNDERSTAND SOMETHING so I was thinking whether it's a good way to write in the topic sentence or not (DZ, Independent)
Using writing strategy	5.76	146 (22)	And maybe if I'm start it I can I think if at this moment if I start it maybe I can have, I can write better, and I will find a good point (FF, Independent)
Detecting writing difficulty with content	3.93	117 (22)	I was thinking how I should link [TEXT], uh, so actually I was not entirely sure what I what I'm writing, this is why I stopped several times, again I…I'm reading on that time because I want to figure out does it make sense or not, so uh, I was taking my time because I knew I had a time (EC, Independent)
Global planning	3.61	94 (21)	I was planning my, uh, my essay before I want to x first that's why it took me a long time before I pressed NEXT, then I read it and… I start sketching the ideas I, I was reading it, and I was thinking OK uh, CAREER now I have to make up, make up my mind like [TEXT], eventually I chose CAREER, it just made more sense to me so I chose that, and I sketched out three ideas that is only two because I already know the third one, and so OK (TT, Independent)

Figure 3 The 10 most frequent subcategories of writing activities reported with the independent task. The passages are marked with the following transcription conventions: ( ) = uncertain transcription; x = incomprehensible word; comma = short pause; capital letters = words that the student has written, was thinking of writing, and/or read directly from his/her text; italics = text read from task; underlined = text read directly from the reading text or heard from the lecture; [ ] = procedural and other behaviors; ? = questioning intonation; [TEXT] = text read from task, reading text, and/or lecture (not included here for test security reasons).

Table 4 Descriptive Statistics for Main Categories of Writing Activities Across Writing Phases for the Integrated Task

	Raw frequency	Percentage out of total for each category			
Writing activity	Total	*M*	*Mdn*	*Min*	*Max*
Interacting with task	174				
Phase 1	94	58.73	55.05	22.22	100.00
Phase 2	35	17.83	18.33	0.00	40.00
Phase 3	45	23.43	20.00	0.00	66.67
Interacting with sources	559				
Phase 1	281	53.96	55.84	20.93	84.62
Phase 2	162	27.59	28.80	4.17	44.19
Phase 3	116	18.43	19.34	0.00	41.67
Planning and organizing	209				
Phase 1	38	22.68	21.11	0.00	66.67
Phase 2	89	39.82	37.98	0.00	83.33
Phase 3	82	37.48	35.82	0.00	75.00
Generating and retrieving	239				
Phase 1	11	4.25	0.00	0.00	30.00
Phase 2	141	55.34	55.55	0.00	88.89
Phase 3	87	40.39	42.85	0.00	100.00
Detecting writing difficulty	284				
Phase 1	28	8.18	7.14	0.00	28.57
Phase 2	123	43.41	41.66	0.00	83.33
Phase 3	133	48.40	50.00	0.00	100.00
Using writing strategy	136				
Phase 1	7	3.45	0.00	0.00	16.67
Phase 2	63	52.54	50.00	14.29	100.00
Phase 3	66	44.00	50.00	0.00	85.71
Evaluating	475				
Phase 1	26	4.52	0.00	0.00	39.29
Phase 2	171	35.62	36.93	16.67	62.50
Phase 3	278	59.84	62.50	21.43	83.33
Revising	223				
Phase 1	7	2.28	0.00	0.00	20.00
Phase 2	67	31.03	30.00	0.00	66.67
Phase 3	149	66.67	70.00	33.33	100.00
Procedural	303				
Phase 1	93	36.15	36.25	7.14	100.00
Phase 2	85	26.00	23.11	0.00	50.00
Phase 3	125	37.84	40.00	0.00	75.00

writing difficulty, and using writing strategy (see Tables D2 and D3 in Appendix D for results of statistical tests). As Tables 4 and 5 show, the participants reported significantly more interaction with task in the first phase of the process than in later phases with both tasks and significantly more interactions with sources in Phase 1 than in Phases 2 and 3 with the integrated task. Planning and organizing was reported more frequently in the last two phases of the writing process than in Phase 1 with the integrated task and significantly more frequently in the first phase than in later phases with the independent task. Specifically, global planning was reported most frequently in Phase 1 with the independent task and in Phase 2 with the integrated task, while local planning was reported more frequently in the last two phases with both tasks.

The proportions of generating and retrieving activities varied significantly across phases for the integrated task but not for the independent task (see Tables D2 and D3 in Appendix D). As Table 5 shows, generating and retrieving activities are distributed almost evenly across the three phases for the independent task, although they tended to decrease in Phase 3. With the integrated task, participants reported significantly more generating and retrieving activities in the last two phases than in Phase 1 (see Table 4). The participants also reported detecting writing difficulties and using writing strategy more frequently in the last two phases than they did in Phase 1 with both tasks, but these differences were significant only for the integrated task (see Table 4 and Tables D2 and D3 in Appendix D). The participants also reported significantly more evaluating and revising activities in the last two phases of the writing process, and particularly in Phase 3, than they did in Phase 1 with both tasks. Finally, the participants reported significantly more procedural activities at the end of the

Table 5 Descriptive Statistics for Main Categories of Writing Activities Across Writing Phases for the Independent Task

	Raw frequency	Percentage out of total for each category			
Writing activity	Total	*M*	*Mdn*	*Min*	*Max*
Interacting with task	112				
Phase 1	76	81.71	91.66	27.27	100.00
Phase 2	14	6.33	0.00	0.00	36.36
Phase 3	22	11.95	0.00	0.00	50.00
Planning and organizing	388				
Phase 1	158	42.14	39.35	26.09	70.00
Phase 2	121	30.93	31.25	10.00	53.33
Phase 3	109	26.92	25.94	0.00	50.00
Generating and retrieving	300				
Phase 1	101	33.53	33.33	0.00	57.14
Phase 2	118	39.21	36.06	20.00	73.33
Phase 3	81	27.25	27.55	0.00	53.85
Detecting writing difficulty	400				
Phase 1	107	24.55	21.53	0.00	62.50
Phase 2	157	39.29	41.87	13.64	54.17
Phase 3	136	36.14	34.31	0.00	66.67
Using writing strategy	162				
Phase 1	46	28.63	25.00	0.00	100.00
Phase 2	65	39.37	33.33	0.00	100.00
Phase 3	51	31.99	28.57	0.00	100.00
Evaluating	645				
Phase 1	126	22.28	20.52	0.00	57.14
Phase 2	213	33.36	34.33	15.38	54.17
Phase 3	306	44.35	47.85	0.00	64.58
Revising	315				
Phase 1	54	19.47	18.01	0.00	50.00
Phase 2	90	32.67	32.18	0.00	66.67
Phase 3	171	47.84	41.15	0.00	88.89
Procedural	382				
Phase 1	108	25.11	25.00	0.00	50.00
Phase 2	123	30.97	31.34	0.00	58.33
Phase 3	151	43.91	40.83	16.67	100.00

writing session with the independent task than they did in other phases. In particular, they reported checking the time more frequently in Phase 3.

The patterns in Table 5 suggest that, with the independent task, the participants tended to read and reflect on the writing task at the beginning of the writing process (i.e., Phase 1), to plan, retrieve, and generate ideas mainly in Phases 1 and 2. They experienced writing difficulties and used writing strategies to address them mainly in Phases 2 and 3 of the writing process. At the end of the writing process, they tended to evaluate and revise their texts and to check the time more frequently. As Table 4 shows, with the integrated task, the participants tended to read and reflect on the writing task and to interact with the sources at the beginning of the writing process and then to plan, generate, and write down their ideas mainly in the last two phases of the writing process. Most of the writing difficulties the participants experienced and the writing strategies they used to address them occurred in the last two phases of the writing process too. At the end of the writing process, the participants tended to evaluate and revise their texts. Overall, all groups, regardless of ELP and keyboarding skill level, seemed to have adopted the same patterns.

Variability in Writing Activities Across English Proficiency Groups

Table 6 displays descriptive statistics for the main categories of writing activities across tasks and ELP groups (see Appendix E for results for the subcategories). Because nonparametric statistical tests do not allow examining the effects of more than one independent variable at a time or the examination of interaction effects, two statistical tests were conducted to compare writing activities across tasks and ELP groups. First, Wilcoxon signed-ranks tests were conducted to compare the proportions of the main categories of writing activities relative to each other between tasks for each ELP

Table 6 Descriptive Statistics for Percentage of Main Categories of Writing Activities by Task and English Language Proficiency Group

Task	Integrated								Independent							
ELP group	Low				High				Low				High			
Writing activity	*M*	*Mdn*	*Min*	*Max*	*M*	*Mdn*	*Min*	*Max*	*M*	*Mdn*	*Min*	*Max*	*M*	*Mdn*	*Min*	*Max*
Interacting with task	5.00	4.29	1.85	8.11	4.93	4.71	1.55	8.05	3.74	4.27	0.00	7.95	2.62	2.42	0.72	5.26
Interacting with sources	19.69	18.29	7.55	32.41	23.47	19.37	14.94	44.33	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Planning and organizing	9.65	10.18	1.84	15.32	7.68	8.04	1.89	16.88	16.82	18.42	6.38	24.56	13.02	13.72	5.36	20.00
Generating and retrieving	9.82	9.25	2.47	18.92	8.49	8.88	0.00	19.55	11.44	12.28	6.84	14.46	11.40	10.52	2.98	21.43
Writing difficulty	10.85	11.11	6.10	17.28	11.43	10.85	2.06	25.29	13.35	15.05	4.11	19.30	15.67	14.28	9.15	23.03
Using writing strategy	4.66	5.00	0.00	8.64	5.57	5.74	0.78	8.96	5.03	5.12	2.41	7.89	6.13	5.93	2.86	9.70
Evaluating	18.55	15.31	8.70	37.74	18.98	19.25	6.02	30.23	24.90	25.80	11.40	37.29	23.67	21.56	10.92	36.90
Revising	9.34	9.00	5.56	14.63	7.91	6.01	0.00	19.38	10.40	8.43	2.63	22.03	13.28	12.85	3.33	21.13
Procedural	12.40	12.88	4.94	18.07	11.49	13.43	1.15	15.50	14.28	16.86	2.54	25.60	14.17	13.76	3.51	25.49

Note. ELP = English language proficiency. Interacting with sources is not indicated for the independent task because there is no source for this type of task.

group separately (i.e., comparisons of writing activities across tasks for each group; see Table D4 in Appendix D). Second, two-sample Kolmogorov-Smirnov tests were conducted to compare the proportions of the main categories of writing activities across ELP groups for each task separately (i.e., comparisons of writing activities between groups for each task; see Table D5 in Appendix D).

Comparisons across tasks within each ELP group (Table D4) indicated that both ELP groups reported significantly fewer planning and organizing activities with the integrated task than they did with the independent task. Additionally, the low ELP group reported significantly more writing difficulties with the independent task ($Mdn = 15.00$) than it did with the integrated task ($Mdn = 11.00$). The high ELP group reported significantly fewer interactions with the task ($Mdn = 4.71$) and significantly more evaluation ($Mdn = 21.56$) and revision ($Mdn = 12.85$) activities with the independent task than it did with the integrated task ($Mdn = 4.71$, 19.25, and 6.00, respectively; see Table 6 and Table D4). There were no significant differences across tasks for both groups in relation to the other main categories of writing activities.

Comparisons of categories between ELP groups for each task (Table D5) indicated that the proportions of main categories of writing activities relative to each other did not differ significantly across ELP groups for both tasks, except for planning and organizing for the independent task. Overall, participants with low ELP reported more planning and organizing activities ($Mdn = 18.42$) than did high ELP participants ($Mdn = 13.72$) with the independent task ($Z = 1.49$, $p < .05$, $r = .32$). See Appendix E for comparisons of subcategories of writing activities across ELP groups.

In terms of subcategories of writing activities (see Appendix E), the low-proficiency group reported checking the time more frequently than did the high-proficiency group with the independent task. The high-proficiency group reported using writing strategy to address writing difficulties more frequently than did the low-proficiency group with both tasks. Additionally, low-proficiency participants reported more interacting with task (mainly reflecting on writing task) with the independent task and more revising activities (particularly revising language) with the integrated task than did the high-proficiency group. The high group reported more interacting with sources (particularly referring to sources) and more evaluating activities with the integrated task and more revising activities (particularly revising language) with the independent task than did the low-proficiency group.

Variability in Writing Activities Across Keyboarding Skill Groups

Table 7 displays descriptive statistics for the main categories of writing activities across tasks and keyboarding skill groups (see Appendix F for results for the subcategories). Because nonparametric statistical tests do not allow examining the effects of more than one independent variable at a time or the examination of interaction effects, two statistical tests were conducted to compare writing activities across tasks and keyboarding skills groups. First, Wilcoxon signed-ranks tests were conducted to compare the proportions of the main categories of writing activities relative to each other between tasks for each keyboarding skill group separately (i.e., comparisons of writing activities across tasks for each group; see Table D6 in Appendix D). Second, two-sample Kolmogorov-Smirnov tests were conducted to compare the proportions of the main categories of writing activities across keyboarding skill groups for each task separately (i.e., comparisons of writing activities between groups for each task; see Table D7 in Appendix D).

Table 7 Descriptive Statistics for Percentage of Main Categories of Writing Activities by Task and Keyboarding Skill Group

Task	Integrated								Independent							
Keyboarding group	Low				High				Low				High			
Writing activity	*M*	*Mdn*	*Min*	*Max*	*M*	*Mdn*	*Min*	*Max*	*M*	*Mdn*	*Min*	*Max*	*M*	*Mdn*	*Min*	*Max*
Interacting with task	4.40	3.87	1.55	8.11	5.53	5.66	1.85	8.05	2.40	2.63	0.00	5.60	3.96	4.27	0.72	7.95
Interacting with sources	21.82	19.37	14.46	44.33	21.34	18.91	7.55	37.74	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Planning and organizing	8.08	8.43	1.84	15.32	9.24	8.52	1.89	16.88	14.33	13.72	5.36	24.56	15.50	15.94	6.82	20.55
Generating and retrieving	8.87	10.30	0.00	19.55	9.44	8.88	2.53	18.92	11.56	11.86	2.98	21.43	11.28	11.97	5.93	17.78
Writing difficulty	10.42	10.85	2.06	17.28	11.87	11.11	5.66	25.29	16.02	15.20	8.47	20.83	13.01	13.04	4.11	23.03
Using writing strategy	5.33	5.15	0.78	8.96	4.90	5.06	0.00	8.53	6.28	6.45	2.86	8.93	4.89	4.54	2.41	9.70
Evaluating	18.45	19.40	6.02	30.23	19.09	17.05	8.75	37.74	22.89	24.46	10.92	37.29	25.69	23.93	18.12	35.59
Revising	9.45	7.46	3.09	19.38	7.80	7.59	0.00	16.09	10.83	11.11	2.63	22.03	12.84	12.28	3.33	21.13
Procedural	13.14	14.63	4.94	18.07	10.75	11.62	1.15	15.09	15.66	15.74	2.54	25.60	12.79	13.76	2.74	23.33

Note. Interacting with sources is not indicated for the independent task because there is no source for this type of task.

Comparisons across tasks within each keyboarding skill group (Table D6) indicated that both keyboarding skill groups reported significantly fewer planning and organizing activities with the integrated task than they did with the independent task. Additionally, the low keyboarding skill group reported significantly more detecting writing difficulties with the independent task ($Mdn = 15.20$) than it did with the integrated task ($Mdn = 10.85$). The high keyboarding skill group reported significantly more evaluating ($Mdn = 23.93$) and revising ($Mdn = 12.28$) activities with the independent task than it did with the integrated task ($Mdn = 17.05$ and 7.59, respectively; see Table D6 and Table 7). There were no significant differences across tasks for both groups in relation to the other main categories of writing activities.

Comparisons of categories between keyboarding skill groups for each task (Table D7) indicated that the proportions of main categories of writing activities relative to each other did not differ significantly across keyboarding skill groups for both tasks. However, as Table 7 shows, generally, the low keyboarding skill group reported more procedural activities and fewer interacting with task activities than did the high keyboarding skill group with both writing tasks. Additionally, the high keyboarding skill group reported more planning and organizing and revising activities than did the low keyboarding skill group, which reported detecting writing difficulties and using writing strategy more frequently than the high keyboarding skill group with the independent task (see Table 7).

In terms of subcategories of writing activities (see Appendix F), the low keyboarding skill group tended to check the time more frequently than did the high keyboarding skill group with both writing tasks. With the independent task, participants with low keyboarding skills experienced more writing difficulties related to language and content and used more writing strategies than did participants with high keyboarding skills who tended to plan, both at the global and local levels, and to evaluate various writing aspects more frequently.

Relationships Between Writing Activities and Text Quality

To examine the relationships between writing activities and text quality, the correlations (Spearman rho) between the proportion of writing activities and scores for each writing task were computed. Generally, as Table 8 shows, participants who obtained higher scores reported more writing activities, particularly with the integrated task. However, neither the total number of reported writing activities nor the percentages of the main categories of writing activities correlated significantly with scores for both writing tasks. Table 8 shows also that participants who obtained higher scores on the integrated task tended to report more interacting with task, detecting writing difficulty, using writing strategy, and evaluating activities and fewer planning and organizing, generating and retrieving, revising, and procedural activities than did those with low scores. Generally, participants who scored higher on the independent task reported more interacting with task, detecting writing difficulties, using writing strategy, and revising activities and fewer activities related to planning and organizing and generating and retrieving than did those with low scores.

Table 8 also displays the correlations between the proportion of writing activities and scores for each writing phase for each task. While the sample size is small, Table 8 reveals some interesting patterns. First, the interacting with task activity and the scores correlated positively in the first phase and negatively in the last phase for the independent task. Second, the interacting with sources activity correlated positively with task scores in Phase 1 and negatively in Phases 2 and 3 for the integrated task. Third, the planning and organizing activity tended to correlate positively with the scores in Phase 1 and

Table 8 Correlations (Spearman rho) Between Percentages of Categories of Writing Activities and Computer-Based Task Scores by Writing Phase

Writing activity	CB integrated task score ($N = 22$)	CB independent task score ($N = 22$)
Grand total (raw frequency)	.37	.13
Interacting with task	.12	.16
Phase 1	−.05	.29
Phase 2	−.01	.00
Phase 3	.09	−.32
Interacting with sources	.06	
Phase 1	.24	N/A
Phase 2	−.22	N/A
Phase 3	−.24	N/A
Planning and organizing	−.32	−.35
Phase 1	.44	.18
Phase 2	−.25	.05
Phase 3	−.23	−.14
Generating and retrieving	−.14	−.10
Phase 1	−.17	−.22
Phase 2	.13	.12
Phase 3	.08	−.04
Detecting writing difficulty	.11	.16
Phase 1	−.07	−.06
Phase 2	−.01	.08
Phase 3	.01	.11
Using writing strategy	.36	.09
Phase 1	−.13	−.11
Phase 2	−.13	.41
Phase 3	.18	.03
Evaluating	.10	−.05
Phase 1	−.22	−.07
Phase 2	−.24	.35
Phase 3	.30	−.04
Revising	−.21	.12
Phase 1	−.32	−.14
Phase 2	.03	.05
Phase 3	.08	.03
Procedural	−.13	−.04
Phase 1	.11	.09
Phase 2	−.27	.07
Phase 3	−.04	−.24

Note. CB = computer based. Interacting with sources is not indicated for the independent task because there is no source for this type of task.

negatively and/or weakly in Phases 2 and 3 for both tasks. Generally, participants who reported planning and organizing their texts more frequently at the beginning of the writing process and less frequently later obtained higher scores. Fourth, the evaluating activity seems to correlate positively with task scores only in Phase 2 for the independent task and in Phase 3 for the integrated task. The revising activity and the scores were negatively correlated in Phase 1 for both tasks. Using the writing strategy activity correlated positively with task scores in Phase 2 for the independent task and Phase 3 for the integrated task.

Summary and Discussion

Writing Activities by Writing Task

This study used stimulated recalls to examine the writing activities prompted by the TOEFL iBT integrated and independent writing tasks as well as the effects of test-taker ELP and keyboarding skills on the frequency and distribution of these activities. The participants reported a wide range of writing activities with both tasks. The most frequently reported activities with both tasks related to evaluating (particularly evaluating language and evaluating local text), planning and

organizing (particularly local planning), detecting writing difficulty (particularly difficulties with language and content), revising (particularly revising language), using writing strategy, and procedural (particularly checking the time). Interacting with task was reported less frequently with both tasks. With the integrated task, the participants reported interacting with sources most frequently. In particular, they reported referring to sources (i.e., taking notes, paraphrasing and summarizing main ideas and details from lecture and/or text) and source-based generating (i.e., generating or retrieving ideas from the reading and/or lecture) frequently. With both tasks, the participants reported more evaluating activities than revising activities (cf. Whalen & Ménard, 1995) and more detecting writing difficulty than using writing strategy (to solve difficulties). Generally, participants who reported experiencing more writing difficulties reported using more writing strategies, and those who reported more evaluation activities reported more revision activities as well with both tasks.

Overall, the activities that the participants reported are part of the writing construct, as they are consistent with expectations regarding the processes that test takers would engage with when responding to independent and integrated writing tasks. This empirical evidence about the actual activities that test takers reported employing can be used to substantiate claims about the validity of inferences based on TOEFL iBT writing scores.

Comparisons across tasks indicated that the participants tended to check the time and interact with the writing task significantly more frequently with the integrated task. They reported that they tended to plan, particularly at the local level, to experience writing difficulties related to language and content and to evaluate language and local text significantly more frequently with the independent task. They also tended to generate and retrieve text (particularly generating or retrieving content from long-term memory, i.e., self-based generating) and to revise language slightly more frequently with the independent task. These findings are not surprising, given that the independent task requires generating and planning content and language as well as writing more extensively than does the independent task, which provides test takers with content, and possibly language and organization, for their texts (cf. Plakans, 2008, 2009). Consequently, the test takers experienced more difficulties related to language and content and had to evaluate and revise their texts more often with the independent task.

Variability in Writing Activities Across the Writing Process

Examination of the distribution of writing activities across the writing process suggests that the participants adopted a linear approach to writing with both tasks. With the independent task, the participants read and reflected on the writing task at the beginning of the writing process and then planned, generated, and wrote throughout the writing process. Consequently, they experienced writing difficulties and used writing strategies to address them throughout the writing process. At the end of the writing process, they tended to evaluate and revise their texts and to check time more frequently. With the integrated task, participants tended to read and reflect on the writing task and to interact with the sources at the beginning of the writing process and then to plan, generate, and write in the second and last phases of the writing session. Most of the writing difficulties the participants experienced and the writing strategies they used to address them occurred in the last two phases of the writing process. At the end of the writing process, the participants tended to evaluate and revise their texts and to check time more frequently. Finally, it seems that all groups, regardless of level of ELP and keyboarding skills, adopted the same linear approach to writing with both tasks.

Previous research suggests that writers are less likely to adopt a linear approach when writing on the computer than when they write on paper (e.g., Haas, 1989; van Waes & Schellens, 2003). While the findings of this study suggest that the frequency of writing activities varied across participants within each phase of the writing process, perhaps because some participants engaged in all writing activities in all phases, the predominant approach is a linear one with both tasks (i.e., interact with task and sources, generate, plan and write, and then evaluate and revise). There are several explanations for the linear approach adopted by the participants in this study. First, it is possible that this is the approach they usually use when writing in L2, perhaps as a result of their learning and writing beliefs and histories. Second, the structure and instructions of the tasks might have encouraged a linear approach to writing. Finally, the time constraints imposed by the test might have led the participants to adopt a linear approach to writing (cf. Hall, 1991).

Variability in Writing Activities Across English Language Proficiency Groups

The major differences between ELP groups concerned planning and organizing, interacting with task, detecting writing difficulties, evaluating, and revising. These differences affected mainly the independent task. First, participants with low

ELP reported significantly more planning and organizing activities than did high ELP participants with the independent task. Second, the low ELP group reported significantly more detecting writing difficulties with the independent task than it did with the integrated task. Third, the high ELP group reported significantly fewer interacting with task activities and significantly more evaluating and revising activities with the independent task than it did with the integrated task.

Additionally, the low-proficiency group reported checking the time more frequently than did the high-proficiency group with the independent task. The high-proficiency group reported using writing strategy to address writing difficulties more frequently than did the low-proficiency group with both tasks. Additionally, low-proficiency participants reported more interacting with task (mainly reflecting on writing task) with the independent task and more revising activities (particularly revising language) with the integrated task than did the participants in the high-proficiency group. The latter group reported more interacting with sources (particularly referring to sources) and more evaluating activities with the integrated task and more revising activities (particularly revising language) with the independent task than did the low-proficiency group.

These patterns suggest that responding to the writing tasks went more smoothly for the more proficient participants, who faced fewer writing difficulties, often were able to address the difficulties they encountered, and, consequently, did not need to worry about time as much as their less proficient counterparts did, particularly with the independent task. More proficient students also seem to be more effective in deciding which activities to engage in during each writing task. For example, they devoted more time to interacting with the writing task and the sources with the integrated task and less time to these activities with the independent task compared to the less proficient group. They also made more evaluations and revisions when they had to generate and revise their own content and language (i.e., with the independent task) than they did with the integrated task. The less proficient group, in contrast, tended to engage in more evaluation than revision activities with the independent task and devoted more time to revising language with the integrated task than they did with the independent task.

Variability in Writing Activities Across Keyboarding Skills Groups

There were two main significant differences across keyboarding skill groups in terms of the proportions of reported writing activities. First, the low keyboarding skill group experienced significantly more detecting writing difficulty with the independent task than it did with the integrated task. Second, the high keyboarding skill group reported significantly more evaluating and revising activities with the independent task than it did with the integrated task. There were also some differences in terms of subcategories of writing activities across keyboarding skill groups that were not significant but are worth mentioning. First, the low keyboarding skill group reported more procedural activities and less interacting with task than did the high keyboarding skill group with both writing tasks. Second, for the independent task, the high keyboarding skill group reported more planning and organizing and revising activities than did the low keyboarding skill group, which reported detecting writing difficulty and using writing strategy more frequently than did the high keyboarding skill group. In terms of subcategories of writing activities, the low keyboarding skill group reported more checking the time than did the high keyboarding skill group with both writing tasks. With the independent task, participants with low keyboarding skills experienced more writing difficulties related to language and content and used more writing strategies than did participants with high keyboarding skills, who tended to plan, both at the global and local levels, and to evaluate various writing aspects more frequently.

It seems that participants with low keyboarding skills were more worried about the time with both tasks, perhaps because they were concerned that they would not be able to complete their responses within the allotted time, given their low typing speed. They also interacted less frequently with the writing task, perhaps because they felt they needed to start planning and writing their responses soon so as not to waste time needed for typing their responses. The effects of keyboarding skills, though small, were more apparent with the independent task, which may have seemed to be more demanding than the integrated task as it required test takers not only to type their responses, but also to generate, plan, organize, and revise their own content and language. Consequently, participants with low keyboarding skills experienced more writing difficulties, particularly in relation to finding ideas, and needed to use more writing strategies with the independent task (compared to both the integrated task and to participants with high keyboarding skills). In contrast, participants with high keyboarding skills were able to devote more time to planning and evaluating their responses with the independent task compared to the participants in the low keyboarding skill group. Perhaps because the integrated task provided test takers with content and language to use in their writing, participants with low keyboarding skills were able

to devote relatively more time to interacting with sources as well as evaluating their texts when responding to this task. It seems thus that writing on the computer had a somewhat negative impact on the writing activities of some participants with low keyboarding skills with the independent task.

Relationships Between Writing Activities and Task Scores

Examination of the relationships between writing activities and text quality indicates that, while the relationships were not significant, perhaps because of small sample size, high-scoring participants tended to interact with the writing task more frequently, to report experiencing more writing difficulties, to use more writing strategies, and to engage in fewer planning and organizing and generating and retrieving activities, with both tasks. Additionally, participants who obtained higher scores on the integrated task tended to report more evaluating and procedural activities than did those with low scores, while participants who scored higher on the independent task reported more revising than did those with low scores. Revising thus seems to have played a more important role with the independent task than it did with the integrated task.

That high-scoring participants interacted more frequently with the writing task (i.e., reading and reflecting on the task) is not surprising, as this is likely to lead to a better understanding of task requirements and, hence, a more relevant and appropriate response. The higher proportion of writing difficulties and strategy use for the high-scoring participant might have occurred because these participants approached writing as a problem-solving activity that involved identifying and addressing several problems (rhetorical, linguistic, etc.), rather than as a knowledge-telling exercise. That high-scoring participants reported fewer planning and generating activities suggests that this group did not need to generate and plan frequently throughout the writing session. As discussed below, while overall planning correlated negatively with scores, planning in the first phase correlated positively with scores. Another possible explanation for the low proportion of planning and generating for high-scoring participants is that these processes were automatized in this group, and so they were reported less frequently. With the integrated task, high-scoring participants reported fewer revision activities, while with the independent task they reported more revision activities than did low-scoring participants. These patterns suggest that revising played a more important role when test takers had to generate their own content and language (i.e., with the independent task) than when they had to summarize the lecture and reading (i.e., with the integrated task).

The correlations between proportions of writing activities and text quality varied depending on when these activities were engaged in during the writing process. In particular, participants who reported interacting with the writing task and planning and organizing their texts more frequently at the beginning of the writing process and less frequently later obtained higher scores with both tasks (cf. van Weijen et al., 2008). This finding is consistent with previous research that shows that skilled writers tend to plan more at the beginning of the writing process (e.g., van der Hoeven, 1999a). Planning early in the writing process may reflect the development of a conceptual representation of the task as a whole, which is an important ingredient in the production of high-quality texts (van der Hoeven, 1999a). With the integrated task, interacting with sources at the beginning of the writing session was associated with higher scores. This might be the case because high-scoring participants started by writing down their ideas in relation to the lecture and reading text and planning their responses at the beginning of the writing session (i.e., immediately after listening to the lecture and reading the text) and then later focused on typing, evaluating, and revising their responses, as the positive correlations between scores and evaluating and revising in Phase 3 suggest. Generally speaking, the correlations between writing activities and task scores, including correlations across writing phases, do not indicate any unexpected patterns.

Overall, the findings of this study indicate that writing activities varied mainly across tasks and, to a lesser extent, across English proficiency groups, as expected by theory and research. Many of the activities reported by the participants (e.g., referring to sources, planning, generating, revising) are in some ways obvious, given theory, previous research, and the task types. As noted above, writing theory and research show that writers do engage in these activities when composing (e.g., Hayes & Flower, 1980; Scardamalia & Bereiter, 1987). Additionally, the TOEFL iBT validity argument rests on the assumption that the linguistic knowledge, processes, and strategies required to successfully complete the writing tasks vary across task types and with proficiency levels in keeping with theoretical expectations (Chapelle, 2008, pp. 336–337). The findings of this study thus provide empirical evidence to substantiate claims about the validity of inferences based on scores on the TOEFL iBT writing section. In contrast, while weak, the effects of keyboarding

skills on some participants' writing activities raise some concerns about the role of these skills in performance on TOEFL iBT writing tasks, particularly for test takers with low keyboarding skills when responding to the independent writing task. As noted above, Barkaoui (2014) found that keyboarding skills had a significant, though small, effect on scores on the independent task, but not on scores on the integrated task. The findings of this study shed some light on why this is the case. Specifically, test takers need not only to type their responses, but also to generate, plan, organize, evaluate, and revise the content, structure, and language of their responses with independent tasks, while integrated tasks provide test takers with both ideas and organization to apply in their writing (cf. Plakans, 2008), thus reducing the cognitive load for students with low keyboarding skills when writing on the computer in response to integrated tasks.

Implications for Further Research

The study has its limitations. In particular, it included small samples of participants and tasks, the writing tasks were not counterbalanced across participants, and data were collected for research purposes rather than in a real test setting. For example, the small sample size perhaps did not allow the detection of real differences across test-taker groups. Additionally, keyboarding skill was defined operationally for this study as typing speed and accuracy when copying, rather than composing, text. A copying task was used in order to estimate typing speed independent of the influence of writing ability or other cognitive abilities and processes. However, the operationalization of keyboarding skills in this study does not encompass what is arguably a broader and more critical skill: word processing. Future studies need to consider other aspects of word processing skills and their relation to performance on CB writing tasks.

Another limitation is that participants might have reported only some of the writing activities they engaged in during the test and/or reported other activities that they thought of during, or because of, the stimulated recall process. As previous research shows, participants can be selective in terms of what they report given the large number of activities they may employ at a given time and/or their awareness of an audience for their reports (Barkaoui, 2011; Cohen, 2011). As some participants mentioned at the end of the stimulated recalls, it was difficult for them to remember or describe in English everything they thought of or did during the test. Other participants felt that doing the stimulated recall with the integrated task somehow affected their performance during the independent task. One participant, for example, reported at the end of the stimulated recall that she "tried to be more structured" when responding to the independent task "so that [she] remember[s] everything later on" during the stimulated recall, while another mentioned that he took notes during the independent task of what he was doing so he could report his activities later in the stimulated recall. Furthermore, the reliability of stimulated recall of previous cognitive processes may in itself present measurement issues, since participants were probed for cognitive processes that had already transpired. Finally, this study examined only the frequency and distribution of writing activities but not their effectiveness, interrelations, or importance for the writing process or test takers' reasons for engaging in these activities (e.g., in relation to their interpretation of text audience and purpose). Additionally, the use of percentages and nonparametric statistical tests, though appropriate given the small sample size and the characteristics of the data, could have affected the findings of the study and prevented the examination of the effects of interactions between task type, ELP level, and keyboarding skill level on the writing activities that the participants employed.

To address some of these limitations, the current dataset could be explored further. In particular, case studies could be conducted with a small subsample of participants (e.g., participants with highest and/or lowest scores) to examine the relationships between their pausing and revision behaviors as recorded by Morae, the quality and sequencing of their writing activities as reported in the stimulated recalls, the characteristics of their final texts, their scores, and their background data (e.g., L1). To obtain a clearer picture of the indirect relationships between writing activities and scores, future analyses could examine how writing activities relate to the text produced in real time by linking the writing activities and decisions reported in the stimulated recalls to the written text as it emerges on the screen and then examining how the characteristics of the final written text relate to scores.

The study points to several areas for further research. First, future studies could compare the writing performance of test takers with different levels of L2 proficiency and keyboarding skills when writing on paper and on the computer to determine whether writing on paper eliminates some of the negative effects associated with low keyboarding skills when writing on the computer. As noted above, writing on the computer seems to call for a different distribution of writing activities compared to writing on paper (Haas, 1989; Lee, 2004; van Waes & Schellens, 2003). Second, comparing test

takers' writing activities under test and nontest conditions can also shed light on the effects of test conditions on test takers' writing performance. Two key differences between writing in academic settings and writing in a test like the one examined in this study are that in real-life settings writers usually (a) have more time to plan, write, and revise their texts and (b) have access to various forms of support such as writing and editing tools (e.g., spell and grammar checkers, thesauruses, dictionaries, references). In the version of the test used in this study, the participants had a limited amount of time to write and had access to only three editing functions (cut, paste, and undo). Future studies could compare the writing activities that L2 writers engage in when they have (a) different amounts of time to complete their responses and/or (b) access to other writing and editing tools. Comparisons of L2 learners' writing activities on the computer in test and nontest settings (e.g., when writing a take-home paper for a course) can enhance our understanding of the effects of testing conditions on test takers' writing performance and provide important evidence concerning the extrapolation inference in the validity argument of TOEFL iBT and similar CB L2 writing tests.

Acknowledgments

I am grateful to the following research assistants who helped with data collection and analysis for this study: Angelpreet Singh, Anna Martinez, Gerrenne Gunthrope, Mara Reich, Reza Ahmadilari, and Xiaohui Zhu. Thanks are also due to the participants and to three anonymous reviewers for their comments on earlier drafts of this report. This study was funded by the TOEFL Research Program at Educational Testing Service. The opinions expressed in the report are those of the author.

Notes

1 Cohen (2011) defined strategies in L2 learning and use as "thoughts and actions, consciously chosen and operationalized by language learners, to assist them in carrying out a multiplicity of tasks from the very onset of learning to the most advanced levels of target-language performance" (p. 7).
2 A process is automated "if it occurs without voluntary control or interferes minimally with other processes" (Torrance & Galbraith, 2006, p. 74).
3 The original plan was to include 24 participants (six students by two keyboarding skill levels [high and low] by two ELP levels [high and low]). However, it was very difficult to identify and recruit (a) students with low ELP and high keyboarding skills and (b) students with high ELP and low keyboarding skills (see Barkaoui, 2014). Every fourth student volunteering for participating in the main study in each group was selected to do the stimulated recalls. This resulted in a sample of 23 participants. However, one participant with low English proficiency and high keyboarding skills did not complete all the tasks and was excluded from the study.
4 Using percentages has its drawbacks as well. In particular, higher percentages might hide lower (absolute) frequencies and vice versa, and all percentages that add up to 100% become interdependent, which can cause interpretation problems and make correlational analyses problematic.
5 The Kolmogorov-Smirnov two-sample test is a nonparametric equivalent of the two-sample *t*-test.
6 The Wilcoxon signed-rank test is a nonparametric equivalent of the matched-pairs *t*-test.
7 The Friedman test is a nonparametric equivalent of a repeated-measures ANOVA.
8 $r = Z/\sqrt{N}$ where N is the total number of observations on which Z is based (Field, 2009, p. 550).

References

Alamargot, D., & Chanquoy, L. (2001). *Through the models of writing*. New York, NY: Kluwer Academic.

Alves, R. A., Castro, S. L., de Sousa, L., & Stromqvist, S. (2007). Influence of typing skill on pause-execution cycles in written composition. In M. Torrance, L. van Waes, & D. Galbraith (Eds.), *Writing and cognition: Research and applications* (pp. 55–65). Amsterdam, The Netherlands: Elsevier.

Baker, E., & Kinzer, C. K. (1998). Effects of technology on process writing: Are they all good? *National Reading Conference Yearbook*, *47*, 428–440.

Barkaoui, K. (2011). Think-aloud protocols in research on essay rating: An empirical study of their veridicality and reactivity. *Language Testing*, *28*(1), 51–75.

Barkaoui, K. (2014). Examining the impact of L2 proficiency and keyboarding skills on scores on TOEFL iBT writing tasks. *Language Testing*, *31*(2), 241–259.

Barkaoui, K., Brooks, L., Swain, M., & Lapkin, S. (2013). Test-takers' strategic behaviors in independent and integrated speaking tasks. *Applied Linguistics*, *34*(3), 304 – 324.

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.

Bourdin, B., & Fayol, M. (1994). Is written language production more difficult than oral language production? A working memory approach. *International Journal of Psychology*, *29*, 591 – 620.

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.

Breland, H., Lee, Y., & Muraki, E. (2004). *Comparability of TOEFL CBT writing prompts*: *Response mode analyses* (Research Report No. RR-04-23). Princeton, NJ: Educational Testing Service. http://dx.doi.org/10.1002/j.2333-8504.2004.tb01950.x

Broekkamp, H., & van den Bergh, H. (1996). Attention strategies in revising a foreign language text. In G. Rijlaarsdam, H. van den Bergh, & M. Couzijn (Eds.), *Theories, models and methodology in writing research* (pp. 170 – 181). Amsterdam, The Netherlands: Amsterdam University Press.

Burke, J. N., & Cizek, G. J. (2006). Effects of composition mode and self-perceived computer skills on essay scores of sixth graders. *Assessing Writing*, *11*, 148 – 166.

Chapelle, C. A. (2008). The TOEFL validity argument. In C. A. Chapelle, M. K. Enright, & J. M. Jamieson (Eds.), *Building a validity argument for the Test of English as a Foreign Language* (pp. 319 – 352). New York, NY: Routledge.

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

Clachar, A. (1999). It's not just cognition: The effect of emotion on multiple-level discourse processing in second-language writing. *Language Sciences*, *21*, 31 – 60.

Cohen, A. D. (1994). English for academic purposes in Brazil: The use of summary tasks. In C. Hill & K. Parry (Eds.), *From testing to assessment: English as an international language* (pp. 174 – 204). New York, NY: Pearson Education.

Cohen, A. D. (2011). *Strategies in learning and using a second language: Research and practice* (2nd ed.). Harlow, England: Longman/Pearson Education.

Cohen, A. D. (2012). Test-taking strategies and task design. In G. Fulcher & F. Davidson (Eds.), *The Routledge handbook of language testing* (pp. 262 – 277). New York, NY: Routledge.

Cohen, A. D., & Upton, T. A. (2007). I want to go back to the text: Response strategies on the reading subtest of the new TOEFL. *Language Testing*, *24*, 209 – 250.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). New York, NY: Academic Press.

Connelly, V., Gee, D., & Walsh, E. (2007). A comparison of keyboard and handwritten compositions and the relationship with transcription speed. *British Journal of Educational Psychology*, *77*, 479 – 492.

Cumming, A. (1989). Writing expertise and second-language proficiency. *Language Learning*, *39*, 81 – 141.

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

Cumming, A., Kantor, R., Powers, D., Santos, T., & Taylor, C. (2000). *TOEFL 2000 writing framework: A working paper* (TOEFL Monograph Series No. 18). Princeton, NJ: Educational Testing Service.

Esmaeili, H. (2002). Integrated reading and writing tasks and ESL students' reading and writing performance in an English language test. *The Canadian Modern Language Review*, *58*, 599 – 620.

Fayol, M. (1999). From online management problems to strategies in written composition. In M. Torrance & G. Jeffery (Eds.), *The cognitive demands of writing: Processing capacity and working memory effects in text production* (pp. 13 – 23). Amsterdam, The Netherlands: Amsterdam University Press.

Field, A. (2009). *Discovering statistics using SPSS* (3rd ed.). Thousand Oaks, CA: Sage.

Gass, S. M., & Mackey, A. (2000). *Stimulated recall methodology in second language research*. Mahwah, NJ: Lawrence Erlbaum.

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

Hall, E. (1991). Variations in composing behaviors of academic ESL writers in test and non-test situations. *TESL Canada Journal*, *8*, 9 – 33.

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* (pp. 1 – 27). Mahwah, NJ: Lawrence Erlbaum.

Hayes, J. R. (2006). New directions in writing theory. In C. A. MacArthur, S. Graham, & J. Fitzgerald (Eds.), *Handbook of writing research* (pp. 28 – 40). New York, NY: Guilford Press.

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

Higgins, J., Russell, M., & Hoffman, T. (2005). Examining the effect of computer-based passage presentation on reading test performance. *The Journal of Technology, Learning and Assessment*, *3*(4). Retrieved from www.jtla.com

Horkay, N., Bennett, R. E., Allen, N., Kaplan, B., & Yan, F. (2006). Does it matter if I take my writing test on computer? An empirical study of mode effects in NAEP. *The Journal of Technology, Learning and Assessment, 5*(2). Retrieved from http://ejournals.bc.edu/ojs/index.php/jtla/issue/view/180

Jamieson, J. M., Eignor, D., Grabe, W., & Kunnan, A. J. (2008). Frameworks for the new TOEFL. In C. A. Chapelle, M. K. Enright, & J. M. Jamieson (Eds.), *Building a validity argument for the Test of English as a Foreign Language* (pp. 55 – 95). New York, NY: Routledge.

Kellogg, R. T. (1996). A model of working memory in writing. In C. M. Levy & S. Ransdell (Eds.), *The science of writing: Theories, methods, individual differences and applications* (pp. 57 – 71). Hillsdale, NJ: Erlbaum.

Krapels, A. R. (1990). An overview of second language writing process research. In B. Kroll (Ed.), *Second language writing: Research insights for the classroom* (pp. 37 – 56). Cambridge, England: Cambridge University Press.

Lee, H. K. (2004). A comparative study of ESL writers' performance in a paper-based and a computer-delivered writing test. *Assessing Writing*, *9*, 4 – 26.

Lee, Y. (2002). A comparison of composing processes and written products in timed-essay tests across paper and pencil and computer modes. *Assessing Writing*, *8*, 135 – 157.

Lee, Y., & Kantor, R. (2005). *Dependability of new ESL writing scores: Evaluating prototype tasks and alternative rating schemes* (Research Report No. RR-05-14). Princeton, NJ: Educational Testing Service. http://dx.doi.org/10.1002/j.2333-8504.2005.tb01991.x

Li, J. (2006). The mediation of technology in ESL writing and its implications for writing assessment. *Assessing Writing*, *11*, 5 – 21.

Lindgren, E. (2005). *Writing and revising: Didactic and methodological implications of keystroke logging*. (Doctoral thesis). Umea University, Sweden. Retrieved from http://umu.diva-portal.org/smash/record.jsf?pid=diva2%3A143740&dswid=−4719

Mateos, M., Martin, E., Villalon, R., & Luna, M. (2008). Reading and writing to learn in secondary education: Online processing activity and written products in summarizing and synthesizing tasks. *Reading and Writing: An Interdisciplinary Journal*, *21*, 675 – 697.

Maycock, L., & Green, T. (2005). The effects on performance of computer familiarity and attitudes towards CB IELTS. *Research Notes*, *20*, 3 – 8.

McCourt Larres, P., Ballantine, J. A., & Whittington, M. (2003). Evaluating the validity of self-assessment: Measuring computer literacy among entry-level undergraduates within accounting degree programmes at two universities. *Accounting Education: An International Journal*, *12*, 97 – 112.

McCutchen, D. (1996). A capacity theory of writing: Working memory in composition. *Educational Psychology Review*, *8*(3), 299 – 325.

McCutchen, D. (2000). Knowledge, processing, and working memory: Implications for a theory of writing. *Educational Psychologist*, *35*, 13 – 23.

Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13 – 103). New York, NY: Macmillan.

Olive, T., & Kellogg, R. T. (2002). Concurrent activation of high- and low-level production processes in written composition. *Memory and Cognition*, *30*, 594 – 600.

Pennington, M. C. (1996). *The computer and the non-native writer: A natural partnership*. Cresskill, NJ: Hampton Press.

Phinney, M., & Khouri, S. (1993). Computers, revision and ESL writers: The role of experience. *Journal of Second Language Writing*, *2*, 257 – 277.

Plakans, L. (2008). Comparing composing in writing-only and reading-to-write test tasks. *Assessing Writing*, *13*, 111 – 129.

Plakans, L. (2009). The role of reading strategies in integrated L2 writing tasks. *Journal of English for Academic Purposes*, *8*, 252 – 266.

Raimes, A. (1987). Language proficiency, writing ability, and composing strategies: A study of ESL student writers. *Language Learning*, *37*, 439 – 468.

Read, J. (1990). Providing relevant content in an EAP writing test. *English for Specific Purposes*, *9*, 109 – 121.

Roca de Larios, J., Manchón, R., Murphy, L., & Marín, J. (2008). The foreign language writers' strategic behaviour in the allocation of time to writing processes. *Journal of Second Language Writing*, *17*(1), 30 – 47.

Roca de Larios, J., Murphy, L., & Manchón, R. (1999). The use of restructuring strategies in EFL writing: A study of Spanish learners of English as a foreign language. *Journal of Second Language Writing*, *8*(1), 13 – 44.

Roca de Larios, J., Murphy, L., & Marin, J. (2002). A critical examination of L2 writing process research. In S. Ransdell & M. L. Barbier (Eds.), *New directions for research in L2 writing* (pp. 11 – 47). Dordrecht, The Netherlands: Kluwer Academic Publishers.

Russell, M., & Haney, W. (1997). Testing writing on computers: An experiment comparing student performance on tests conducted via computer and via paper-and-pencil. *Education Policy Analysis Archives*, *5*(3), 1 – 18.

Russell, M. (1999). Testing on computers: A follow-up study comparing performance on computer and on paper. *Education Policy Analysis Archives*, *7*(20). Retrieved from http://epaa.asu.edu/epaa/v7n20

Sasaki, M. (2000). Toward an empirical model of EFL writing processes: An exploratory study. *Journal of Second Language Writing*, *9*, 259 – 291.

Sasaki, M. (2002). Building an empirically-based model of EFL learners' writing processes. In S. Ransdell & M.-L. Barbier (Eds.), *New directions for research in L2 writing* (pp. 49 – 80). Dordrecht, The Netherlands: Kluwer Academic Publishers.

Scardamalia, M., & Bereiter, C. (1987). Knowledge telling and knowledge transforming in written composition. In S. Rosenberg (Ed.), *Advances in applied psycholinguistics: Reading, writing, and language learning* (Vol. 2, pp. 142 – 175). Cambridge, England: Cambridge University Press.

Shaw, S. (2005). Evaluating the impact of word processed text on writing quality and rater behaviour. *Cambridge ESOL: Research Notes*, *22*, 13 – 19.

Shaw, S. D., & Weir, C. (2007). *Examining writing: Research and practice in assessing second language writing*. Cambridge, England: Cambridge University Press.

Slattery, P. J., & Kowalski, R. (1998). On screen: The composing processes of first-year and upper-level college students. *Computers and Composition*, *15*, 61 – 81.

Standards Australia. (2001). *Keyboarding speed tests* (3rd ed.). Sydney, Australia: Author.

Stevenson, M., Schoonen, R., & de Glopper, K. (2006). Revising in two languages: A multi-dimensional comparison of online writing revisions in L1 and FL. *Journal of Second Language Writing*, *15*, 201 – 233.

Taylor, C., Jamieson, J., Eignor, D., & Kirsch, I. (1998). *The relationship between computer familiarity and performance on computer-based TOEFL test tasks* (TOEFL Research Report. No. TOEFL-RR-61). Princeton, NJ: Educational Testing Service. http://dx.doi.org/10.1002/j.2333-8504.1998.tb01757.x

Torrance, M., & Galbraith, D. (2006). The processing demands of writing. In C. A. MacArthur, S. Graham, & J. Fitzgerald (Eds.), *Handbook of writing research* (pp. 67 – 80). New York, NY: Guilford Press.

van der Hoeven, J. (1999a). Differences in writing performance: Generating as an indicator. In M. Torrance & D. Galbraith (Eds.), *Knowing what to write: Conceptual processes in text production* (pp. 65 – 77). Amsterdam, The Netherlands: Amsterdam University Press.

van der Hoeven, J. (1999b). Children's composing: A study of the relationships between planning and revision skill, writing processes, text quality and linguistic skills. In G. Rijlaarsdam, H. van den Bergh, & M. Couzijn (Eds.), *Theories, models and methodology in writing research* (pp. 108 – 120). Amsterdam, The Netherlands: Amsterdam University Press.

van Waes, L., & Schellens, P. J. (2003). Writing profiles: The effect of the writing mode on pausing and revision patterns of experienced writers. *Journal of Pragmatics*, *35*, 829 – 853.

van Weijen, D., van den Bergh, H., Rijlaarsdam, G., & Sanders, T. (2008). Differences in process and process-product relations in L2 writing. *International Journal of Applied Linguistics*, *156*, 203 – 226.

Wang, L., Eignor, D., & Enright, M. K. (2008). A final analysis. In C. A. Chapelle, M. K. Enright, & J. M. Jamieson (Eds.), *Building a validity argument for the Test of English as a Foreign Language* (pp. 259 – 318). New York, NY: Routledge.

Weigle, S. C. (2002). *Assessing writing*. Cambridge, England: Cambridge University Press.

Weir, C. J. (2005). *Language testing and validation: An evidence based approach*. Basingstoke, England: Palgrave.

Weir, C. J., O'Sullivan, B., Jin, Y., & Bax, S. (2007). Does the computer make a difference? The reaction of candidates to a computer-based versus a traditional hand-written form of the IELTS writing component: Effects and impact. *IELTS Research Report, 7,* 311 – 347. Canberra, Australia: British Council & IDP Australia.

Whalen, K., & Ménard, N. (1995). L1 and L2 writers' strategic and linguistic knowledge: A model of multiple-level discourse processing. *Language Learning*, *45*(3), 381 – 418.

Wolfe, E. W., Bolton, S., Feltovich, B., & Niday, D. M. (1996). The influence of student experience with word processors on the quality of essays written for a direct writing assessment. *Assessing Writing*, *3*, 123 – 147.

Wolfe, E. W., & Manalo, J. R. (2005). *An investigation of the impact of composition medium on the quality of TOEFL writing scores* (Research Report No. RR-04-29). Princeton, NJ: Educational Testing Service. http://dx.doi.org/10.1002/j.2333-8504.2004.tb01956.x

Appendix A
Description of Typing Skills Test

Each participant completed two online typing tests (through www.assesstyping.com) to assess their typing speed and accuracy in English. Each typing test consisted of typing a 200-word passage, presented at the upper half of the computer screen, into a blank text box located at the lower half of the screen (cf. Higgins et al., 2005). The participants were instructed to type each text as quickly and as accurately as possible within 2 minutes. The students did not have access to any editing functions when typing the texts, but they were instructed to do a practice test before doing the actual typing tests. The online test provides three measures of typing skills (cf. Horkay et al., 2006; Russell, 1999):

- *Gross Typing Speed* is calculated by dividing the total number of keystrokes (i.e., characters, spaces and punctuation marks) by test duration (2 minutes) to obtain gross speed in keystrokes per minute (KPM). Gross speed in KPM is then divided by the standard word length to get typing speed in WPM. The commonly accepted standard word length in English typing tests (e.g., Standards Australia, 2001) is five keystrokes, including spaces and punctuation marks (i.e., 5 KPM = 1 WPM). This measure is not adjusted for typing errors.
- *Net Typing Speed* is typing speed adjusted for typing accuracy. It is computed by (a) subtracting the number of incorrect words times word length (i.e., 5) from the total number of keystrokes to get the total net keystrokes for the whole test duration and then (b) dividing the total net keystrokes by the test duration to get the net typing speed in KPM. This is then divided by the standard word length (i.e., 5 keystrokes) to obtain net typing speed in WPM.
- *Accuracy Percentage* is calculated by computing the rate of net keystrokes to gross keystrokes. Five keystrokes are deducted for each mistyped word, regardless of the number of mistakes in it. The error penalty is the number of words typed incorrectly times the word length (i.e., 5). For example, 10 incorrectly typed words result in an error penalty of 50. An accuracy percentage of 75% means that three quarters of the words (that the student typed in 2 minutes) were typed correctly. Accuracy is rounded to the next lower whole score. For example, if the accuracy is 97.7%, it is reported as 97% (Standards Australia, 2001).

The following is an example of how the three measures are computed. If a test taker completed a typing test with 500 keystrokes in 2 minutes and 10 incorrectly typed words. The results are as follows:

- Gross Typing Speed: 500 keystrokes/2 = 250 KPM; 250 KPM/5 = 50 WPM.
- Net Typing Speed: 500 – 50 error penalty (10 mistakes x word length of 5)/2 = 450/2 = 225 KPM/5 = 45 WPM.
- Accuracy Percentage: 450/500 x 100 = 90%.

TypingMaster.com (personal communication, November 29, 2010) reported that the mean gross typing speed for a sample of 15,000 test takers was 35 WPM (*SD* = 10). TypingMaster.com recommended using a net typing speed of 40 WPM and an accuracy percentage of 95% as a cut-off score, with everyone typing above these cut scores being considered to have high typing speed and everyone below these cut scores being considered to have low typing speed. Following this recommendation, two cut scores were set for this study. First, to be classified into the high typing skills group, a test taker needed to achieve a net typing speed of 40 WPM or more. In order to distinguish typing skill groups, a decision was made to include in the low typing speed group only those volunteers with a net typing speed that was one *SD* below the cut score for the high skills group. Consequently, only volunteers with net typing speed of 30 WPM (i.e., 40 WPM – *SD* 10) or less were included in the low typing speed group.

Appendix B
Stimulated Recall Instructions (Adapted From Gass & Mackey, 2000)

Instructions to Research Participant

Before Doing the Writing Task

In this study, I am interested in learning what you think about as you carry out the two writing tasks administered on the computer. To do this, I am going to first record your writing process on the computer. After each task you complete, I am going to play back the recording on the computer and ask you to recall and say out loud everything that came into your mind while completing each writing task.

After Finishing Each Writing Task

We are going to watch a video of your writing session. We are interested in what you were thinking **at the time you were writing**. We can see what you were doing and writing by looking at the video, but we don't know what you were thinking then. So, I'd like you to watch the video and to tell me what you were thinking, what was in your mind **at that time** while you were writing. As you watch the video, try to put your mind back into the task. I am interested in finding out **what you were thinking when you were writing**, and it does not matter at all to me if those thoughts were silly or profound. *If you want to point at something on the screen please use the mouse* (not your finger) to point at it (so the mouse movement can be recorded too).

It is important that you do not plan or try to explain to me what you are thinking, and it is important that you keep talking all the time. If you are silent for any period of time, I will remind you to keep talking.

I am going to put the mouse on the desk here, and you can pause the video any time that you want. So, if you want to tell me something about what you were thinking at that time, you can click the pause button. If I have a question about what you were thinking or what you have written, then I will push pause and ask you to think back and tell me what you were thinking at that time. You can pause the video as often as you want. Anytime you remember something, say it; interrupt me; stop the video if you want.

I will record what you say on the computer using this microphone and this digital recorder.

Do you understand what I am asking you to do? Do you have any questions?

Now we are going to watch the recording of your writing process for the writing task you just did. I'd like you to tell me what you were thinking **when you were completing the task**, **NOT** what you are thinking now. I am interested in **what was in your mind** from the point when you read the instructions for the test up until the time when you finished the task. Please do not think about what you should have thought or done. Again, I am interested in knowing **what you were thinking at the time you were writing the essay**, **NOT** what you think about it now.

Please go ahead and tell me what you can remember. [Researcher then starts the video]

Instructions to Researcher

After reading the instructions to the participant, model stopping the video and asking a question. For example, choose a segment and stop the video. Ask your question. If the participant stops the video, listen to what s/he says. If you stop the video, ask something general, like:

- *What were you thinking here/at this point/right then?*
- *Can you tell me what you were thinking at that point?*
- *Is there anything else that comes to your mind?*
- *Do you remember anything else about what you were thinking at that moment?*

Use only open-ended prompts/questions.

Make sure to ask the student what they were thinking even during the lecture and reading (before they start writing) with the integrated task and when reading the prompt in the independent task.

If the participant starts to talk about what s/he is thinking now (e.g., that s/he made a mistake), try to maintain orientation to time of writing, for example, by saying "were you thinking that at the time?" Keep him/her focused on the time when s/he did the writing. Emphasize thoughts **during the writing**, not interpretation now.

If the participant says "I don't remember," accept the comment and move on. If the participant cannot recall the item at once, do not ask any other questions. Continue to watch the video. Try not to focus or direct participants' attention beyond "what were you thinking then/at that time." It may also be useful to direct participants' attention to the pauses and revisions they make by saying something like:

- *I see you stopped writing, what were you thinking then?*
- *I see you changed (added, deleted, reordered, etc.) the text, can you tell me what you were thinking then?*
- *Can you tell me your thoughts when you paused (or made a change)?*

If the participant begins to talk without pausing the video, pause the video and angle the mouse toward the participant so that s/he can release the pause when s/he is finished talking.

Additionally, you should not give concrete reactions to participants' responses. Back-channeling or nonresponse are preferable. For example: *mhm, I see, good, uh-huh, ok,*

Appendix C
Coding Scheme for Stimulated Recalls With Examples

Code	Definition	Examples (student code, task)
Interacting with task		
Reading test instructions		I read the direction and after the question [arranges the scratch papers] sorry when I read the direction I couldn't understand am I requested to write essay or a paragraph short paragraph. (EC, Integrated)
		First thing I reading the direction reading direction and what (I need to do). (TN, Independent)
Reading the writing task	Reading or rereading the task	Mmm I listen to the question and then I read it again that time I'm reading it again and thinking how should I start. (SA, Integrated)
		I was reading carefully the question because uh I remember one of the directions said they don't ask for my personal opinion but just the information from the lecture and the paragraph, so I yeah was reading carefully the question. (DF, Integrated)
Reflecting on the writing task	Analyzing the writing task in terms of its requirements (e.g., text purpose, audience, and structure) and/or evaluating the task in regard to its difficulty/easiness or interest	So, I thought the topic is, good for me because it is it wasn't hard for write something. (EC, Independent)
		The summarizing the word summarizing like it bothers me because summarizing means not writing the whole thing again it's just writing the main points so that's why I was doing. (CL, Integrated)
Interacting with sources		
Referring to sources	Referring to reading and/or lecture (e.g., taking notes from lecture/text, summarizing main idea of text/lecture)	Yeah I'm looking for the paragraph to find information, to add my paragraph, so I, actually I wrote exactly same thin- same thing with the paragraph same thing in my uh paragraph. (EF, Integrated)
		Then I was like carefully listening and writing the points what was not in the text like some there was like 2 or 3 things he said said other than in the text so I was like writing them down. (GL, Integrated)
Reflecting on sources	Reflecting on, analyzing, and/or evaluating the reading and/or lecture	I just wait uh uh and the uh the organization of the lecture, is no, I mean the, uh, the organi, organi is very well so I just they, uh how to say they separate they separate the the uh each idea of the topic very clear (DL, Integrated).
		When like reading I just first like went scan like went through what this is about before like going through the whole uh text […] they have like 3 theories and I was thinking of that and keeping that in mind because I I was thinking that they might ask a question regarding these theories so I was like going through every the theories and from and then again I like felt very comfortable once it because it looks easy and the text is it's not hard to read or hard to understand it's easy and then I it's OK and then I wrote some point there. (GL, Integrated)
Integrating sources	Establishing connections between ideas in the reading and the lecture	And through the reading and writ- through the reading and listening I found out the differences which is uh the listening part is refute the reading part […] I think which the listening part will refute the reading part. (DZ, Integrated)
		So at first I thought the speech would be different than the paragraph but, then I I figured out it it talks about same things uh same things with the paragraph. (EC, Integrated)
Checking comprehension	Checking or confirming comprehension of reading text/lecture	I found it was, uh the lecturer said when they use this uh uh he calls [TEXT] uh it's not possible for them to know they don't know always that how far their destination is they cannot find it without compass thing [laugh], so uh actually I think I forgot what the lecturer meant by this, I was trying to say something for myself and uh match that with. (ST, Integrated)
		In the in the first part I think it's ok for me for I I understand the meaning yeah I understand the meaning of the first part. (PH, Integrated)

Appendix C Continued

Code	Definition	Examples (student code, task)
Detecting comprehension difficulty	Referring to a comprehension problem/difficulty in relation to the reading and/or lecture	I read the from the beginning to the to the end, uh I stopped here because I I tried to read again what the topic about because it's not a familiar topic to me so some word I can't understand. (DL, Integrated)
		I'm trying to (pause) here because I'm just don't know what what exactly the text is about. (DZ, Integrated)
Using comprehension strategy	Verbalizing the use of a reading/listening problem-solving strategy to address a comprehension difficulty	I have to reread x from the reading material because uh mmm because of the listening, if I if I summarize really good words from reading passage I can prepare more about listening so I concentrated to choose uh the mmm the word. (SM, Integrated)
		So I was the the, after I after I finish reading the first thing yeah and uh I'm try to read one more time yeah to understand more so, and at that time I'd uh I was concentrate about x reading about the paragraph. (PH, Integrated)
Reacting to sources	Reacting in personal and affective way to reading and/or lecture or referring to personal or emotional states	I thought the lecture is longer uh than I assumed but uh so so because the lecture is not so long uh I was a little surprised in this time. (HT, Integrated)
		Maybe in this time I read everything a little bit boring [laughs]. (KS, Integrated)
Mining	Rereading text for ideas, words, phrases, etc. to use them in own writing	I decided to add how how can how [TEXT] so I look looked at pa-passage again to find specific information yes so I decided to add the hippohippocampal region in their brain xxx I decided to add this information. (EC, Integrated)
		Mmm at this moment, I am reading the first first paragraph and try to thinking trying to get the main idea the main point and try to find some mmm useful words to then I can use in my essay. (FF, Integrated)
Planning and organizing		
Global planning	Planning and choosing the purpose, audience, content, and structure of the writing in regard to units longer than a paragraph	I first I start the the the art- answer is about the lecture summarize of the lecture and explain how the lecture begins but I think it's hard to write three hundred words so I think whether I should put some summary about the lecture too. (WG, Integrated)
		Then Mmm how can I prepare my idea for essay so [TEXT] [laughs] making making society better and that's a concern to society. (SM, Independent)
Local planning	Planning and/or choosing what or how to write the next clause or sentence	I stopped to writing a couple minutes because I was thinking what what I should write what I should write what I should write how should I how I should link my reasons to the to the topic so … (EC, Independent)
		I'm kind of organized it well and uh, to make sure I wrote the examples clearly, and the you can see the the first theory I made from the text part I only wrote one sentence to describe it. (DZ, Integrated)
Introduction planning	Planning of the introduction	I just try to think about how to uh, begin my introduction, Uh I try to uh make the connection between the topic of the essay to the introduction here. (DL, Independent)
		I was writing the reason why I'm agree with this statement but I delete it then I write it again and then I still delete it because I'm not sure sh- whether I should write the reason in the introduction introduction so I'm trying to make it uh general to this topic but I was just not sure what should I make so. (DZ, Independent)
Conclusion planning	Planning of the conclusion	Mmm when I when I was writing the conclusion I feel that it's much similar (with) the beginning, I wanted to make it make it different, I was thinking how to make it different. (WG, Integrated)
		I come to uh, the last paragraph to write a conclusion, and it's just like a the introduction but I paraphrasing paraphrase it. (TT, Independent)

Appendix C Continued

Code	Definition	Examples (student code, task)
Generating and retrieving		
Plan retrieving	Retrieving an already constructed plan	I was thinking my significant what should I write for my significant part, I was looking at my paper that I that I wrote down something and so I started to write my first paragraph, uh and so I started to explain my first reason. (EC, Independent)
		There I was I know the answer but I was excited and like thinking of the thinking in of best way of putting the theory into short form that's I was like looking at the notes and thinking. (GL, Independent)
Self-based generating	Generating an idea without any stimulus or retrieving appropriate information from long-term memory	I had this idea that at least from my experience people around me tend to think about the second option so that's what I put uh THEY SPEAK ABOUT LEARN WHAT IS USEFUL things like that uh so I put that MY OPINION PEOPLE TEND TO BELIEVE uh of course I I mmm through my head I had this thought that probably it's not all the people but at least from my experience so I uh yeah like I have clearly the idea I should put this in my opinion. (DF, Independent)
		I start I brainstorm some uh different kind of ideas like what should I write anything I was I was think about. (DZ, Independent)
Text-based generating	Generating an idea related to what has been written	And I reread what I am writing to continue. (DL, Independent)
		Something came up uh when I wrote the (before) sentence so so and I wanted to uh write more words so I put I add this I added this sentence here. (HT, Independent)
Source-based generating	Generating an idea based on the reading and/or lecture or retrieving information from text and/or lecture	I starting to write the uh start started to write it and first I try to mmm write the yes the main point of the … lecture uh that the … people cannot people cannot [TEXT], so I try to write the main reason. (HT, Integrated)
		And I was like thinking there was the brain part it's in the in the passage they say they have a special kind of memory in their they have part in the brain that they makes them more unique other than anyone else I was thinking about that. (GL, Integrated)
Detecting writing difficulty		
Difficulty with content	Referring to a problem/difficulty related to content (not knowing what to write/say)	Uh I was looking at this paragraph again to see how how can I conclude it like how should I write the like what what information I didn't cover in the previous uh mmm the sentences. (DZ, Independent)
		And now I begin my first argument in this essay, I just thinking about how to write the topic sentence here, and that's take me a little bit. (DL, Independent)
Difficulty with language	Referring to a problem/difficulty related to language (grammar, vocabulary, punctuation, etc.)	I was just uh fixing the mistake but uh uh I I mainly just fix the INTERESTING and INTERESTED because I'm not sure about this So I just I'm finding the word "INTERESTING" or "INTERST" to make sure that I use the right one. I think about change the word "EXCITING" or "EXCITED" That's the same thing, I'm not sure. (DL, Independent)
		Oh yeah and just like BY MISTAKE, mmm yeah I was still thinking how to make sure that MISTAKE would clearly refer to the previous phrase uh I could (see it) I think I have some doubts if it's not reflected here because I write anyways (then) I couldn't erase so but I was having some doubts about letting that (SILLY) there. (DF, Independent)
Difficulty with rhetoric	Referring to a problem/difficulty related to organizing paragraphs, sentences, or ideas and/or balance (style, register)	Just thinking about but I don't know what I need to write in the writing test x is different from what I was study in school because I didn't write about summarize the (point) in the lecture I didn't study about this and in my school I just studied about we had a topic sentence we had a topic and just write about this in the way you want, and I don't have any knowledge about summarize. (TN, Integrated)

Appendix C Continued

Code	Definition	Examples (student code, task)
		Then just try try to maybe after this I just write down without second re- without second reason but second is not clear and the time is up uh I am thinking what what's the structure for this. (SM, Independent)
		I was thinking do I need should I find another way to make it shorter and ea- easier to understand but I just wrote the many words. (DZ, Integrated)
		And again I use the word , which I take it from the question, I wanted here to mention that the SOCIETY NEEDS DIFFERENT KINDS OF JOBS and I was thinking how can I how can I say it, I I in fact I was thinking that I'm using the word WHICH a lot but I couldn't replace it with any other word at that point. (MA, Independent)
Using writing strategy	Verbalizing the use of a problem solving strategy to address a writing difficulty	I can't understand the idea here so I try to uh read the paragraph again, and read the note again, and then I try to remember what I listening from lecture, but this is very difficult to me, at that time yeah. (DL, Integrated)
		I just use my experience and something I imagine some uh some use some something I just imagine it uh to write to write this, maybe this this point I know fifty percent so I try to finish. (FF, Integrated)
Evaluating		
Evaluating local text	Evaluating part of the generated text (e.g., a sentence, word), with or without revision	I were thinking that GLAD is not a formal word, and VARIOUS is not good. (WG, Independent)
		I wrote the last paragraph I was reading the paragraph, I found a lots of mistakes there. (SA, Independent)
Evaluating global text	Evaluating the generated text at a global level (paragraph and above) , with or without revision	I just look at the whole paragraph whole pa- the essay again to see whether it looks good and it's only one minute so so I cannot write any other information. (DZ, Independent)
		I just finishing it up I knew that everything was correct. (MM, Integrated)
Evaluating content	Evaluating content or ideas written or not written yet, with or without revision	So I I decided to change yeah idea yeah I decided to change uh change my explanation about the second paragraph because I (thought) it is not it was not enough so I change my sentence. (EC, Integrated)
		I think about the example I think uh maybe it's a little difficult because uh the the point is very sh- uh specific so is not very general so I choose the specific uh example for supporting is to play chess. (JZ, Independent)
Evaluating language	Evaluating language, with or without revision	I add something here because I put ALSO in the end but I think it's better to put in the in the front so I change a little bit. (DZ, Independent)
		OK I erased ACTUALLY [laughs], it's not necessary. (DF, Independent)
Evaluating rhetoric	Evaluating the organization and/or balance (style, register) of the text, with or without revision	Here I wanted I wanted to add this idea but I think it's better to mention it as my first example here I think that it sounds logically if I add it first and the consequence is, here. (MA, Independent)
		I realized that this paragraph wasn't fitting in over here it should come afterwards so I decided to have another paragraph before that with some other reasons so make it easy (to) make it x better. (MM, Independent)
		Uh because [laugh] I'm just because I use a lot of CAN THEY CAN DO SOMETHING I CAN DO SOMETHING so I'm trying to use BE ABLE TO so instead of CAN maybe this more different (it's a) little bit different but just want to change some words. (DZ, Independent)
		I was trying to see about how I could phrase it in a way which explains what I'(ll be) talking about that doesn't give off the whole meaning so that people have to read the paragraphs and. (MM, Independent)
Evaluating text length		I was aware that for sure it is less than three hundred much shorter but I thought maybe a shorter complete essay is better than a longer but unfinished without conclusion essay. (MA, Independent)
		Then I decided to add a clarifying sentence to that, and after which I realized that the maximum word of the essay is (531) the essay (can't be) bigger than that, as so as soon as I realize I went to see what section I could remove. (MM, Independent)
Reacting to own writing	Reacting in personal and affective way to own writing	I think I uh had finished uh I have enough word and the time is enough so maybe I think uh I'm happy at that time. (JZ, Independent)
		I think I was lost in [laugh] between summarizing and writing. (ST, Integrated)

Appendix C Continued

Code	Definition	Examples (student code, task)
Revising		
Revising content	Revising the informational content of the text at a global or local level	Uh OK this one is like IF YOU DO NOT HAVE ENOUGH KNOWLEDGE X THING but then, later on I thought OK it's not appropriate, uh uh again, it just repeating the previous sentence if you want to be FINANCIAL ANALYST, yeah you need a strong foundation, if you say again, you do not have enough knowledge you won't be able to do it it's just repeating it, so so uh I delete it and start a new idea, like I chose three x rather than just repeating it. (TT, Independent)
		I don't have much time here to change everything so, here I thought just add a few more sentences, if I uh choosing why (pursuing you passion makes you successful) say something about that. (SP, Independent)
Revising language	Revising language (i.e., spelling, grammar, vocabulary, punctuation and format, or phrasing)	So at that time I still have a mistake this word I think that's were eliminated And I realize that's maybe at this time and then I change it to uh ELIMINATE? Elimi-, and then I try to find a word that I write in the first and then I change that. (DL, Integrated)
		I was like going through the passage what I have written and (I) just correcting the I in the beginning I started with the simple (thing) and I corrected that and I was going through my words the grammar I mean the spellings, correcting some mistakes. (GL, Integrated)
Revising rhetoric	Changing the order of paragraphs or sentences and/or revising text to make it more (or less) appropriate for the intended reader (e.g., style, politeness, genre)	Yeah I was again like I said I realized that this paragraph wasn't fitting in over here it should come afterwards so I decided to have another paragraph before that with some other reasons so make it easy (to) make it x better. (MM, Independent)
		I guess after writing it I thought this is not the right place for this sentence to be it should come at some later stage in the write-up that I push it back and edit. (SP, Independent)
		At the first time I want to use the study but I change the (do) after finally I choose the one yes as I say I want to avoid using the same word. (KS, Independent)
		So just I try to explain my support and I think x is too strong so I put LIKELY LIKELY LESS LIKELY ah but it's the wrong sentence at that time I think the sentence is too strong so I think at the time I think. (SM, Independent)
Procedural		
Describing actions	Referring to the writing and reading procedures	Yeah I think I move to the listening. (DZ, Integrated)
		So sometimes I stopped sometimes I typed again. (EC, Independent)
Verbalizing a proposition	Verbalizing or saying the content the writer has written or intends to write	Uh this is my first idea and then I try to write the second one and the last one, FEEL MORE EXCITING that 's my second idea here. (DL, Independent)
		And then I wrote my first point there, THAT PEOPLE ARE DIFFERENT AND NOT EVERYONE IS SAME AND THEY HAVE DIFFERENT SKILLS so it was my first point and I was writing that. (GL, Independent)
Checking the time	Checking the time	Mmm at this moment I don't care the the time because I know at this moment I have a lot of time to writing so I think I can finish it before the time. (FF, Independent)
		Yeah I was looking at the time I was thinking how many minutes I have left so then start writing. (CL, Independent)

Note. The passages are marked with the following transcription conventions: ( ) = uncertain transcription; x = incomprehensible word; comma = short pause; capital letters = words that the student has written, was thinking of writing, and/or read directly from his/her text; italics = text read from task; underlined = text read directly from the reading text or heard from the lecture; [ ] = procedural and other behaviors; ? = questioning intonation; [TEXT] = text read from task, reading text, and/or lecture (not included here for test security reasons).

Appendix D
Results of Nonparametric Statistical Tests

Table D1 Wilcoxon Signed-Ranks Tests for Comparisons of Main Categories Across Tasks

Main category	*Z*	Exact sig. (2-tailed)
Interacting with task	−2.16	.03
Planning and organizing	3.52	.00
Generating and retrieving	1.61	.11
Detecting writing difficulty	2.19	.03
Using writing strategy	0.41	.69
Evaluating	2.94	.00
Revising	2.26	.02
Procedural	1.09	.28

Table D2 Friedman Tests for Comparisons of Main Categories Across Writing Phases by Task

Task	Integrated ($n = 22$)			Independent ($n = 22$)		
Main category	Chi-square	*df*	Exact sig.	Chi-square	*df*	Exact sig.
Interacting with task	23.17	2	.00	31.16	2	.00
Interacting with sources	19.55	2	.00	N/A	N/A	N/A
Planning and organizing	9.46	2	.00	7.42	2	.03
Generating and retrieving	29.47	2	.00	2.36	2	.31
Detecting writing difficulty	25.28	2	.00	5.37	2	.07
Using writing strategy	29.04	2	.00	1.89	2	.39
Evaluating	32.64	2	.00	12.88	2	.00
Revising	34.46	2	.00	9.63	2	.01
Procedural	3.95	2	.14	12.03	2	.00

Note. Interacting with resources does not apply to the independent task because this task does not require reading or listening.

Table D3 Wilcoxon Signed-Ranks Tests for Comparisons of Main Categories Across Writing Phases by Task (Only for Significant Results From Table D2)

Task	Integrated ($n = 22$)		Independent ($n = 22$)	
Main category and phase	*Z*	Exact sig. (2-tailed)	*Z*	Exact sig. (2-tailed)
Interacting with task				
Phase 1 vs. Phase 2	−3.92[a]	.00	−4.14[a]	.00
Phase 1 vs. Phase 3	−3.38[a]	.00	−4.01[a]	.00
Phase 2 vs. Phase 3	−0.70	.48	−1.72	.09
Interacting with sources				
Phase 1 vs. Phase 2	−3.39[a]	.00		
Phase 1 vs. Phase 3	−3.68[a]	.00		
Phase 2 vs. Phase 3	−2.65[a]	.01		
Planning and organizing				
Phase 1 vs. Phase 2	−1.98	.05	−2.60[a]	.01
Phase 1 vs. Phase 3	−2.09	.04	−2.86[a]	.00
Phase 2 vs. Phase 3	−0.31[a]	.76	−0.84[a]	.40
Generating and retrieving				
Phase 1 vs. Phase 2	−3.92	.00		
Phase 1 vs. Phase 3	−3.87	.00		
Phase 2 vs. Phase 3	−1.83[a]	.07		
Detecting writing difficulty				
Phase 1 vs. Phase 2	−3.95[a]	.00		
Phase 1 vs. Phase 3	−3.89	.00		
Phase 2 vs. Phase 3	−0.65	.52		

Table D3 Contiuned

Task	Integrated ($n = 22$)		Independent ($n = 22$)	
Main category and phase	Z	Exact sig. (2-tailed)	Z	Exact sig. (2-tailed)
Using writing strategy				
Phase 1 vs. Phase 2	−4.02	.00		
Phase 1 vs. Phase 3	−3.73	.00		
Phase 2 vs. Phase 3	−0.79[a]	.43		
Evaluating				
Phase 1 vs. Phase 2	−4.02	.00	−2.35	.02
Phase 1 vs. Phase 3	−4.04	.00	−3.07	.00
Phase 2 vs. Phase 3	−3.10	.00	−1.95	.05
Revising				
Phase 1 vs. Phase 2	−3.83	.00	−2.25	.02
Phase 1 vs. Phase 3	−4.02	.00	−2.82	.00
Phase 2 vs. Phase 3	−3.17	.00	−1.72	.09
Procedural				
Phase 1 vs. Phase 2			−1.35	.18
Phase 1 vs. Phase 3			−2.69	.01
Phase 2 vs. Phase 3			−2.19	.03

Note. A Bonferroni correction was applied to pairwise comparisons across phases so that only effects with $p < .017$ are considered significant.

[a] Based on positive ranks. Results for other categories are based on negative ranks.

Table D4 Wilcoxon Signed-Ranks Tests for Comparisons of Main Categories Across Tasks by English Language Proficiency (ELP) Group

ELP group	Low ($n = 11$)		High ($n = 11$)	
Main category	Z	Exact sig. (2-tailed)	Z	Exact sig. (2-tailed)
Interacting with task	−0.89[a]	.37	−1.96[a]	.05
Planning and organizing	−2.63	.01	−2.81	.00
Generating and retrieving	−1.17	.24	−1.75	.08
Detecting writing difficulty	−1.99	.05	−1.60	.11
Using writing strategy	−0.89	.37	−0.76	.45
Evaluating	−1.33	.18	−2.40	.02
Revising	−1.24	.21	−2.13	.03
Procedural	−0.71	.48	−0.80	.42

[a] Based on positive ranks. Results for other categories are based on negative ranks.

Table D5 Two-Sample Kolmogorov-Smirnov Test for Comparisons of Main Categories Across English Language Proficiency Groups by Task

	Most extreme differences				
Writing activity	Absolute	Positive	Negative	Kolmogorov-Smirnov Z	Exact sig. (2-tailed)
Integrated task					
Interacting with task	.27	.27	−.18	.64	.81
Interacting with sources	.27	.27	−.18	.64	.81
Planning and organizing	.27	.09	−.27	.64	.81
Generating and retrieving	.36	.09	−.36	.85	.46
Detecting writing difficulty	.36	.18	−.36	.85	.46
Using writing strategy	.27	.27	.00	.64	.81
Evaluating	.36	.36	−.09	.85	.46
Revising	.45	.18	−.45	1.07	.21
Procedural	.18	.18	−.18	.43	.99
Independent task					
Interacting with task	.36	.09	−.36	.85	.46

Table D5 Continued

Writing activity	Most extreme differences			Kolmogorov-Smirnov Z	Exact sig. (2-tailed)
	Absolute	Positive	Negative		
Planning and organizing	.64	.09	−.64	1.49	.02
Generating and retrieving	.27	.27	−.18	.64	.81
Detecting writing difficulty	.27	.27	−.18	.64	.81
Using writing strategy	.36	.36	.00	.85	.46
Evaluating	.45	.18	−.45	1.07	.21
Revising	.45	.45	−.09	1.07	.21
Procedural	.36	.18	−.36	.85	.46

Table D6 Wilcoxon Signed-Ranks Tests for Comparisons of Main Categories Across Tasks by Keyboarding Group

Keyboarding skill group	Low ($n = 11$)		High ($n = 11$)	
Main category	*Z*	Exact sig. (2-tailed)	*Z*	Exact sig. (2-tailed)
Interacting with task	−1.90[a]	.06	−0.82[a]	.41
Planning and organizing	−2.81	.00	−2.71	.01
Generating and retrieving	−1.28	.20	−1.83	.07
Detecting writing difficulty	−1.99	.05	−1.25	.21
Using writing strategy	−0.54	.59	−1.32	.19
Evaluating	−1.25	.21	−2.50	.01
Revising	−0.80	.42	−2.22	.03
Procedural	−0.89	.37	−0.71	.48

[a] Based on positive ranks. Results for other categories are based on negative ranks.

Table D7 Two-Sample Kolmogorov-Smirnov Test for Comparisons of Main Categories Across Keyboarding Groups by Task

Main category	Most extreme differences			Kolmogorov-Smirnov *Z*	Exact sig. (2-tailed)
	Absolute	Positive	Negative		
Integrated task					
Interacting with task	.27	.27	−.09	.64	.81
Interacting with sources	.18	.18	−.18	.43	.99
Planning and organizing	.18	.18	−.09	.43	.99
Generating and retrieving	.27	.27	−.27	.64	.81
Detecting writing difficulty	.18	.18	−.09	.43	.99
Using writing strategy	.27	.09	−.27	.64	.81
Evaluating	.27	.18	−.27	.64	.81
Revising	.18	.09	−.18	.43	.99
Procedural	.36	.00	−.36	.85	.46
Independent task					
Interacting with task	.36	.36	.00	.85	.46
Planning and organizing	.27	.27	−.18	.64	.81
Generating and retrieving	.18	.18	−.09	.43	.99
Detecting writing difficulty	.45	.09	−.45	1.07	.21
Using writing strategy	.45	.09	−.45	1.07	.21
Evaluating	.36	.36	−.18	.85	.46
Revising	.27	.27	−.09	.64	.81
Procedural	.27	.09	−.27	.64	.81

Appendix E
Descriptive Statistics for Subcategories of Writing Activities by Task and English Language Proficiency Group

Table E1 Integrated Task

English language proficiency group	Low ($n = 11$)				High ($n = 11$)			
Writing activity	*M*	*Mdn*	*Min*	*Max*	*M*	*Mdn*	*Min*	*Max*
Interacting with task								
Reading test instructions	1.31	1.21	0.00	2.70	1.42	1.14	0.00	3.77
Reading the writing task	1.65	1.23	0.00	5.43	1.44	1.48	0.00	3.90
Reflecting on writing task	2.04	2.17	0.00	5.41	2.06	1.50	0.00	4.65
Interacting with sources								
Referring to sources	4.47	4.34	1.89	7.32	7.04	6.97	2.99	14.94
Reflecting on sources	2.04	2.17	0.00	5.41	2.06	1.50	0.00	4.65
Integrating sources	2.20	1.25	0.00	6.48	1.54	1.14	0.00	4.51
Checking comprehension	1.97	1.80	0.00	5.06	1.29	1.03	0.00	4.72
Comprehension difficulty	3.47	2.70	0.00	11.39	3.82	3.44	0.00	9.43
Comprehension strategy	2.07	1.26	0.00	6.48	2.81	2.25	0.00	9.28
Reacting to sources	1.31	1.26	0.00	2.50	1.25	0.74	0.00	6.19
Mining	2.68	2.53	0.94	5.41	3.80	2.83	0.78	8.27
Planning and organizing								
Global planning	2.00	2.40	0.00	4.35	2.49	2.29	0.00	5.19
Local planning	5.79	6.09	1.23	8.86	4.07	4.59	0.78	7.79
Introduction planning	0.81	0.61	0.00	3.60	0.46	0.00	0.00	2.60
Conclusion planning	1.04	0.00	0.00	3.80	0.64	0.00	0.00	3.45
Generating and retrieving								
Plan retrieving	1.81	1.20	0.00	5.00	3.27	3.89	0.00	6.90
Self-based generating	1.08	1.21	0.00	3.60	1.49	0.00	0.00	8.25
Text-based generating	0.31	0.00	0.00	2.53	0.07	0.00	0.00	0.78
Source-based generating	6.61	6.52	0.00	13.51	3.64	3.44	0.00	10.53
Detecting writing difficulty								
Content	3.98	3.60	.00	9.88	2.94	2.29	0.00	13.79
Language	6.26	6.17	2.70	11.96	7.68	7.79	2.06	16.30
Rhetoric	0.60	0.00	0.00	1.80	0.80	0.00	0.00	3.45
Using writing strategy								
Using writing strategy	4.66	5.00	0.00	8.64	5.57	5.74	0.78	8.96
Evaluating								
Local text	5.61	4.50	0.00	13.21	4.43	3.44	1.03	10.08
Global text	1.44	1.35	0.00	3.70	1.10	1.14	0.00	3.10
Content	2.26	1.26	0.00	7.41	1.76	1.55	0.00	3.45
Language	4.77	3.79	1.25	15.09	4.04	3.10	1.15	9.30
Rhetoric	1.18	1.23	0.00	2.83	1.38	0.77	0.00	4.65
Text length	1.85	1.35	0.00	6.75	3.39	1.03	0.00	14.73
Reacting to own writing	1.40	1.25	0.00	3.80	2.85	2.23	0.00	11.32
Revising								
Content	1.94	1.35	0.00	5.06	1.10	0.75	0.00	5.43
Language	6.80	6.52	2.53	11.04	6.66	5.42	0.00	14.94
Rhetoric	0.59	0.00	0.00	2.44	0.14	0.00	0.00	1.55
Procedural								
Describing actions	3.52	2.83	0.00	9.78	3.17	1.55	0.00	9.28
Verbalizing a proposition	3.07	2.70	0.00	8.49	2.54	2.96	0.00	6.72
Checking the time	5.80	5.52	0.00	10.81	5.78	5.97	1.15	13.18

Table E2 Independent Task

English language proficiency group	Low ($n = 11$)				High ($n = 11$)			
Writing activity	*M*	*Mdn*	*Min*	*Max*	*M*	*Mdn*	*Min*	*Max*
Interacting with task								
Reading test instructions	0.20	0.00	0.00	2.27	0.29	0.00	0.00	0.98
Reading the writing task	1.49	1.13	0.00	5.48	0.78	0.87	0.00	1.68
Reflecting on the writing task	2.04	1.85	0.00	4.80	1.54	1.11	0.00	4.39
Planning and organizing								
Global planning	4.51	4.54	1.06	8.77	3.47	2.63	0.00	10.00
Local planning	9.84	10.95	2.27	18.28	8.02	8.33	2.86	17.78
Introduction planning	1.01	0.87	0.00	4.39	0.76	0.92	0.00	1.43
Conclusion planning	1.44	1.75	0.00	2.74	0.76	0.92	0.00	1.45
Generating and retrieving								
Plan retrieving	1.60	2.12	0.00	4.27	0.69	0.00	0.00	2.90
Self-based generating	8.45	9.63	0.85	11.02	9.45	8.77	2.98	17.14
Text-based generating	1.13	1.06	0.00	2.41	0.72	0.00	0.00	4.29
Source-based generating	0.23	0.00	0.00	1.75	0.53	0.84	0.00	1.21
Detecting writing difficulty								
Content	4.80	3.50	0.80	10.53	4.89	4.34	0.70	10.00
Language	6.34	6.81	1.37	11.70	9.66	8.88	4.29	18.79
Rhetoric	2.21	1.75	0.00	5.60	1.11	0.92	0.00	4.24
Using writing strategy								
Using writing strategy	5.03	5.12	2.41	7.89	6.13	5.93	2.86	9.70
Evaluating								
Local text	8.01	8.47	3.61	10.96	7.92	7.24	0.00	13.89
Global text	1.10	1.07	0.00	5.08	1.51	1.11	0.00	6.14
Content	3.91	3.50	0.00	8.60	2.48	2.38	0.00	7.02
Language	5.81	5.37	0.88	9.65	6.88	6.66	0.00	13.89
Rhetoric	2.21	1.06	0.00	8.47	1.53	1.40	0.00	3.39
Text length	2.11	1.69	0.00	5.56	1.74	1.22	0.00	4.93
Reacting to own writing	1.73	0.80	0.00	10.84	1.58	0.70	0.00	6.14
Revising								
Content	1.48	0.92	0.00	5.93	2.36	1.68	0.00	7.89
Language	8.26	7.22	0.88	17.81	10.29	10.18	3.33	16.20
Rhetoric	0.64	0.00	0.00	3.39	0.62	0.59	0.00	1.69
Procedural								
Describing actions	1.02	0.84	0.00	3.61	0.60	0.00	0.00	1.79
Verbalizing a proposition	7.94	7.52	0.00	17.60	10.10	9.09	0.60	22.69
Checking the time	5.31	4.25	0.00	13.64	3.46	1.68	0.00	11.43

Appendix F
Descriptive Statistics for Subcategories of Writing Activities by Task and Keyboarding Skill Group

Table F1 Integrated Task

Keyboarding skill group	Low ($n = 11$)				High ($n = 11$)			
Writing activity	*M*	*Mdn*	*Min*	*Max*	*M*	*Mdn*	*Min*	*Max*
Interacting with task								
Reading test instructions	0.92	0.77	0.00	2.47	1.81	1.29	0.74	3.77
Reading the writing task	1.14	0.77	0.00	5.43	1.95	1.55	0.00	3.90
Reflecting on the writing task	2.33	2.17	0.78	5.41	1.76	1.14	0.00	4.65
Interacting with sources								
Referring to sources	5.61	6.76	2.45	7.32	5.90	4.05	1.89	14.94
Reflecting on sources	2.33	2.17	0.78	5.41	1.76	1.14	0.00	4.65
Integrating sources	1.71	1.23	0.00	4.51	2.04	1.25	0.00	6.48
Checking comprehension	1.52	1.80	0.00	3.61	1.73	1.25	0.00	5.06
Comprehension difficulty	3.04	2.45	0.00	9.28	4.25	3.44	0.00	11.39
Comprehension strategy	2.45	1.80	0.00	9.28	2.42	1.48	0.00	6.60
Reacting to sources	1.66	1.55	0.00	6.19	0.89	0.92	0.00	2.50
Mining	3.58	3.26	0.78	8.27	2.90	2.53	0.94	8.15
Planning and organizing								
Global planning	2.55	2.43	0.00	4.65	1.94	1.35	0.00	5.19
Local planning	4.47	4.65	0.78	8.11	5.39	5.40	1.48	8.86
Introduction planning	0.72	0.00	0.00	3.60	0.55	0.00	0.00	2.60
Conclusion planning	0.33	0.00	0.00	1.22	1.35	0.77	0.00	3.80
Generating and retrieving								
Plan retrieving	1.85	1.20	0.00	6.02	3.23	3.89	0.00	6.90
Self-based generating	1.85	1.21	0.00	8.25	0.72	0.00	0.00	2.60
Text-based generating	0.08	0.00	0.00	0.90	0.30	0.00	0.00	2.53
Source-based generating	5.08	5.42	0.00	12.05	5.18	3.77	0.00	13.51
Detecting writing difficulty								
Content	2.83	3.06	0.00	9.88	4.09	2.59	0.00	13.79
Language	6.82	6.09	2.06	11.96	7.13	6.48	2.70	16.30
Rhetoric	0.76	0.77	0.00	1.80	0.64	0.00	0.00	3.45
Using writing strategy								
Using writing strategy	5.33	5.15	0.78	8.96	5.00	5.06	0.00	8.53
Evaluating								
Local text	5.09	4.34	1.03	10.08	4.95	3.70	0.00	13.21
Global text	1.16	0.90	0.00	3.70	1.38	1.29	0.00	3.10
Content	2.12	1.55	0.00	7.41	1.90	1.55	0.00	4.72
Language	4.19	3.60	1.50	9.30	4.62	3.79	1.15	15.09
Rhetoric	0.87	0.77	0.00	2.44	1.69	1.26	0.00	4.65
Text length	3.19	1.03	0.00	14.73	2.06	1.14	0.00	10.34
Reacting to own writing	1.80	1.80	0.00	6.19	2.46	1.29	0.00	11.32
Revising								
Content	1.97	1.21	0.00	5.43	1.08	0.94	0.00	5.06
Language	7.04	6.30	2.06	12.40	6.42	6.25	0.00	14.94
Rhetoric	0.44	0.00	0.00	2.44	0.29	0.00	0.00	1.89
Procedural								
Describing actions	4.01	4.81	0.75	9.78	2.68	2.50	0.00	7.79
Verbalizing a proposition	3.06	3.00	0.00	6.72	2.55	1.25	0.00	8.49
Checking the time	6.06	5.97	0.00	13.18	5.51	5.06	1.15	11.32

Table F2 Independent Task

Keyboarding skill group	Low ($n = 11$)				High ($n = 11$)			
Writing activity	*M*	*Mdn*	*Min*	*Max*	*M*	*Mdn*	*Min*	*Max*
Interacting with task								
Reading test instructions	0.16	0.00	0.00	0.98	0.33	0.00	0.00	2.27
Reading the writing task	0.89	0.84	0.00	2.63	1.38	1.13	0.00	5.48
Reflecting on the writing task	1.34	0.98	0.00	4.80	2.24	2.54	0.00	4.55
Planning and organizing								
Global planning	3.77	2.77	1.06	10.00	4.21	4.34	0.00	8.77
Local planning	8.59	8.33	2.86	18.28	9.27	9.09	2.27	17.78
Introduction planning	1.01	0.92	0.00	4.39	0.76	0.87	0.00	1.71
Conclusion planning	0.95	1.06	0.00	2.40	1.25	1.21	0.00	2.74
Generating and retrieving								
Plan retrieving	0.91	0.00	0.00	2.63	1.38	0.87	0.00	4.27
Self-based generating	9.33	8.77	2.98	17.14	8.57	9.42	0.85	15.56
Text-based generating	1.05	0.92	0.00	4.29	0.80	0.84	0.00	2.41
Source-based generating	0.24	0.00	0.00	0.98	0.52	0.00	0.00	1.75
Detecting writing difficulty								
Content	5.60	6.45	0.80	10.53	4.09	3.50	0.70	9.09
Language	8.53	7.52	4.24	14.81	7.47	6.81	1.37	18.79
Rhetoric	1.87	1.68	0.00	5.60	1.44	1.13	0.00	5.13
Using writing strategy								
Using writing strategy	6.28	6.45	2.86	8.93	4.89	4.54	2.41	9.70
Evaluating								
Local text	7.96	8.47	0.00	13.89	7.97	7.69	3.61	13.56
Global text	1.05	0.00	0.00	5.08	1.56	1.36	0.00	6.14
Content	3.06	2.38	0.00	8.60	3.33	2.56	0.00	7.02
Language	6.24	5.37	0.00	13.89	6.45	6.66	3.51	10.17
Rhetoric	1.66	0.92	0.00	8.47	2.07	2.17	0.00	4.39
Text length	1.67	1.42	0.00	5.56	2.19	1.75	0.00	4.93
Reacting to own writing	1.21	0.84	0.00	5.95	2.10	0.00	0.00	10.84
Revising								
Content	1.49	0.92	0.00	5.93	2.35	1.20	0.00	7.89
Language	8.64	9.24	0.88	13.10	9.91	9.64	3.33	17.81
Rhetoric	0.70	0.59	0.00	3.39	0.57	0.00	0.00	1.75
Procedural								
Describing actions	0.46	0.00	0.00	1.79	1.16	0.85	0.00	3.61
Verbalizing a proposition	9.57	7.52	0.60	22.69	8.47	9.09	0.00	22.22
Checking the time	5.61	5.35	0.00	12.04	3.16	2.40	0.00	13.64

Suggested citation:

Barkaoui, K. (2015). *Test takers' writing activities during the TOEFL iBT® writing tasks: A stimulated recall study* (TOEFL iBT Research Report No. 25, ETS Research Report No. RR-15-04). Princeton, NJ: Educational Testing Service. 10.1002/ets2.12050

Action Editor: Gary Ockey

Reviewers: This report was reviewed by the Research Subcommittee of the TOEFL Committee of Examiners.