The Effect of Task Type on Accuracy and Complexity in IELTS Academic Writing

Nguyễn Thúy Lan*

Faculty of English Teacher Education, VNU University of Languages and International Studies, Phạm Văn Đồng, Cầu Giấy, Hanoi, Vietnam

Received 30 August 2014
Revised 23 January 2015; Accepted 06 March 2015

Abstract: IELTS is one of the most popular international standardized tests of English language proficiency. Its two academic writing tasks are crucially different in cognitive and linguistic demands, but to date, few studies have compared the influence of their different task demands on test-takers’ performance. In second language research (L2) area, two contrasting theories on task demands are the Limited Attentional Capacity Model which predicts a worse linguistic performance on a more complex task and the Cognition Hypothesis which expects a better performance on a more demanding task. My study examines the effect of task type as an important factor of task complexity on L2 writing in a testing condition. The study was a single-factor, repeated-measures design which compares the performance of 30 L2 writers on task 1 and task 2 of the IELTS Academic writing subtest. The candidates’ writing samples were analyzed using a range of discourse measures focusing on accuracy and complexity. The findings showed that low demanding task (task 1 - graph description) elicited a significantly better performance in terms of accuracy than high demanding task (task 2 - argumentative essay). Meanwhile, the latter was more complex in terms of grammatical subordination and lexical variation. The current study contributes exploratory findings to the body of knowledge on L2 writing by investigating task complexity embedded in different task types. The use of discourse measurement of accuracy and complexity revealed some IELTS candidates’ language problems related to genre writing. The gained knowledge may help teachers manipulate task features to channel learners’ attention to the area in which they fail.

Keywords: Language testing, writing assessment, IELTS, task type, genre writing, discourse measurement, accuracy, complexity.

1. Introduction

1.1. Context of the study

IELTS, the International English Language Testing system, is an international standardized test of English language proficiency. IELTS plays an important role in many people’s life as it involves critical decisions such as admission to universities or immigration. The IELTS writing tasks are designed to be “communicative and contextualized for a specified audience, purpose, and genre”, which
reflects the growing focus of second language (L2) writing research on genres/task [1: 2].

Studies have compared the effect of different genres on learners’ writing performance, but few have investigated into the impact of visual description (Task 1) in contrast with argumentative essays (Task 2). In addition, the previous genre-related studies are mostly classroom-based, but similar investigations in a testing situation, especially in IELTS writing, are still scarce [2]. Furthermore, in SLA research area, two contrasting theories on attentional resources, i.e. the Limited Attentional Capacity Model and the Cognition Hypothesis, have been often examined by manipulating task complexity along planning, here-and-now variables, task prompts and draft availability; meanwhile, few studies investigate task complexity embedded in different task type. Finally, IELTS is a high-stakes test, so it is essential to diagnose candidates’ possible difficulties to prepare them better. However, despite extensive research concerning the test in general, few studies specifically focus on its writing component [1]. The additional problem is that the IELTS analytic assessment scale does not give much information for predicting candidates’ language problems. As noted by Mickan [3], it is difficult to identify specific lexicogrammatical features that distinguish different band scores. Storch [4] also confirms that analytical scores are often collapsed to yield a single score, losing diagnostic value.

This study is thus motivated by (i) the lack of research comparing the effect of graph description with that of argumentative essays on L2 writing in a testing condition, (ii) a small number of studies that examine two models of attention by examining task complexity in different task type, and (iii) the need to have more research on the IELTS writing component with a detailed diagnostic tool to predict its candidates’ language problems.

1.2. Aim and scope

The aim of the present study is to examine the effect of task type as one important aspect of task complexity on L2 writers’ performance in IELTS academic writing. To achieve this aim, I compare L2 writing samples on task 1 and task 2 of the IELTS Academic writing subtest. Data for the study was collected through an IELTS simulation test at a language centre of a large research university in Hanoi, Vietnam. The study evaluates L2 writing by using a range of discourse-analytic measures focusing on the accuracy and complexity. It does not analyse the writings in terms of arguments, organization and cohesion, which is the focus of another study.

1.3. Underpinning theories of research on tasks in second language acquisition (SLA)

1.3.1. Task complexity and attentional resources

Extensive research into the effect of task demands on SLA has been strongly influenced by two models of attention, namely Skehan and Foster’s Limited Capacity Hypothesis [5] and Robinson’s Cognition Hypothesis [6]. Both models emphasize the significant role of attention and L2 learners’ use of their attentional resources in completing tasks. However, the two models differ in their hypotheses about the effect of increasing task complexity on language production.

Skehan and Foster [5] adopt information processing perspectives on the nature of language learning. They hypothesize that language learners’ limited attentional capacities influence pervasively their focus during
meaning-oriented communication. In other words, language learners cannot attend to everything equally at the same time, and attending to one aspect may mean the neglect of others. The three areas competing for attention are complexity, accuracy and fluency. According to Skehan and Foster [5], actual performance largely depends on learners’ priority, task characteristics and task conditions. In regards to the relationship between task content and performance, Skehan and Foster [7] argue that when a cognitively complex task requires significant focus on content, less attention would be allocated to linguistic form. Consequently, the complexity and accuracy of the linguistic output will decrease. They also claim that when resources are available in performing cognitively demanding tasks, learners only could prioritise either accuracy or complexity, but not both.

In contrast to Skehan and Foster’s Limited Attentional Capacity Model, Robinson’s Cognition Hypothesis claims that learners’ attentional resources are multiple and non-competing [6], [8], [9]. Under the influence of both information processing and interactional perspectives of L2 task effects, the Cognition Hypothesis proposes that cognitively more demanding tasks might push learners to produce more accurate and more complex language [10]. These tasks are thought to promote more linguistic awareness and consequently trigger greater linguistic complexity and higher accuracy to meet greater functional demands [11].

1.3.2. Dimensions and variables of task complexity

Both Limited Attentional Capacity Model and the Cognition Hypothesis distinguish a number of dimensions and variables of task complexity that influence L2 learners’ performance.

In the Limited Attentional Capacity Model, Skehan and Foster [7] differentiates between three main aspects of task complexity: communicative stress, code complexity and cognitive complexity. Communicative stress is concerned with performance condition. Code complexity refers to the linguistic demands of the task. Cognitive complexity is related to task content and the structuring of task material. With regards to cognitive complexity, he states that familiarity of information (i.e. the extent to which the task allows learners to draw on their own available content schema) has no impact on accuracy and complexity but improves fluency. In contrast, when the task requires learners to interact with each other, there is a gain in accuracy and complexity at the expense of fluency [12].

Robinson [8] distinguishes task complexity, task difficulty and task conditions. Task complexity (cognitive factors) refers to the “attentional, memory, reasoning and other information processing demands imposed by the structure of the task on the language learner” [8:29]. He also suggests that task complexity can be manipulated along resource-directing and resource-depleting dimensions. The resource-directing dimensions can increase or decrease the functional demands on the language user. Tasks which require learners to describe and differentiate few elements and relationship (+few elements) or/and describe events happening now in a shared context (+here-and-now) are said to consume less attentional resources than tasks which involve different elements and relationship (-few elements), entail displaced references (-here-and-now) and need reasons to support statements (-no reasoning demands) [10].
second task design factors are resource-depleting dimensions such as +/- planning time (with or without planning time), +/-single task (single task or multiple tasks), +/- prior knowledge (with or without prior knowledge). According to Robinson, manipulating task complexity along those dimensions can result in “a depletion in attentional and memory resources”, reducing fluency, accuracy and complexity on the more complex tasks [8: 35]. Unlike task complexity, task difficulty (learner factors) are the differences in resources learners draw on in responding to task demands (e.g. gender, familiarity), and task conditions are participant factors such as one-way or two-way communication and communicative goals.

The two models of attention above have prompted a number of task-based studies on SLA Studies related to the impact of task complexity on L2 learners’ performance will now be reviewed.

1.4. Current debates

1.4.1. The effects of task complexity on L2 written performance

The body of literature on the effects of task complexity on L2 written performance is mainly based on Robinson’s Cognition Hypothesis and Skehan and Foster’s Limited Capacity Model. However, these task-based studies differ in their support for one of the two models.

The first group of studies seems to show more support for Robinson’s multi-resources view of attention. Ishikawa [13] manipulated [+here and now] dimensions of Japanese EFL learners’ narrative writing. The main finding was that more complex tasks pushed learners to produce higher accuracy and syntactic complexity, but no improvement was seen in linguistic complexity. Kuiken and Vedder [11] concerned the effect of task complexity on linguistic performance by looking at the letter writing of 75 Dutch learners of French and 84 Dutch learners of Italian. Two writing tasks were assigned in which cognitive complexity was manipulated by giving six requirements in the complex and three in the non-complex condition. They discovered that the more complex letters (with six requests) prompted higher accuracy but not higher linguistic complexity. Ong and Zhang [14] manipulated task complexity along both resource-depleting dimensions (planning time, the provision of ideas and structure) and resource-directing dimensions (draft availability). Their study explored the effects of task complexity on fluency and lexical complexity of 108 EFL students’ argumentative writing. Their findings lent more support to Robinson’s Cognition Hypothesis than Skehan and Foster’s Limited Attentional Capacity Hypothesis. No trade-offs as suggested by Skehan and Foster were observed; increased lexical complexity and fluency did not compete. When task complexity was increased along planning time continuum, higher fluency and greater lexical complexity were seen. Increasing task complexity through the provision of ideas and macro-structure promoted significantly lexical complexity but no effect on fluency. The manipulation of task complexity along the provision of draft produced no significant differences in fluency and lexical sophistication.

The second group of studies is more in line with Skehan and Foster’s predictions. Ellis and Yuan [15] reported findings on the effects of three types of planning (no planning, unpressured online-planning, pre-task planning) on 42 Chinese learners’ written narratives based on a series of pictures. Pre-task planning was
found to have remarkably positive influence on fluency, syntactic complexity and little influence on accuracy; meanwhile writers in no planning condition were faced with negative consequences in fluency, complexity and accuracy compared to planning group. The researchers explained that planning helped learners in setting goals, organizing the text and preparing the propositional content, thus reducing pressure on the central executive working memory and enhancing confidence during task performance. Ellis and Yuan’s findings pointed into the direction of Skehan and Foster’s Model.

1.4.2. The effect of task type on L2 performance

There have been a number of studies on the intervening effect of task type as one important aspect of task complexity. Most of them support the Limited Attentional Capacity Hypothesis.

Mohsen, Mansoor & Abbas Eslami define writing genre to be “the name given to the required written product as outlined in the task rubric” [16: 206]. Ong & Zhang claim that the requirement of a particular genre determines test-takers’ linguistic choice for their answers [14]. Task type are also said to be crucial in determining “if writers are able to automatize certain features of writing tasks or deal with additional cognitive load to process those aspects” [15: 170]. For example, according to Foster and Skehan [17], argumentative writing is more complex than descriptive writing in that it requires writers to generate reasoning meanwhile descriptive writing has a clear inherent structure, requiring writers to describe individual actions or characters [16].

Most of the studies on genre writing converge on that argumentative writing is the most cognitively demanding writing task and that Skehan and Foster’s Limited Attentional Capacity Hypothesis gives a better explanation of L2 writers’ performance.

Way, Joiner and Seaman [18] compared 937 writing samples of 330 novice learners of French on three tasks (descriptive, narrative, expository). They assessed the quality, fluency, syntactic complexity and grammar accuracy of the writing. Results indicated that the descriptive writing which involved the description of participants’ family, class, pastimes was the easiest, and the expository writing which required students to write a letter about American teenagers and their role in society and family, their views on education and politics, their goals for future was the most difficult. Concerning the main focus of the present study, the findings also seem to support Skehan and Foster’s model by stating that descriptive task was the longest and of the highest quality. In contrast, expository essays were the shortest and had the lowest score.

Mohsen, Mansoor and Abbas Eslami [16] investigated the role of task type in the writing performance of 168 Iranian undergraduate English majors. The two task types were an argumentative writing task and an instruction writing task. Findings showed that the instruction essays, which were considered to have lower cognitive and linguistic demands than the argumentative essays, elicited higher fluency and greater accuracy. In contrast, participants in the argumentative essay group performed significantly better in terms of complexity.

Lu [19] recently reported a large scale corpus study which used 14 complexity measures as objective indices of college-level ESL learners’ language development. The study looked at 3678 essays by Chinese students; the
linguistic complexity was assessed in the length of production, sentence complexity, subordination, coordination and particular structure. With respect to the effect of genre on the participants’ writing, results showed that the syntactic complexity of argumentative essays was higher than narratives.

**Genre writing research in IELTS testing conditions**

The aforementioned studies were carried out mostly in a classroom context, and there is little investigation into the impact of task type in a writing test condition, especially IELTS writing. O'Loughlin and Wigglesworth [2] noted that the writing assessment area needed a great deal more attention to critical intervening factors, of which writing task is one. Among few attempts at exploring the impact of writing task type in IELTS context, most of the studies focus on either task 1 or task 2, leaving the comparison between two tasks an underresearched area.

O'Loughlin & Wigglesworth [2] examined how the task difficulty in IELTS Academic Writing Task 1 was influenced by the amount of information provided and the presentation of information to the candidates. Four tasks differing in the amount of information were assigned to 210 students in Melbourne or Sydney enrolled in the course English for Academic Purposes. The analysis of written texts revealed that the tasks giving less information, i.e. they are cognitively easier to process, generated more complex language. This partially supports the Limited Attentional Capacity Hypothesis.

In one rare effort to look at both IELTS writing tasks, Banerjee, Franceschina, and Smith [20] set to see how competence levels, as shown in IELTS band scores, were corresponding to L2 developmental stages. These researchers tried to document typical linguistic features shown in Task 1 and Task 2 written texts of 275 Chinese and Spanish test takers. They looked at the defining characteristics of bands 3-8 in terms of cohesive device use, vocabulary richness, syntactic complexity and grammar accuracy. The effects of L1 and writing task type were also examined. These authors claimed that task type had significant effects on candidates’ writing performance. The impacts of two tasks on vocabulary richness were different. They found that task 1 induced higher lexical density, and task 2 had higher lexical variation as measured by type-token ratio. In their findings, task 2 scripts also tended to elicit fewer high-frequency words. Although these researchers also examined the effect of task type by comparing L2 writers’ performance in two IELTS writing tasks, they did not approach the task differences from task complexity perspective. Their findings are consequently descriptive of IELTS candidates’ typical writing features in each task.

**1.5. Summary of gaps in the literature**

A brief review of the literature in the research area suggests that to date, few researchers have investigated the different effects of task type as a crucial factor of task complexity on L2 writing in IELTS Academic Writing subtest across three areas of fluency, accuracy and complexity. Therefore, the present study has been carried out in an effort to bridge this research gap.

**1.6. Research questions**

The following research questions have been formulated to examine the influence of task
type as a factor of task complexity on complexity and accuracy in IELTS Academic writing:

1. Does task type influence the accuracy of EFL learners’ written products in a simulated IELTS test?

2. Does task type influence the complexity of EFL learners’ written products in a simulated IELTS test?

(EFL learners are learners of English as a foreign language. They are different from ESL learners – learners of English as a second language in that ESL learners will use English as the second official language in their country while EFL learners will use English as a foreign language.)

2. Method

2.1. Design

The study is a single-factor, repeated-measures design which aims to explore the effects of two task types i.e. graph description and argumentative essay on learners’ writing performance. This was congruent with the focus of the study: comparing how two different tasks influence the same group of participants. Repeated-measures design also afforded the opportunity to work with a limited number of participants within the scope of a small-scale minor thesis. This approach has been adopted in a number of similar task-based studies, e.g. [16], [11], [9], [2].

2.2. Instruments

The participants were assigned two IELTS Academic Writing tasks from an IELTS practice tests book as these tasks are stated to represent the tasks in actual IELTS examinations [21]. These writing tasks were included in the participants’ second progress test within an IELTS preparation course. Task 1 required the participants to summarize the information and make comparisons where relevant; the information was presented in a bar graph about gender differences in different levels of post-school qualification in Australia in 1999. This task was considered a simple type of task 1 in IELTS Academic Writing as it included fewer than 16 pieces of information following O’Loughlin and Wigglesworth’s classification (see Appendix A) [2: 92]. The participants were asked to write at least 150 words in 20 minutes.

In Task 2, the participants were asked to discuss both sides of the following statement “The Internet is an excellent means of communication”, but “it may not be the best place to find information”. They were required to give reasons and relevant examples in their responses (see Appendix A). This topic was of general interest and did not require expert knowledge to avoid giving certain participants an advantage. Research evidence shows that the task related to candidates’ discipline would boost their performance [22], [23], [24]. Task 2 essay had to consist of at least 250 words, and there was a time limit of 40 minutes.

Different levels of task complexity of two IELTS writing tasks

Although all previous studies agree that the argumentative essay is the most demanding writing task, there have been few studies that investigate the differences in task demands between graph description and argumentative essay in terms of task complexity in IELTS tests. Thus, I use Skehan (1996)’s criteria for task grading, i.e. code complexity and cognitive complexity to argue that task 1 – the graph
description has lower cognitive and linguistic demands than task 2 – the argumentative essay. This would serve as the basis for my analysis of the effects of different complexity levels of different task type on L2 writing performance in light of the Limited Attentional Capacity Hypothesis and Cognition Hypothesis.

Skehan’s [5] first criterion, code complexity, includes vocabulary load and variety. Regarding this aspect, the graph description task would require a more limited range of vocabulary than the argumentative essay. Yu, Rea-Dickins and Kiely [25] claimed that learners were trained to describe concrete contrasts in data presented in bar graphs by using language of comparison, e.g. higher, lower, greater than, less than. Skehan’s second criterion, cognitive complexity, covers two areas: cognitive familiarity and cognitive processing. With respect to the first area, cognitive familiarity, the graph description task would be more familiar to the participants of the present study than the argumentative task. The structure of the graph description task was more predictable as IELTS candidates were aware of the principles of “cognitive naturalness” when people produced bars to depict comparisons [27]. Moreover, it would be easier to familiarize intended potential test-takers with the discourse genre of task 1 because task 1 only covers several types of visual input such as graphs, charts, diagrams as compared to limitless topics of task 2.

Regarding the second area, cognitive processing, the graph description task involved a smaller amount of online-computation than the argumentative essay task for the following three reasons. First, the graph description task required less reasoning; the participants were only asked to summarize main features and compare where possible. The argumentative essay, on the other hand, involved complicated reasoning to establish causality and justification of beliefs which was claimed to be cognitively more challenging than tasks without those demands [8]. Second, in terms of input material, task 1 provided the participants with visual aids and exact figures that they could draw on to organize their description. However, when completing task 2, the participants had to draw on their own resources to come up with ideas and supportive reasons to defend their positions. Finally, the information given in task 1 was more interconnected and had a clearer inherent structure than task 2, which tended to have an arbitrary organization of the content.

An investigation into the rating criteria of two tasks also suggests that a less amount of cognitive process is required in task 1. Both tasks are assessed on lexical resource, grammatical range and accuracy criteria. Task 1 scripts are assessed according to task fulfilment, coherence and cohesion; task 2 scripts are assessed according to task response (making arguments) [1]. Test-takers can be considered to have fulfilled task 1 by describing and comparing the main information; meanwhile, task 2 requires them to do a more challenging task of making arguments and supporting their positions. Robinson [8] asserts that the tasks that require learners to give reasons to establish causality and justification of beliefs are more complex than the task without these demands. Uysal also criticized that the criteria “coherence and cohesion” of task 1 causes “rigidity and too much emphasis on paragraphing” [1: 371].

Based on the above-discussed criteria, I argue that task 1 – the graph description has lower cognitive and linguistic demands than task 2 - the argumentative essay.
2.3. Participants

The study involved the participation of 30 EFL learners at the aforementioned language centre. There were two sampling criteria: (i) they must be non-native speakers of English, and (ii) they must have no experience of taking the actual IELTS test but are planning to take the IELTS test in the near future. The assumption for the first criteria was that all of the participants speak Vietnamese as their first language in a non-English speaking context. The purpose of the second criteria was to control the effect of different amounts of IELTS training that the participants may have received before joining the study, and the researcher anticipated that these participants who were planning to take IELTS would be more engaged with this research project. To this end, 30 participants were sampled from the IELTS preparation class with the target band score of 5.0-6.0. This was the lowest-level IELTS preparation course at the centre, which included learners with virtually no previous IELTS training or experience. All of the participants were students at the same university; their majors were Law, Technology, Economics and Science. As these participants were placed in the same class based on the scores of their placement test, they were supposed to have approximately the same proficiency level. Each chosen participant was referred to by a number to ensure their anonymity.

3. Analyses and results

3.1. Analytical procedures

As claimed by Storch [4], the IELTS analytical assessment scale does not give much information for predicting candidates’ language problems. She also confirms that analytical scores are often collapsed to yield a single score, losing diagnostic value [4]. It is difficult to identify specific lexicogrammatical features that distinguish different band scores [3]. The unsuitability of the IELTS rating scale for diagnostic purposes motivated the present study to use the discourse measures of complexity and accuracy which are believed to be more specific indicators of learners’ language proficiency level [19]. As defined by Skehand and Foster [5], complexity refers to size, richness and diversity of linguistic resources. It reflects “speakers’ preparedness to take risks and restructure their interlanguages” [5: 2]. Accuracy means the ability to produce the language appropriately in relation to the rule system of the target language.

For the use of the chosen discourse measures, all writings were coded for T-units, clauses and errors. A T-unit is defined as “one main clause plus whatever subordinate clauses happen to be attached or embedded within it” [4: 107]. The participants’ scripts were also coded for independent and dependent clauses. An independent clause is one clause that can stand on its own, and a dependent clause is defined as one that augments an independent clause with additional information but cannot stand alone [26]. There has been disagreement among researchers about how to code for a dependent clause. In this study, dependent clauses contained a finite or non-finite verb and at least one clause element such as subject, object, complement or adverbial [16]. The following examples were taken from the data. The first example contains one T-unit which is composed of two clauses (separated by a slash as shown): an independent clause and a finite dependent clause beginning with “that”. The second one comprises one T-unit which
contains an independent clause separated from a non-finite dependent clause beginning with “achieved”:

*It is undoubtedly true/ that the Internet plays an important role in our modern life.*

The bar chart illustrates the proportion of 5 post-school qualifications/ achieved by males and females in Australia in 1999.

To assess accuracy, the study used the proportion of error-free t-units to t-units (EFT/T), error-free clauses to clauses (EFC/C) and the total number of errors per total number of words (E/W). The last measure was used to account for the T-units containing multiple errors [4]. The participants’ writings was coded for errors using Chandler’s [27] error taxonomy which categorize errors into syntax errors (e.g. word order, incomplete sentences), morphology errors (verb tense, subject-verb agreement, use of articles) and lexis errors (word choice). Errors in spelling, punctuation and capitalization were not counted to avoid overestimation of errors due to unclear handwriting [4]. The following errors from the data illustrated Chandler’s categorization.

Grammatical complexity was measured by the ratio of dependent clauses per clause (DC/C) as the level of embedding and subordination is believed to demonstrate syntactic sophistication [4]. Following [28] and [4], the measure of lexical variation was a type/token ratio (i.e. the number of different lexical words over the total number of lexical words per one script) and the proportion of academic words to total words. For the analysis of lexical variation, I used the corpus linguistic program Compleat Lexical Tutor v.6.2. This program has been empirically validated in peer-reviewed papers [29], and Diniz [30] confirmed that the unique features of this corpus program could help researchers analyse the lexical complexity of different texts. All the written scripts were inputted into the program which would, in turn, give the statistics about type/token ratios and the percentage of words from the writings appearing in the academic word list (AWL). AWL developed by Coxhead [31] comprises 570 headwords and over 3000 words in total, representing about 10% of the most commonly used academic words.

Once the data had been collected in the form of number of words per T-unit, proportion of error-free t-units to t-units (EFT/T), error-free clauses to clauses (EFC/C), the total number of errors per total number of words (E/W) (measures of accuracy), ratio of dependent clauses per clause (DC/C) (measure of grammatical complexity), type-token ratio and percentage of academic words (measures of lexical complexity), means were calculated for each aspect of each task. In the next step, given the fact that the same group of participants performed two different writing tasks, Paired sample t-tests were run to find the differences between Task 1 and Task 2 with regards accuracy and complexity respectively. The t-test results were analyzed in relation to the means to identify the task with higher performance. The alpha for achieving statistical significance was set at 0.05 [11], [16]. The effect sizes were defined as “small, \( d = 0.2\)”, “medium, \( d = 0.5\)”, and “large, \( d = 0.8\)” [32: 25].

To ensure inter-coder reliability in coding, randomly chosen four writings of Task 1 and four writings of Task 2, representing over 13% of the total sample of 60 writings, were coded by a second researcher. As advised by Polio (1997) [33], specific guidelines were created defining and exemplifying T-units, clauses, and errors. To check for intra-coder reliability, a random sample of 8 writings (four of each task)
were coded a week after the first coding by the researcher. Simple percentage agreement was used to show reliability scores. Inter-rater reliability for T-unit, clause analyses and error counts were 95%, 93% and 88% respectively. Intra-rater reliability for T-units and clause identification were 97% and 96% respectively, while the rate was 89% for error analysis.

3.2. Results of the study

3.2.1. Research question 1: Does task type influence the accuracy of EFL learners’ written products?

Three variables were assessed to measure the accuracy of the participants’ language use in two tasks. Table 3 shows that the mean score of E/W of task 1 (M = 0.04) was lower than that of task 2 (M = 0.07), indicating that on average, the learners made fewer mistakes in task 1 than task 2. While the ratio of error free T-units to T-units (EFT/T) in task 1 was higher than that of task 2 (M = 0.44 and 0.30 respectively), the proportion of error-free clauses to clauses (EFC/C) were roughly the same for two tasks.

The results of paired sample t-tests revealed that there is a statistically significant difference for complexity in two measures (E/W: t(29) = -5.237, p < 0.001; EFT/T: t(29) = 4.013, p = 0.001). The effect size of E/W is rather small at d = -0.21, and the effect size of EFT/T was medium (d = 0.76). Regarding the third measure of EFC/C, the difference between two tasks was not significant (EFC: t(29) = -0.471, p = 0.642).

The above analysis suggests that the participants performed significantly better in terms of accuracy in the graph description task than in the argumentative essay task. In other words, the language used in the graph description task was more accurate than in the argumentative essay task.

Table 1. Results for accuracy (n=30)

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>SD</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>E/W</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Task 1</td>
<td>0.04</td>
<td>0.2</td>
<td>0.01</td>
<td>0.10</td>
</tr>
<tr>
<td>Task 2</td>
<td>0.07</td>
<td>0.04</td>
<td>0.03</td>
<td>0.17</td>
</tr>
<tr>
<td>EFT/T</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Task 1</td>
<td>0.44</td>
<td>0.18</td>
<td>0.13</td>
<td>0.88</td>
</tr>
<tr>
<td>Task 2</td>
<td>0.30</td>
<td>0.19</td>
<td>0.00</td>
<td>0.69</td>
</tr>
<tr>
<td>EFC/C</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Task 1</td>
<td>0.57</td>
<td>0.17</td>
<td>0.28</td>
<td>0.89</td>
</tr>
<tr>
<td>Task 2</td>
<td>0.58</td>
<td>0.14</td>
<td>0.30</td>
<td>0.85</td>
</tr>
</tbody>
</table>

3.2.2. Research question 2: Does task type influence the complexity of EFL learners’ written products?

To measure the grammatical complexity, the proportion of dependent clauses of all clauses (DC/C) was used. As Table 3 shows, the mean score of DC/C was higher in task 2. The paired sample t-test confirmed that this difference was statistically significant (t(29) = 4.681, p < 0.001), and the difference was broad (Cohen’s d = 1.37).

Lexical complexity or variation was measured by the type-token ratio and the percentage of words used in the participants’ scripts that appeared on the Academic Word List (AWL). The results in table 3 show that in
writing task 1, the participants produced a lower proportion of different lexical words of all lexical words (type-token ratio). The paired sample t-test proved that this difference between two tasks in terms of type-token ratio was significant (t(29) = 4.88; p < 0.001), and the difference was of small size (Cohen’s d = 0.13). With regard to the second measure, the descriptive results show that task 1 was written with a slightly higher percentage of academic words (6.51%) than task 2 (6.03%). However, the paired sample t-test indicated that this difference was not statistically significant (t(29) = 0.99; p = 0.34).

From the above analysis, the argumentative essay was produced with more complex language than the graph description task as illustrated in the higher level of grammatical subordination (DC/C) and the greater variety of lexical words (type-token ratio).

### Table 2. Results for complexity (n=30)

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>SD</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>DC/C</td>
<td>0.41</td>
<td>0.12</td>
<td>0.20</td>
<td>0.58</td>
</tr>
<tr>
<td>Task 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Task 2</td>
<td>0.55</td>
<td>0.08</td>
<td>0.38</td>
<td>0.63</td>
</tr>
<tr>
<td>TYPE/TOKEN</td>
<td>0.49</td>
<td>0.05</td>
<td>0.41</td>
<td>0.58</td>
</tr>
<tr>
<td>Task 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Task 2</td>
<td>0.55</td>
<td>0.04</td>
<td>0.48</td>
<td>0.64</td>
</tr>
<tr>
<td>AW/W</td>
<td>6.51</td>
<td>1.79</td>
<td>3.49</td>
<td>10.81</td>
</tr>
<tr>
<td>Task 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Task 2</td>
<td>6.03</td>
<td>1.83</td>
<td>3.56</td>
<td>12.74</td>
</tr>
</tbody>
</table>

### 3.3. Summary of key outcomes

In summary, the quantitative analysis of the language in the two tasks suggested that the graph description task (task 1) elicited a significantly better performance in terms of accuracy than the argumentative essay task (task 2). Meanwhile, the language used in the argumentative essay task was more complex in terms of grammatical subordination and lexical variation than the graph description task.

### 4. Discussion

Given the comparability in the first language, general language proficiency and writing ability, the participants’ writings in two tasks were significantly different in terms of accuracy and complexity. These findings were consistent with other studies in the field and consolidated the fact that task type had a tremendous impact on writers’ performance. One possible explanation was suggested by Mickan and Slater (2003) [34] that the specification of a particular type of task determined test-takers’ choice of linguistic elements for their answers.

#### 4.1. Research question 1: Does task type influence the accuracy of EFL learners’ written products?

The first research question addressed the effects of task type on the accuracy of learners’ written output. The evaluation of accuracy levels was based on the examination of three measures: the proportion of error-free t-units to t-units (EFT/T), error-free clauses to clauses (EFC/C) and the total number of errors per total number of words (E/W). Results show that our participants produced more accurate language in the less demanding task, i.e. the graph description, with respect to E/W and EFT, and no significant difference was seen between two tasks in terms of EFC/C.

Regarding the lower E/W and higher EFT/T of task 1, the significant results emerged because the bar graph description task requires simpler grammatical structures which mostly involve comparative structures, whereas the argumentative essay task entails a wider range of more complex grammatical structures to state, exemplify and justify a position on an issue. As my participants achieved more
effective control over the simpler grammatical structures, their language produced in the bar graph description was more accurate than in the argumentative essay. Another possible explanation was that the chance of committing errors would increase with the length and complexity of the required output [13]. This seems particularly true for the argumentative essay that exerted more word limits than the graph description task. This higher accuracy level in a low demanding task is consistent with most previous studies [15], [17], [18]. The explanation for this effect given by those researchers was that less challenging tasks enable participants to allocate more attentional resources toward monitoring accuracy than do the high demanding tasks which may direct more of their attention to conveying meaning first.

Regarding the non-significant result of the ratio of error-free clauses to all clauses (EFC/C), one point needs to be mentioned: a T-unit may contain multiple clauses. At the clause level, the accuracy between two tasks may be roughly the same. However, as a T-unit in the argumentative essay comprised more clauses, this might increase the number of T-units with errors. The post hoc paired sample t-test showed that the average number of clauses per T-unit in the graph description (1.77) was significantly lower than that in the essay (2.25) with t(29) = -5.04, p < 0.001.

4.2. Research question 2: Does task type influence the complexity of EFL learners’ written products?

The second research question dealt with what effect task type exerted on learners’ writing performance. The structural complexity was evaluated by the proportion of dependent clauses of all clauses (DC/C), and the lexical variation was measured by the type-token ratio and the percentage of words used in the participants’ scripts that appeared on the Academic Word List (AWL). From the analysis, our participants produced more complex language in the high cognitively demanding task (the argumentative essay) than the graph description task as illustrated in the higher level of grammatical subordination (DC/C) and the greater variety of lexical words (type-token ratio).

The conclusions about the argumentative essay’s higher lexical and syntactic complexity were consonant with those reported by Mehnert [35], Robinson [6], [8], Ong and Zhang [14], Banerjee, Franceschina and Smith [20], and Lu [19]. However, while most of the previous studies reported gains in lexical complexity based on lexical sophistication measures, the present study conducted lexical range measures. The findings of the current study may be accounted for by the task demands. The requirement of the argumentative essay for elaborated content and justified opinions induced the participants to produce more complex messages that entailed the use of more advanced vocabulary and structures to transfer. In contrast, the graph description task only called for describing and comparing information, thus directing mental effort at cohering information presented in the graph and not involving diverse lexical items. This finding partially supports Robinson’s Cognition Hypothesis.

5. Implication

The findings of the current study seemed to lend stronger support for Skehan and Foster’s Limited Attentional Capacity Hypothesis than
Robinson’ Cognition Hypothesis by confirming the trade-off phenomenon between accuracy and complexity when L2 learners performed a more cognitively demanding task. The overall results showed that an increase in task cognitive demands led learners to allocate their limited attentional resources to the complexity of language with which accuracy could not keep pace, i.e. “accuracy last approach” [7: 189]. It is possible to conclude that there is indeed a relation between task complexity and linguistic performance. Increasing task complexity may lead learners to produce a text which more syntactically complex and lexically varied but not necessarily more accurate.

The findings of the current study are believed to yield some implications. First, while most of the previous studies examined the two models of attention by manipulating task complexity along planning, here-and-now variables, task prompts and draft availability, my study chose to investigate task complexity embedded in different task type. Second, given the limited research on comparing the effect of diagram interpretation with argumentative essays on L2 writers’ performance in a testing condition, the present study also contributed exploratory findings to the body of knowledge on L2 writing. Third, it was also believed that the study addressed the need of more investigation into writing components of IELTS examination [1]. Finally, the use of discourse measurement of accuracy and complexity in this study might give a deeper insight into IELTS candidates’ language problems related to genre writing. Thus, some pedagogical implications need consideration to help curriculum developers and teachers of IELTS preparation courses to prepare their students better for this high-stakes test. To promote accuracy and complexity of both low and high demanding tasks, IELTS trainers should instruct learners how to plan the content and vocabulary of their writing. Given the assumption that learners may fall behind on at least one area, i.e. accuracy or complexity, due to their limited attentional capacity, teachers should manipulate task features to channel learners’ attention to the area in which they fail. For the graph description task (lower demanding task), learners’ language complexity might be improved by instructing them how to plan the vocabulary of their description as well as providing them with more synonymous vocabulary and structures to diversify their expression. For high demanding tasks, trainers should provide learners with more information about the grammatical structures and expressions relevant to the assigned tasks. It is advisable that IELTS curriculum developers integrate reading and listening materials as input for writing topics of argumentative essays. More exposure to the target like language of similar topics might help L2 writers increase the fluency and language complexity of their essays.

6. Limitations and agenda for further research

This small-scale study is largely exploratory, so its results should be cautiously interpreted. Firstly, the study looked at only Vietnamese students, so the sample did not represent the population of typical IELTS candidates with various first languages. The sample of the present study had a similar proficiency level; meanwhile, a wider range of language abilities was seen in real-life IELTS test-takers. A future study should include participants with different proficiency levels and various first language background. Due to
the scope of the study, some important aspects of
demanding task (task 1 – graph description) was more effective in
promoting accuracy of learners’ written output;
meanwhile, higher demanding task (task 2 –
argumentative essay) could induce more
syntactically complex and lexically varied
language. These overall results were more
compatible with the Limited Attentional
Capacity than the Cognitive Hypothesis in that
a less challenging task produced a more
accurate linguistic performance. Moreover, the
findings also proved that accuracy and
complexity compete with each other when
attentional resources need to be allocated to
perform complex tasks.

7. Conclusion

The purpose of the present study was to
provide an insight into the effect of task type
with different task complexity levels on L2
writers’ performance in task 1 and task 2 in
IELTS Academic writing subtest. To achieve
this aim, I looked at 60 writing samples (30 of
each task) of 30 Vietnamese students in an
IELTS simulation test at a language centre of a
large Vietnamese university. The study was
limited to an evaluation of accuracy and
complexity of the writing samples. The results
showed that low demanding task (task 1 –
graph description) was more effective in
promoting accuracy of learners’ written output;
meanwhile, higher demanding task (task 2 –
argumentative essay) could induce more
syntactically complex and lexically varied
language. These overall results were more
compatible with the Limited Attentional
Capacity than the Cognitive Hypothesis in that
a less challenging task produced a more
accurate linguistic performance. Moreover, the
findings also proved that accuracy and
complexity compete with each other when
attentional resources need to be allocated to
perform complex tasks.

References

design in IELTS Academic Writing Task 1: The
effect of quantity and manner of presentation of
information on candidate writing IELTS:
International English language testing system
(Research Reports 2003 B2 - IELTS: International
English language testing system (Research
Reports 2003 (pp. 89-130). Canberra: Australian
Capital Territory: IELTS.
the assessment of academic writing International
English Language Testing System Research
Reports B2 - International English Language
Testing System Research Reports (pp. 59-88).
Canberra: IELTS Australia Pty.
second language (L2) medium university on the
development of L2 writing. [Article]. Journal of
Second Language Writing, 18, 103-118. doi:
10.1016/j.jslw.2009.02.003.
Task Structure and Processing Conditions on
Narrative Retellings. Language Learning, 49(1),
93-120.


Anh hướng của thể loại viết tối độ chính xác và phục tập của ngôn ngữ trong bài viết học thuật của đề thi IELTS

Nguyễn Thùy Lan

Khoa Sự phẩm tiếng Anh, Trường Đại học Ngoại ngữ, ĐHQGHN,
Phạm Văn Đồng, Cầu Giấy, Hà Nội, Việt Nam

Tóm tắt: IELTS là một trong những bài thi tiêu chuẩn quốc tế phổ biến nhất dùng để kiểm tra năng lực tiếng Anh. Hai nhiệm vụ viết của bài thi rất khác nhau về yêu cầu từ ngữ và ngôn ngữ. Tuy nhiên, cho đến nay, chưa có nghiên cứu nào so sánh ảnh hưởng của những thể loại viết khác nhau đối với bài viết của thí sinh dự thi. Trong lĩnh vực nghiên cứu ngôn ngữ thứ hai có hai lý thuyết trái ngược nhau về yêu cầu của nhiệm vụ ngôn ngữ: Mô hình khả năng tập trung hạn chế (the Limited Attentional Capacity Model) - mô hình này cho rằng những nhiệm vụ phức tạp sẽ làm giảm chất lượng ngôn ngữ của người viết - và Giả thuyết Nhận thức (the Cognition Hypothesis) – giả thuyết cho rằng nhiệm vụ phức tạp hơn làm tăng chất lượng ngôn ngữ của người viết. Bài viết dưới đây đánh giá ảnh hưởng của thể loại viết – một yếu tố tạo nên độ phức tạp của nhiệm vụ viết (task type) đối với chất lượng ngôn ngữ của người viết trong bối cảnh một bài thi. Nghiên cứu được tiến hành là nghiên cứu một yếu tố, tài liệu được lấy trong một số bài viết của 30 người học tiếng Anh để chỉ việc viết trong bài thi IELTS. Bài viết của đối tượng nghiên cứu được phân tích thông qua việc sử dụng một loạt những công cụ đo độ chính xác và phục tập của diễn ngôn. Kết quả cho thấy nhiệm vụ viết dễ hơn (bài 1 – miêu tả bằng biểu) sẽ cho ra ngôn ngữ chính xác hơn. Trong khi đó, nhiệm vụ viết phức tạp hơn (bài 2 – viết luận) cho ra ngôn ngữ với ngữ pháp phức tạp hơn và từ vựng đa dạng hơn. Bài viết này đồng góp những kết quả bước đầu trong lĩnh vực nghiên cứu ảnh hưởng của độ phức tạp của nhiệm vụ ngôn ngữ trong kỹ năng viết bằng ngoại ngữ. Việc sử dụng công cụ đo độ chính xác và phục tập của diễn ngôn đã hỗ trợ những vấn đề ngôn ngữ của thí sinh IELTS. Với giả định là người học có thể biết yêu cầu ở một phương diện: độ chính xác hay phục tập của ngôn ngữ do sự hạn chế trong việc chú ý tới cả hai phương diện khi viết, giáo viên nên có những hoạt động dạy học phù hợp để giúp người học cải thiện điểm yếu của họ.

Từ khóa: Kiểm tra ngôn ngữ, đánh giá kỹ năng Viết, IELTS, nhiệm vụ ngôn ngữ, thể loại viết, công cụ đo diễn ngôn, độ chính xác, độ phức tạp (của ngôn ngữ).
Appendix A: IELTS Writing Tasks

PROGRESS WRITING TEST 03

You should spend about 20 minutes on this task.

The chart below shows the different levels of post-school qualifications in Australia and the proportion of men and women who held them in 1999.

Summarise the information by selecting and reporting the main features, and make comparisons where relevant.

Write at least 150 words.

<table>
<thead>
<tr>
<th>Post-school qualifications in Australia according to gender 1999</th>
</tr>
</thead>
<tbody>
<tr>
<td>Skilled vocational diploma</td>
</tr>
<tr>
<td>Undergraduate diploma</td>
</tr>
<tr>
<td>Bachelor’s degree</td>
</tr>
<tr>
<td>Postgraduate diploma</td>
</tr>
<tr>
<td>Master’s degree</td>
</tr>
</tbody>
</table>

0% 10% 20% 30% 40% 50% 60% 70% 80% 90% 100

Females  Males

WRITING TASK 2

You should spend about 40 minutes on this task.

Write about the following topic:

It is generally believed that the Internet is an excellent means of communication but some people suggest that it may not be the best place to find information.

Discuss both these views and give your own opinion.

Give reasons for your answer and include any relevant examples from your own knowledge or experience.
Appendix B: Error Taxonomy

Paragraph: Not indenting with each new speaker in dialogue
Sentence structure: I suggest you to go with me.
Meaning unclear: I planned how to divide my cabins in a good way.
Awkward: He is the brother of her.
Idiom: She asked me to lay the table.
Fragment: Because I did not want to go.
Run-On: She told me about it, I answered her question, she left.
Punctuation: I wanted her to come, because she knew the way.
Capitalization: She said Love was very important.
Delete: She told to me her answer.
Insert: He is listening music.
Wrong word: He is becoming to mature.
Word order: I and my sister came.
Redundant/Repetitious: She is a famous and well-known singer.
Wrong form: It has stopped to rain.
Subject-verb agreement: She have much to do.
Verb tense: If I know, I would tell you.
Verb voice: When it be happens, we will know.
Plural: Two woman came.
Article: We saw dog. A dog was named Fido. I like the education.
Pronoun: She is a friend of myself.
Word division: She is my fri-
end
Spelling: freind