The Effects of Time Allocation on Korean College Students’ Performance of Drafted and Timed Essays

Sang-Keun Shin
(Ewha Womans University)


This paper investigates the effects of time constraints on writing performance of Korean college students by comparing the overall quality of two types of writing samples—the term paper with a 5-week time allowance and the in-class timed essay with 30-minute time constraints. Thirty-nine college students majoring in English language education participated in the study, and their compositions were compared in terms of 5 criteria. To ensure the compatibility of the two different types of tasks, two textbook passages whose topic was similar to each other were presented to the students. The writing task in both conditions was to build a lesson plan in an essay format explicitly detailing the rationale for each activity. The t-test results suggest a meaningful difference in the areas of content, language use, and mechanics; but not in organization and vocabulary. The results indicate that there are limitations on measuring students’ writing ability under timed conditions, which was also validated by survey responses collected at the end of the experiment. Implications for writing assessment are offered.

I. INTRODUCTION

With the development of the awareness that the best way to evaluate the writing abilities of an examinee is to make the examinee write a piece of work of his or her own, timed essays have established themselves as a main method of writing assessment (Hughes, 2003; Weigle, 2002). In most situations, timed essays are conducted by giving the examinee a topic and setting a time limit for the examinee to write an essay based on his or her own experience or opinions.

The validity of timed essays was assumed to be very high when they were first
introduced because the writing abilities of the examinee are directly evaluated. However, further studies on the second language composing processes have shed light on the limitations of timed essays. In most cases the second language composing process is a non-linear, exploratory, and generative process whereby writers discover and reformulate their ideas as they attempt to approximate their intended meaning (Zamel, 1982, 1983). The writing process is also recursive instead of linear because writers are constantly cycling through planning, writing, revising, and editing stages. The problem is that many characteristics of real-life writing are not maintained in the timed essay (Belanoff, 1994; Casanave, 2004; Horowitz, 1993; Young-Ju Lee, 2006; Ruth & Murphy, 1998; Sanders & Littlefield, 1975; Sang-Keun Shin, 2005). It is often the case that the examinee taking a timed essay test has to think about something to write about the assigned topic he or she has not thought about before, write, revise and complete an essay in a short period of time.

There have been many efforts to overcome the limitations of time essays, such as allowing the examinee to select a topic (Spaan, 1993), introducing a process-focused evaluation method where the examinee can receive feedback (Yeonsuk Cho, 2003; Young-Ju Lee, 2006), and providing reading materials (Johns, 1991; Lewkowicz, 1997). Attempts have also been made to enhance the usefulness of timed essays by increasing the amount of time allowed to examinees. Increasing the amount of time can especially be seen as an attempt to allow sufficient time to the examinee so that he or she goes through the most natural writing process possible, thereby allowing test users to infer test takers’ writing abilities in a more accurate way. From the perspective of writing assessment, there is a need to examine whether a difference in the amount of time allocated actually leads to different writing performance. If there is no difference in the quality of writing when additional time is allotted, there is no need to allocate more time for examinations.

There have been many studies to examine whether the difference of time allocated on a writing test leads to significant differences in the writing performance of examinees, and their results are contradictory (e.g., Biola, 1982; Elder, Knoch, & Zhang, 2009; Livingston, 1987; Powers & Fowles, 1996; Wild, Durso, & Rubin, 1982).

It should be noted, however, that since most of the prior studies were conducted under the large frame of timed essays, there were limitations to examining the effects of difference of time on the writing test results. In order to properly understand the effects of time difference on writing performance, the situation of a timed essay should be compared with an actual writing situation, not another examination situation (Powers & Fowles, 1996; Read, 1991). As Read (1991) puts it, substantially increased time allocation for completing a test task does not alter the fact that the students are being required to write under constraints that do not normally apply to the writing process. Since various elements other than time can have an effect on the study results, one may argue that such confounding variables should be controlled to enhance the internal validity of the studies.
However, since writing tests aim to help test users to infer the writing abilities of an examinee in an actual situation, there is a need to compare the results of something written in a timed essay situation and in a real life situation (Bachman, 1990; Bachman & Palmer, 1996; Messick, 1989). Furthermore, most prior studies compared holistic scores and thus failed to reveal the areas of writing that are especially affected by the difference of allocated time. In order to accurately understand the effects of difference of time allocation on the different aspects of writing, analytic scoring should be used rather than holistic scoring and examine which aspect of writing is more susceptible than others to time limits.

The present study aimed to fill the gap in the literature by comparing pieces of writing produced in timed essay situations and at-home situations. Specifically, it investigated whether 30-minute timed essays students wrote as a part of their final exam elicited different scores from those achieved on their term paper produced over a 5-week period. After completing their timed essays, the participants completed a short questionnaire designed to elicit their opinions of the two writing conditions. The ecological validity of the study is believed to be very high because pieces written by examinees for course grades, not for research purposes, were analyzed.

The research questions for the study are as follows.

1. Is there a difference in the quality of writing between term papers and 30-minute timed essays?
2. What are writers' perceptions of the two different composing conditions?

II. LITERATURE REVIEW

The biggest advantage of a timed essay compared to an indirect form of writing assessment is that the writing abilities of the examinee are evaluated based on what the examinee has actually written. Most large-scale writing examinations pose a general topic and ask the examinee to write about the topic in three or four paragraphs based on his or her experience and knowledge under time constraints. For example, Test of Written English (TWE) gives 30 minutes, and the International English Language Testing System allows 60 minutes for two writing tasks of 150 and 250 words, respectively.

As mentioned above, the validity of timed essay tests has been called into question in that people can take much more time to write in real-life situations than in timed-essay situations. Many attempts have been made to examine whether scores improve if more time is given for timed essays. The results of the major studies are summarized below.

Hale (1992) examined whether the rank of examinees changed when the time given on TWE changed from 30 minutes to 45 minutes. The correlation coefficient was estimated to examine whether the ranks of examinees were the same, and the results showed that there
was no big difference in rank. However, there was a significant difference in examination scores. The mean scores on the 6-point TWE scale were found to be higher by about 0.25 to 0.3 points under the 45-minute condition than the 30-minute condition, indicating that the provision of additional time produced a modest increase in scores.

Powers and Fowles (1996) examined whether there was a difference in the results of the Graduate Record Exam writing test when the time allowed was changed from 40 to 60 minutes. Three hundred two university students were divided into two groups, and each was asked to write about two different topics. The results were holistically scored on a scale of 1 to 6. The examinees were also asked to report whether they wrote quickly, at a normal speed, or slowly, and the researchers examined whether there was any relationship between the participants’ writing styles and the effects of the amount of time. The results showed that, on average, test performance was significantly better when the examinees were given 60 minutes instead of 40 minutes. However, there was no interaction between the participants’ writing style (fast vs. slow) and time limits. That is, examinees who identified themselves as slow writers did not benefit any more from generous time limits than did their quicker counterparts. In addition, most examinees responded that 60 minutes was more appropriate for the exam than 40 minutes.

Yeonsuk Cho (2003) compared the placement results and the textual quality of timed essays and workshop-based essays produced by 57 graduate ESL writers. In the workshop, the writers were allowed to revise their first draft, incorporating feedback from two peer examinees. The results showed that for the majority of examinees, the workshop helped them receive higher placement results. The results also showed that the workshop essays received higher ratings in content, organization, and linguistic features whereas lower scores in source attribution.

Knoch and Elder (2010) compared the levels of writing of thirty examinees that had been timed for 30 minutes and 55 minutes, respectively. First, in order to examine if the ranks of the examinees were the same for the two writing environments, the correlation coefficient was estimated. While all the correlations were statistically significant, their strength varied somewhat across rating categories. In order to find out if there was a statistically significant difference in the scores received by examinees in the two different environments, paired samples t-test were conducted. The results showed that there were no statistically significant differences between the two performance conditions on any of the rating criteria. Meanwhile, the results of detailed discourse analyses yielded mixed results and thus did not allow definitive conclusions to be drawn about the effects of time limits on the writers’ performance. Specifically, the longer writing condition yielded better quality performance on 5 variables whereas the students produced a higher density of linking devices under the short condition. Finally, the performance on 7 variables was unaffected by the time factor.
The Effects of Time Allocation on Korean College Students’ Performance of...

Only a few prior studies have compared timed essays with essays produced at home. Caudery (1990) asked 24 Cypriot students to write a piece for 40 minutes and then to write another argumentative essay for 1 class hour and completed it at home over a period of 2 days. He reported that there were no significant differences in performance across the two tasks.

Kroll (1990) conducted a study with a similar design. She compared the qualities of writings that were done in a classroom for 60 minutes, and for 10 to 14 days at home. The participants were 25 English as a Second Language writers enrolled in special sections of freshman composition for international students. Each writer contributed four essays: two written in a 60-minute time frame and two written over a longer period time. The results showed that the piece done at home had a slightly higher score than the piece produced in the classroom, but there was no statistically significant difference.

As shown above, the results of previous studies are somewhat contradictory. In terms of the design of study, most of the studies have limitations in that they compared writers’ performance in a timed essay environments and did not attempt to compare timed essays with pieces of writings students worked on for a long time. To date few studies have adopted such a research design, and extant studies have primarily focused on a piece that was a part of writing class assignments, not a high-stakes piece like something a student would write for a term project or a final exam. In order to overcome the limitations of previous studies, this study compared a final paper students worked on over a period of 5 weeks and a 30 minute timed essay that was part of a final exam. The participants’ perceptions of the two different writing environments were also studied.

III. RESEARCH METHOD

1. Participants

The participants of the study were 39 students who were enrolled in the English language teaching methods course at a university in Seoul, including 33 juniors and 6 seniors. All of the participants were majoring or double-majoring in English language education. All of them stated in the background information form that they were quite familiar with lesson planning activities which were the main writing task of the present study.

2. Writing Task

The topics and genres of the two writing tasks were controlled to enhance the internal
validity of the study. The tasks for both cases were to prepare a lesson plan for a reading passage on cultural differences from grade 9 English textbooks. The task required them to present one pre-reading activity, two while-reading activities, and one post-reading activity for the given passage. The writers were asked to describe the activities in detail and to provide the rationale of each proposed activity. The lesson plan was to be in essay form.

As for the final paper, the task was presented in the course syllabus at the beginning of the semester, and students were provided with a passage in Week 10 and asked to submit the paper in Week 15. The students wrote their timed essay as a part of their final exam within 30 minutes. The text presented for the final paper was on differences in food culture, and the topic for the final exam was on gestures across cultures.

3. Rater Training

The writing of the participants was scored on a scale of 1 to 4 in five areas: contents, vocabulary, language use, organization and mechanics. Two instructors in the college English program were asked to evaluate the essays based on the well-accepted scale developed by Jacobs, Zinkgraf, Wormuth, Hartfiel, and Hughey (1981). Both of them had extensive experience in teaching English composition at the college level. Rater training was undertaken before they started to rate the essays. First, the writing tasks were presented to the raters, and the raters were asked to brainstorm reading activities for the two reading passages. Then they discussed the appropriateness of the activities with the researcher. Next, they scored nine activities that appeared most frequently in the students’ essay (three for each stage) and six activities that appeared the least (two activities for each stage). The raters were not informed of the number of the students who proposed each activity. As a final step, they rated 10 lesson plans produced by students in different sections of the same course, and the results were compared. In case the raters’ raw scores varied by more than 2 points, each rater first provided a rationale for why they assigned the particular score and then resolved the discrepancy through discussions.

After the rater training was successfully complete, the pieces written by the participants were copied and then given to the raters to be scored individually. The average score of the two raters for each of the criteria was calculated as the final score. When the two raters had differences of more than 2 points, the researcher reviewed the lesson plan as a third rater and the final score was based on the average score of the third score and the score most similar to the third score.

4. Data Analysis

The Pearson product moment correlations between the two raters in five areas...
ranged between 0.89 and 0.97, demonstrating high inter-rater reliability. Paired samples t-tests were conducted to examine whether there was a statistically significant difference in the students’ performance in the two writing conditions. The significance level was set at .05. The participants were asked to answer open questions on whether they thought the piece of writing they wrote properly reflect their writing ability and why. Their responses were carefully analyzed and categorized into common themes in order to identify the writers’ perceptions of the two writing environments.

IV. RESULTS AND DISCUSSION

1. The Effects of Time Limits on the Quality of Writing

In order to examine the effects of time limits on writing performance, the final papers and timed essays were scored in five areas and the descriptive statistics are presented in Table 1.

<table>
<thead>
<tr>
<th>TABLE 1</th>
<th>Descriptive Statistics of Essay Scores</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Term project</td>
</tr>
<tr>
<td></td>
<td>M</td>
</tr>
<tr>
<td>Contents</td>
<td>2.744</td>
</tr>
<tr>
<td>Organization</td>
<td>3.487</td>
</tr>
<tr>
<td>Vocabulary</td>
<td>2.410</td>
</tr>
<tr>
<td>Language use</td>
<td>2.795</td>
</tr>
<tr>
<td>Mechanics</td>
<td>3.872</td>
</tr>
</tbody>
</table>

As Table 1 shows, the scores for the final paper were higher than those from the 30 minute timed essay in all five scoring areas. The biggest difference in score was found in mechanics, and the scores were most similar for organization. Vocabulary scores were lowest out of the five scoring areas for both writing situations.

Paired samples t-tests were conducted to find out if there was a statistically significant difference in the levels of writings in the two environments, and the results are shown in Table 2.
As shown in Table 2, there was a statistically significant difference in three scoring areas: contents, language use and mechanics, but there were no significant differences in organization and vocabulary. The results suggest that the effects of writing time are different across the components of writing.

First, when it comes to the content of their writing, the most noteworthy difference was that the writers provided much more detailed description of their activities in their term papers than in their timed essays. For example, as can be seen Figure 1, one student started her term project with an overall description of the lesson and then went on to describe her pre-reading activity. Meanwhile, the same writer did not describe the overall structure of the lesson and briefly presented her pre-reading activity, as can be seen in Figure 2.

There was also a significant statistical difference in language use and mechanics. The students made more errors in verb tenses and subject-verb agreement in timed essays than in their term projects. They did not employ as many cohesive devices as they did in their final projects. They also misspelled considerably more words and made more capitalization mistakes in timed essays. The writers rarely made such mistakes in essays composed on computer.

### FIGURE 1
Sample Term Paper

The overall goal of this lesson is to encourage the students to acknowledge cultural diversity regarding foods. Also, the students will be able to comprehend the logical structure of the text, especially cause and effect. Thus, this class will mainly focus on reading skill, though listening and speaking activity will be included.

To begin the lesson, the teacher brings out the subject and shares some experience of eating in local restaurant in India while traveling there. Several pictures taken during the trip will be shown as well, to get the students be familiar with the topic. The students will come to know that Indians use right hand to eat and left hand in the bathroom. Also, students can share their experience of exotic meals.

Announcing today’s topic: ‘food and culture’, the teacher will show the students short video clips of documentary films about eating habits of diverse countries. These films
will include people eating meats, insects, or rice, and eating with their hands, chopsticks, etc. Then the students will guess and briefly discuss what people in the video clips are eating, what they are using to eat, and what makes difference in eating habits or customs of each country. This activity will generate the students' interests by watching lively and realistic material to the reading contents.

FIGURE 2
Sample Timed Essay

In pre-reading stage, students are asked to tell about the taboo, politeness, and body languages that contains certain meaning in Korea. For example, it is not allowed to wave hand to say hello to grandparents, whereas it is acceptable to friends. Students are free to say anything about these things, and it is allowed to use mother tongue. Then the teacher explains that there are situations, locations, and any context that some behaviors are allowed, but some others are not. Here, the teacher prepares students before they read the passage about different customs, beyond within Korea.

There was no big difference for organization, and this appears to be because of the characteristic of the task. Since the students had to provide about pre-, while- and post-reading activities, they did not have to worry about how to structure their essays. It should be noted, however, that transitions between paragraphs were much smoother in the term projects than in the timed essays. Whereas the students wrote transition sentences between the paragraphs in their term papers, they heavily relied on the 'in pre (while, and post)-reading stage' phrase in their timed essays.

Lastly, there was no big difference for vocabulary, suggesting that vocabulary is an area that does not easily improve even if more time and reference materials are provided. This finding is consistent with the study results that examinees had the most difficulty with vocabulary when writing (Leki & Carson, 1994; Uzawa & Cumming, 1989).

2. Perceptions of the Examinees

When asked whether they believe different writing environments had an effect on their writing performance, 84% of the students stated that the final paper they submitted reflected their writing competence better than the timed essays, and the rest of the participants thought that there was no difference in the qualities. None of the students responded that their timed essays better reflect their writing competence.

The 32 students, who believed that different time conditions led to the differences in the overall writing quality, were asked to explain any differences they experienced while producing their lesson plans in the two situations. Their responses are summarized in Table 3.
The participants stated that they preferred the actual writing situation at home because they had ample time to come up with the activities which they thought were the main focus of their task. They also pointed out that they were able to search books for activities, and they were also able to evaluate their activities and came up with new activities if necessary. This finding is consistent with the observation that "academic writing outside of testing situation typically involves the use of source materials as input" (Weigle, 2002, p. 52), suggesting that writing tests using "bare prompts" (Hamp-Lyons & Kroll, 1997, p. 13) is rather inauthentic.

TABLE 3
Reactions of Examinees

<table>
<thead>
<tr>
<th>Opinion</th>
<th>Number of responses</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Contents</strong></td>
<td></td>
</tr>
<tr>
<td>There was enough time to select activities that corresponded to the characteristics of the text.</td>
<td>18</td>
</tr>
<tr>
<td>I was able to take a look at the activities in books.</td>
<td>6</td>
</tr>
<tr>
<td>I was able to use a wider variety of expressions without repeating the same expressions.</td>
<td>11</td>
</tr>
<tr>
<td>I was able to find appropriate expressions in text books or reading materials, etc.</td>
<td>5</td>
</tr>
<tr>
<td>I was able to look up appropriate expressions on the internet or online dictionary and used them in the essay.</td>
<td>31</td>
</tr>
<tr>
<td>I was able to re-read my essay and revise expressions.</td>
<td>13</td>
</tr>
<tr>
<td><strong>Language use</strong></td>
<td></td>
</tr>
<tr>
<td>It was convenient because the computer automatically identified spelling or grammatical mistakes.</td>
<td>29</td>
</tr>
<tr>
<td>I was able to edit my English expressions.</td>
<td>25</td>
</tr>
<tr>
<td>There was not enough time to review my sentences. I was too busy finishing the essay in time.</td>
<td>17</td>
</tr>
<tr>
<td><strong>Mechanics</strong></td>
<td></td>
</tr>
<tr>
<td>It was inconvenient because I rarely hand write anything these days.</td>
<td>9</td>
</tr>
<tr>
<td>I was able to go back and forth to write the essay because revising was so easy on computer.</td>
<td>7</td>
</tr>
</tbody>
</table>

The students' responses also revealed that time issue was inextricably interwoven with paper and computer modes. They reported that revising and editing were easy when they worked on the computer. They were able to work on different parts of the essay, going back and forth. When they wrote a timed essay, however, they had to write it in a linear fashion. One student explained the difference as follows.

When I wrote the timed essay, I knew that I would not be able to make huge changes like copying a whole paragraph and then paste it in a different place. So I wrote
everything very carefully. But when I wrote the term paper on computer, everything was so quick and easy.

This finding is consistent with Young-Ju Lee (2002) which reported that the writers focused on the phrase and sentence level issues when they wrote their timed essay on paper.

The difference in mechanics scores may partly be due to the fact that word-processing programs helped them to avoid spelling and small grammatical mistakes. This finding is consistent with that of Kenworthy (2006). He also found that his students made use of the built-in spell-checking tool of word processing programs, which led to the significant reduction of spelling errors.

While working on the final paper, they looked up vocabularies and expressions in the internet or online dictionary. One student described her composing processes as follows.

Whenever I was not sure whether a particular expression was ok, I typed in the expression in the web browser’s search box and examined how many instances there were. Sometimes I compared the relative frequency of two or three expressions and chose the one with the highest frequencies.

This finding gives an important clue that explains why they performed better in terms of vocabulary and language use in the drafted writing situations than in the timed essay situations.

Meanwhile, those who stated there was no big difference between their two essays reported that they put off writing the paper until the last minute, and thus they did not really have time revising and even proofreading their papers carefully. They did not make the most of the additional time they were given. Kroll (1990) also reported a similar problem. She speculated that some of her students might have spent less amount of time on their out-of-class papers than they did on their timed essay and noted that we have to teach our students to give adequate time to writing tasks and “to make the time spent more productive” (p. 153).

V. CONCLUSION

This study aimed to examine the validity of timed essays and compared essays written for a final course paper and a timed essay which was a part of a final exam in terms of contents, organization, vocabulary, language use and mechanics. The results of the paired samples t-tests revealed a significant statistical difference for contents, language use and
mechanics, but there was no difference in organization and vocabulary. The results showed that there are limitations for a timed essay to infer the actual writing abilities of a subject. Many factors have to be evaluated comprehensively to determine the usefulness of tests for a given purpose (Bachman & Palmer, 1996), and there is no best test method. However, the results of this study showed that test users may not be able to accurately infer test takers’ writing performance in real-life situations based on their performance on timed essays. This is not to say that timed essay exams are not authentic form of writing. This is because students are often required to take timed essay examinations in their content courses (Hale et al., 1996; Weigle, 2002). This is to say, however, that timed essay testing is inauthentic in that academic writing outside of the testing situation is not timed and typically involves the use of source materials as input.

Since the participants were able to employ various composing strategies, including reading relevant materials and consulting dictionaries, which they were not able to do so in timed essay situations, one may argue that these are confounding variables that make the comparisons this study has made problematic. One can also claim that a timed essay reflects the pure writing abilities of examinees. It should be noted, however, that the purpose of writing assessment is to infer test takers’ ability to write in non-test situations, not their performance in timed essay contexts. Since the present study examined the validity of timed essay tests by comparing the students’ performance in drafted writing situations and timed essay situations, it is expected to provide a truer picture of the effects of composing environments on writing performance. However, the small sample size of this study should be taken into consideration when evaluating these results. Given that the present focused only on lesson planning tasks, the impact of the time limits on other forms of writing also warrants further study. Finally, future studies may compare textual qualities of essays written in two different composing environments.

REFERENCES


Lee, Young-Ju. (2006). The process-oriented ESL writing assessment: Promises and


The Effects of Time Allocation on Korean College Students’ Performance of ...

Applicable levels: tertiary
Key words: writing assessment, composing process, timed essay, drafted writing

Sang-Keun Shin
Dept. of English Education
Ewha Womans University
11-1 Daehyun-dong, Seodaemun-ku
Seoul, 120-750, Korea
Tel: (02) 3277-4648
Email: sangshin@ewha.ac.kr

Received in December, 2010
Reviewed in February, 2011
Revised version received in March, 2011