

THE EFFECTS OF TYPED VERSUS HANDWRITTEN ESSAYS ON STUDENTS'  
SCORES ON PROFICIENCY TESTS

By

Erika Lessien

A THESIS

Submitted to  
Michigan State University  
in partial fulfillment of the requirements  
for the degree of

Teaching English to Speakers of Other Languages - Master of Arts

2013

**ABSTRACT**

# THE EFFECTS OF TYPED VERSUS HANDWRITTEN ESSAYS ON STUDENTS' SCORES ON PROFICIENCY TESTS

By

Erika Lessien

Previous researchers (Lam & Pennington, 1995; Lee, 2004) have investigated the difference between language learners' L2-writing-test scores when the learners are required to type essays compared to when they must handwrite them. The outcomes have been mixed, and this may be because the researchers did not investigate whether L2 proficiency impacts score differences. Therefore, in this study I will investigate the score differences in advanced versus intermediate-level English-language learners on handwritten versus typed essay tests. Sixty-one students, from three different proficiencies, were asked to handwrite one essay and type another from prompts retired from the university's English-language placement test. Two trained raters rated the essays using the university's placement test rubric. Using a multiple linear regression, I compared score differences across the conditions (handwriting versus typing) and between the groups (intermediate versus advanced English-language learners). I found that there is a significant difference for the advanced students, and their scores were much greater for the typed condition than for the handwritten condition. This study sheds light on the effects L2 essay test conditions have on L2 test program outcomes; programs that even today use handwritten essays to assess language learners' academic-writing ability

## ACKNOWLEDEMENTS

I would like to take this opportunity to say thank you to the people who helped and encouraged me in the writing of this thesis.

First and foremost, I would like to thank my thesis advisor, Dr. Paula Winke, for all of the time she devoted to this project and for teaching me how to use SPSS and encouraging me throughout the course of this project. Without her help and support it would not have been possible. I would also like to thank my second reader, Dr. Charlene Polio, for all of valuable comments and suggestions she gave. I would like to thank the MA TESOL program for their financial support throughout this project. I would like to thank Mike Kramizeh, the head of the language laboratory at Michigan State University, for the use of the language laboratory during my data collection.

I have received much support from my classmates and colleagues and I am grateful to them for their help. I would like to give a special thank you to Jack Drolet and Peter Sakura for giving up so much of their time in helping me complete several important steps during this project. I would also like to thank Dr. Dan Reed and Andy McCullough for helping me during my very beginning planning stages. A special thank you to Justin Cubilo for helping me work through parts of this project. Also, thank you to Laura Ballard and Lorena Valmori who helped me with some important last minute changes.

Finally, I would like to thank my mom for always being there to encourage me and to help me when I needed it, and my grandparents for always supporting and encouraging me in all of my endeavors throughout my life.

## TABLE OF CONTENTS

LIST OF TABLES .....	iii
LIST OF FIGURES .....	iv
<b>CHAPTER 1:</b>	
<b>INTRODUCTION .....</b>	<b>1</b>
<b>CHAPTER 2:</b>	
<b>LITERATURE REVIEW .....</b>	<b>3</b>
Theories and models on assessing writing .....	3
Studies on the use of paper-based versus computer-based assessments .....	6
<b>CHAPTER 3:</b>	
<b>METHODS .....</b>	<b>15</b>
Participants .....	15
Materials .....	16
Procedure .....	17
Study Approval .....	17
Setting .....	17
Computer Set-up .....	18
Data Collection .....	18
Essay Rating .....	20
Analysis .....	21
<b>CHAPTER 4:</b>	
<b>RESULTS .....</b>	<b>23</b>
Research Question 1 .....	23
Research Question 2 .....	26
Research Question 3 .....	29
<b>CHAPTER 5:</b>	
<b>DISCUSSION AND CONCLUSION .....</b>	<b>31</b>
Categorical Changes .....	31
The Effects of Proficiency, L1, and Preference .....	35
Student Perceptions .....	38
Implications .....	39
Limitations .....	41
Directions for Future Research .....	42

<b>APPENDICES</b> .....	45
<b>Appendix A:</b> .....	46
<b>Appendix B:</b> .....	47
<b>Appendix C</b> .....	48
<b>Appendix D:</b> .....	50
 <b>REFERENCES</b> .....	 53

## LIST OF TABLES

Table 1: <i>Participants' Backgrounds</i> .....	16
Table 2: <i>Rater Reliability</i> .....	23
Table 3: <i>Descriptive Statistics for Writing Medium</i> .....	24
Table 4: <i>Paired Samples t-test</i> .....	24
Table 5: <i>Word Counts by Level</i> .....	26
Table 6: <i>Regression Factors and Their Effect</i> .....	27
Table 7: <i>Correlation of Scores</i> .....	28
Table 8: <i>Student Perceptions: Typed versus Handwritten</i> .....	30
Table 9: <i>Student Perceptions: Which Essay Will Be Scored Higher</i> .....	30
Table 10: <i>Rubric</i> .....	50

## LIST OF FIGURES

Figure 1: <i>Assessing Writing</i> .....	4
Figure 2: <i>Test Scores vs. Profeciency</i> .....	29

## **CHAPTER 1: INTRODUCTION**

When students learn to compose essays in a second or foreign language, they often do so, at first, by handwriting their essays in or outside of class. As they advance in the language, they may move to a computer-format for essay composition, but at which level of proficiency that change occurs most likely varies depending on the language-learning program, their overall computer literacy, and the differences between the students' first or native language (their L1) and the language being learned (the L2). However, computers are the norm for essay composition in most higher-educational, academic settings. Thus, an important question is the following: if a program is using an essay test to assess academic-writing skills, should the test format comprise handwritten or computer-processed essays, or does it matter? And how does that decision relate to the test takers' proficiency and computer literacy in the language being assessed?

In general, there is a need to understand the effect that word-processing has on language learners' essay test scores. In this study I investigate the score differences of 61 advanced versus intermediate-level ESL students, comparing their handwritten versus typed essay test scores. The learners were asked to handwrite one essay and type another, using prompts retired from a university's English language placement test. In addition, students responded to a survey with closed and open-ended questions to investigate their views of typed versus handwritten essay tests in order to understand which they prefer and why. Using a multiple linear regression, I compare score differences across the conditions (handwriting versus typing) and between the groups (intermediate versus advanced test takers). This research is important because programs continue today to use handwritten essays to assess student-academic writing ability, which I hypothesize might



be problematic theoretically (in relation to the construct being measured) and practically, and in particular for advanced-level students used to composing essays using a word processor.

## CHAPTER 2: LITERATURE REVIEW

In this chapter I discuss previous research on writing assessment and computer-based versus paper-based writing assessment in particular. In the first section I discuss theories and models of assessing writing. In the second section I discuss previous literature conducted on the use of paper-based versus computer-based assessments. Third, I discuss a gap in the literature and this study's research questions that aim to address that gap.

### **Theories and models on assessing writing**

Bachman and Palmer (2010) wrote that *test usefulness* is the most important thing to consider when designing and developing tests, be they tests of writing or any other construct in L2 assessment. They included six factors under the umbrella of a test's *usefulness*: reliability, construct validity, authenticity, interactiveness, impact, and practicality. In this section I focus most on the area of construct validity. In other words does the assessment accurately and appropriately measure the construct it is meant to. In university placement tests, the construct is typically about the students demonstrating that they have knowledge of the English language in relation to academic writing. So, for these types of tests it is important that they accurately measure not just writing, but academic writing. Because, most academic writing is done on a computer, it seems that it would be important for the test to be representative of this in order to have construct validity.

Figure 1.

*Assessing Writing*

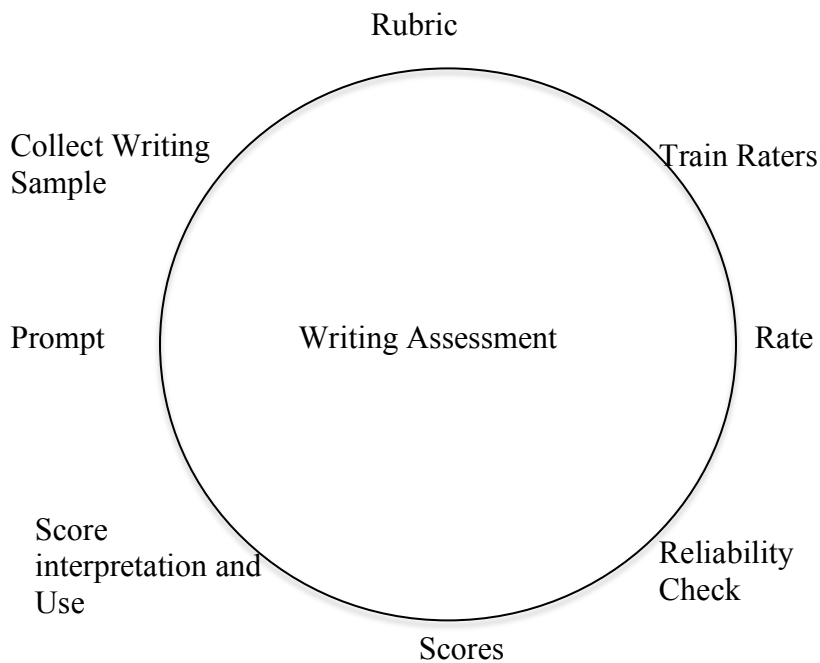


Figure 1 demonstrates that there are many factors to consider when designing a writing assessment, including the topic, the time limits, the discourse mode, the genre, the writing mode (typing or handwriting), the testing conditions, rater inconsistencies, scoring procedures (holistic or analytic), and traits to be scored (language use, spelling, ect...) (Schoonen, 2005). As Gerbril (2009) found, all of these factors can affect the generalizability and validity of a test. Each factor must demonstrate some kind of validity and each must correctly contribute to the writing assessment in order for the assessment to be reliable and valid. The current study deals mostly with the potential effects of the writing mode (typing or handwriting) on students and investigates whether these two types of writing are both generalizable for the student or if one is more accurate than the other.

There are two main types of tasks used in timed writing test situations, the independent writing task, in which students are given a generic prompt and asked to compose an essay discussing their ideas about the prompt, and the integrated writing task where students are given a listening passage and a reading passage and asked to discuss the information in relation to the passages. Despite the criticisms that independent writing tasks do not accurately measure the academic writing construct, (Gerbil & Plakans, 2009; Weigle, 2002; Hamp-Lyons & Kroll, 1996) many universities use still using direct writing tasks for proficiency tests. Hamp-Lyons and Kroll (1996, p. 18) call this method a *snapshot* method because it takes a quick picture of a students writing ability. This is likely because integrated writing tasks require more time to both create and administer (Plankans, 2010). Also, according to Plankans, integrated writing tasks can, for some test-takers, be considered more complicated. Independent writing tasks also do not assume background knowledge and are easier to create and administer. Also, in defense of independent writing tasks, Gebril (2009) found that students do preform similarly on integrated and independent tasks.

Whichever prompt they use, it is common for universities to use handwriting as the medium for composition. There are several reasons that universities, at least in the U.S., are hesitant about switching from paper-and-pencil proficiency tests to computer-based tests. One such reason is that many universities feel that since many languages do not use a Roman alphabet, language learners with those L1s may not be familiar with the English keyboard and thus may have lower scores that do not reflect their true essay-writing ability. A second reason is the demand that such writing tasks can place on the

university. Having typed essays requires the use of large computer labs, which some universities do not have access to.

### **Studies on the use of paper-based versus computer-based assessments**

This effect of computers on students' test scores seems to be a controversial topic. Researchers investigating this topic have not come to a clear consensus on whether or how computers affect test scores. Concomitantly, research that has been conducted on the differences between typed and handwritten essay tests has been inconclusive thus far. On one hand, Benesch (1987) found no difference between test scores that students received on handwritten versus typed tests. On the other hand, Lam and Pennington (1995) found that there was a positive effect for typing: that is, in their study, scores increased when students were allowed to type their essays. Lee (2004) also found a positive effect when students typed essays compared to when they handwrote them. She conducted a study looking at the quality of written products that were typed versus those that were handwritten. Lee used handwritten, typed, and transcribed essays (to control for the effects of raters seeing the handwritten script) to see the differences in the scores and if the differences were due only to the fact that they were typed. She found that typed essays, in general, were scored higher than handwritten essays, but that the transcribed essays were in some cases scored lower than typed essays. It is important to note here, however, that a majority of these studies (Benesch, 1987; Lam & Pennington; even Lee, 2004) are fairly old. Those that found that there was no difference between typed and handwritten essays (i.e., Benesch, 1987) were conducted before computers were as widely used as they are today. This means that in the last few decades, the effect of

computers on students writing may have changed. Prior studies need replicating with current, more computer-literate students.

Of importance is a study by Wolfe, Bolton, Feltovich, and Niday (1996). The authors looked at the difference in scores when students were allowed to choose to either compose essays on a computer or on paper. This study investigated whether performance on a writing assessment is comparable when examinees are given the choice between composing essays using a word processor and handwriting essays. They found that students with weaker English-language proficiency did better on the handwritten exam, and students with higher English-language proficiency did worse on the handwritten exam. The findings of this study must be considered with care because the authors looked at score differences when students were able to *choose* their writing medium. That is, the test takers were able to select the group and writing condition to which they would be assigned; this made the study a non-randomized, between-group design. More preferable would be to compare the differences of handwriting versus typing using a randomized study (in which test takers were randomly assigned to type or handwrite). Even better than that would be to have a between-groups study design in which each student would type and handwrite essays; a researcher could then compare the results without the confound of writing-type preference. It is difficult to say what the results would be for the individual students without being able to compare the essays for the same student.

Qualitative differences in students' writing that may arise from typing instead of handwriting should also be investigated, and some researchers have looked at this. For example, Schwartz, Fitzpatrick, and Huot (1994) found that students who used word

processors were producing both longer and richer texts. Whithaus, Harrison, and Midyette (2008) looked at raters' attitudes towards typed versus handwritten essays, and whether the types of errors produced differed between handwritten and typed essays. Like Wolfe and Manalo (2004) and Wolfe, Bolton, Feltovich, and Niday (1996), students were allowed to pick their writing medium. Raters noticed that students' proofreading abilities seemed to decrease on typed essays. Raters also noticed more spelling and grammar errors on the handwritten essays, though the handwritten essays were still considered stronger. Raters also noticed a difference in organization; the raters noted that in general typed essays were organized better than the handwritten essays. While these results are interesting, much care needs to be taken when considering them, primarily because different students typed or handwrote the essays (each student did one or the other, not both) and they self-selected the way in which they would compose their essays. There was no real way to know if the differences were student-based (better-abled students chose to type) or if the score differences could be attributed to the writing format.

Typed essays have many other advantages over handwritten essays. According to Susser (1994), the writing process itself is beneficial to students' writing. Individuals learn to write when they pre-write, edit, and revise the writing. In timed-writing situations, like those on proficiency tests, however, it is difficult for students to have the time and the means to pre-write, edit, and revise without affecting their essays in another way. For example, Harrington (2000) stated that the appearance of the essay could have an effect on the score that the rater will give the essay. Unfortunately, in agreement with this situation, many raters in Harrington's study claimed that how the essay looked and

how easy the essay was to read caused them to change the way they rated. One implication from Harrington's study could be that editing on handwritten essays may be harmful to students' scores rather than helpful, as it should be. This is because to edit on paper, students must erase or cross out the parts they want to delete and add new sentences or words, and sometimes they use arrows or other marks indicating the additions or subtractions of text. This can lead to messy-looking essays. Students are also limited by the amount of space available for editing. Typed essays may be a solution to this problem. Although learners will still have time limits on their essays, they will not have the space limitations, nor will they risk a lower grade based on the appearance of their writing. This claim that editing on handwritten essays can be harmful to students and that editing on paper essays is difficult to do is supported by Daiute (1985) who wrote that computers better allow students to fully incorporate the writing processes when composing short, timed essays.

Not only do raters notice that handwriting essays can be harmful to students' scores, but the students themselves seem to be aware of the harmful affects that can be caused by changing their paper-based essays too much. Students tend to be reluctant to make changes to their handwritten essays because of the fear of making it look less neat (Daiute 1985). Daiute suggested that students seem to be much more willing to make changes to their timed essays when they are able to write them on a computer. Li (2005) found that the number of revisions was significantly higher for computer-processed responses than for handwritten responses. Li noted that not only do students appear to be more willing to make changes to their typed essays, but they have also been found to spend significantly more time searching for words or phrases, evaluating written texts,



and have more *decision-making episodes* (as Li wrote, these are when students engage in pre-planning, in process planning, consider spelling, reason about linguistics choices, and search for correct words or phrases.) in computer writing. Overall, Li's study showed that participants were able to revise at higher levels when using a computer compared to using paper.

Another consideration that needs to be made when considering whether students' writing proficiencies should be tested using a computer-based test or a handwritten one is *authenticity*. For a proficiency test to accurately measure a student's academic writing ability, it is important to take into consideration the authenticity of the assessment. Authentic assessments are those that represent how students would be expected to perform in real-life situations. Bachman (1990) and Bachman and Palmer (2010) claimed that authenticity plays an important role that works together with construct validity to determine how the construct definition and the domain of generalization will affect the way in which a test score will be interpreted.

Over the past few decades, using authentic assessments is becoming more and more important and central to language learning (Douglas, 2000; Lewkowicz, 2000). Because most writing in university and professional settings is done on computers, it is important that the way in which a student's writing ability is measured is representative of how they will need to perform in the real world, or in this case, the university world. With the emphasis on computers today, it seems more realistic for students to perform writing tasks, especially those determining their academic proficiency, on a computer rather than with a pencil and paper. Lee (2004) agreed that computer-based tests allow

for a more authentic writing experience, and thus an increased chance of success for students who are academically prepared and able.

Finally, studies have looked at learner perceptions of typed versus handwritten essays. Many have found that most students feel more comfortable typing rather than handwriting essays. This is most likely due to the fact that, in their home countries, most students type their assignments. Lee (2004) reported that students in her study said they believed that the computer-based-writing medium was more likely to place them into the correct ESL class. This is especially important in large-scale, university placement tests because there are many more problems with students when they do not feel that the placement test could accurately place them into the correct level. If nothing else, Lee's study showed that students find computer-based tests have more face-validity than paper-based tests. Students also reported that they preferred typing to handwriting essays because they feel typing is more familiar, produces more legible text, and is convenient. Students also reported that their errors were easier to spot in a typed essay over a handwritten essay (Whithaus, Harrison, & Midyette 2008).

While previous studies have looked at the differences in scores between paper-based and computer-typed writing assignments, no study has yet looked at the effect of proficiency on the differences in scores between paper-based and computer-based writing assignments. In the current study I will look at score differences across typed essays versus handwritten essays and in comparison to proficiency scores. This will help me determine if there is a positive or negative effect depending on proficiency and also to determine if this effect would be significant enough to change the placement of students in English language classes.

## **The gap in the literature and this study's research questions that aim to address that gap**

As reviewed above, the research topic of how the writing context (handwritten versus typed) affects essay-test outcomes is not new. Several studies (Lee 2004; Lam & Pennington 1995; Wolfe & Manalo 2005) have looked at differences in scores between handwritten versus typed essays, but these studies' results were limited in scope. Few of them compared scores between the same students (Wolfe & Manalo, 2005; Wolfe, Bolton, Feltovich, Niday, 1996; Whithause, Harrison, & Midyette 2002) (these studies used between-group designs, with one group producing handwritten essays, and the other group producing typed essays; scores across groups were then compared. However none looked at the differences in the scores as related to the proficiency levels of the students while comparing scores for the same student. Moreover, none looked at *computer-processed* essays; instead, the researchers whose work is discussed above investigated computer-typed versus handwritten essays. I make this distinction because computer-processing is different from typing on a computer in that when a student computer-processes an essay, he or she may use tools that allow for copying text, moving it to another part of an essay, and deleting text (King, Rohani, Sanfilippo, & White 2008). If using a program such as Microsoft Word, the writer may check and edit, to a certain extent, his or her basic grammar and spelling using Word's grammar and spell check features. Thus, the present study compares computer-processed (using Microsoft Word) and paper-based scores within the same student while also making comparisons among students with different proficiencies. This is important because in the real world's academic context, students do not just use a computer to type their academic work; they

use computer-processing tools, including spell checkers and grammar checkers. Thus, to better approximate the real-world construct of academic essay-writing, I believe tests of academic essay-writing should allow students to use processing tools that are available to them (and that current students use) in the real context. The effect of this method of essay-composition (computer-processed essays) in relation to paper-and-pencil-produced essays has not yet been investigated. A logical step is to do so.

The following research questions will be investigated:

1. Do the results of a placement test differ between a paper-and-pencil based test and a computer-processed test
2. Are there specific areas of writing (measured via scores from the analytic rubric) that seem to improve with one of the writing mediums?
3. Do these results differ when proficiency is taken into consideration, that is, (a) for low-level test takers and (b) for higher-level test takers?
4. Do test takers feel more comfortable computer-processing or handwriting their essay tests?

In relation to research question 1, I hypothesize that there will be no significant difference between paper-and-pencil-based tests and computer-processed tests. I believe that for research question 2 there will be difference based on category looked at as described on the rubric. I also draw on research by Whithaus, Harrison, and Midyette (2008) when they said that raters commented on increased mechanical and spelling errors in typed essays rather than handwritten ones, however, overall they found that the essays increased in quality. I also draw on research by Lee (2004) who found that scores increase in the areas of organization and content. Thus I believe that students will have

increased mechanical and spelling errors on the typed essay compared to the hand written essay and that their scores will go up in the areas of organization and content on the typed essay. Also, like Whithaus, Harrison and Mdyette (2008) found I believe that there will be increased mechanical errors in the typed essays. I thus hypothesize that raters will perceive typed essays as better for high proficiency students but will notice more errors in the writing of lower proficiency students. In relation to research question 3, I hypothesize that low-level test takers will perform worse on the placement test when they must use Microsoft Word due to their unfamiliarity with the English keyboard. On the other hand, I hypothesize that higher-level test takers will perform better on the computer-processed test because they will be more familiar with computers, the English keyboard, and in computer-processing their English-language essays. For research question 4, I have the following hypothesis. Based on results from Lee (2004), I expect most students to be more comfortable with typing over handwriting essays. When divided by proficiency, however, it may be that lower-level students will be less comfortable using a computer to compose essays while students of a higher proficiency may prefer it.

### **CHAPTER 3: METHODS**

In chapter four I discuss the materials used and the data collection procedure. I will conclude this chapter with a discussion of the ways in which the data were analyzed.

#### **Participants**

Participants of this study consisted of 61 ESL students recruited from the low-intermediate level of an intensive English program (IEP) as well as advanced students from an English academic program (EAP) at a large Midwestern university in the United States. It is worth mentioning here that there are two additional levels between the low intermediate and the advanced level students. Students in the same university's Master of Arts in Teaching English to Speakers of Other Languages also participated. The IEP program is intended for students who do not meet the TOEFL score requirements to be enrolled in an academic program, as well as students who want to further their English study and do not plan to enroll in academic classes in the future. The EAP program is meant for students who have completed the IEP program and consists of a series of four courses of which students may take only the specific skills that they need and can take academic courses at the same time. The Master's in Teaching English as a Second Language program is a regular master's program, and the students in this group did not take any classes at the English language center before being accepted to the university.

To recruit participants I visited the classes to explain the project and pass out a sign up sheet where possible. When I was unable to visit a class, I gave a flyer to the teacher so students could sign up to meet at one of six specified time slots. A total of 61 students participated in this study. An exact breakdown of the participants can be seen in

Table 1. As can be seen in this table a majority of the participants were native Chinese speakers and between the ages of 18 and 21.

Table 1.

*Participants' Backgrounds*

	Male	Female	Total
Number of Participants	29	32	61
Average Age	21	22	21
Average Age of Starting English	11	10	11
Level of Study			
IEP	14	10	24
EAP	13	16	29
Master's Student	2	6	8
Native Language			
Chinese	21	28	49
Korean	0	1	1
Bhasa	0	1	1
Arabic	7	2	9
Spanish	1	0	1

## Materials

For this study I used two retired prompts (Appendix A) from an English language test proficiency test at a large Midwestern university. Both the instructions and the format of the placement test were followed as closely in order to keep the task as similar to the actual test as possible. The differences are that in the actual test students would only have to write one essay and they would be required to handwrite the essay rather than type it.

Aside from the essay prompts, I also included a background questionnaire (Appendix C) that asked participants information about their age, sex, language learning background, and time at the University. I also employed an exit questionnaire (Appendix

D) consisting of eight items, five Likert scale items (four of which asked for additional explanation), and three two-choice items that asked for additional explanations. The exit questionnaire asked about the students' perceptions of each task and about their satisfaction with their original level placement.

Finally, the raters used the original rubric (Appendix E) that had been developed and used for the placement test at the university. I chose this rubric not only because it is the actual rubric used to rate the proficiency test, but also because it has several other benefits. The rubric is an analytic rubric consisting of different categories that are scored independently and added together. An analytic rubric is useful because it reflects the different aspects of a test takers writing ability (Weir, 2005) and gives a fuller picture of the differences between the essays due to the variables. Knoch (2009), Weigle (2002) and Bachman and Palmer (1990) also pointed out that analytic rubrics tend to be more reliable and show a greater range in score differences.

## **Procedure**

### **Study Approval**

The university IRB approved the study prior to the start of data collection. The IRB's Policy required every participant to sign a consent form stating the purpose of the study and the risks, benefits, means of ensuring privacy, and the procedures associated with it. Therefore, the students signed a copy of the consent form at the beginning of the session. I passed out and went over the consent form with participants answering questions as they arose. I gave the students a copy of the consent form to take with them after completing the task.

### **Setting**



Testing took place in a language learning computer lab. The lab was equipped with 36 iMac computers running Mac OS X Snow Leopard. Computers were arranged in six rows, each with six computers in it. Seats were arranged so that participants would be facing the front of the room when seated at their computers. Participants sat at every other computer to ensure that there was no cheating. At the front of the room there was a teacher's station connected to a projector, which displayed the amount of time remaining in the test.

### **Computer Set-up**

Each computer had one Microsoft Word document open on it. The document was blank and students were instructed to type their name and the date at the top of the document. No other windows or programs were open, and students were instructed not to touch anything else on the computer but the Microsoft word document. Students were allowed to adjust the size of the document to ensure they could see it and work with it comfortably.

### **Data Collection**

Participants came to the computer lab on the day and time they signed up for and were randomly placed into one of four experimental groups. Each group received their first essay at the same time; however, one group was handwriting while the other group was typing. The method of production, handwritten or typed, was switched for the second essay. In order to keep the essay format as similar to the placement test as possible, students had 35 minutes to write an essay on each of the topics. The experimental groups were arranged in these four conditions:

1. Type Essay A, Write Essay B

2. Type Essay B, Write Essay A
3. Write Essay A, Type Essay B
4. Write Essay B, Type Essay A

Participants were also asked to fill out a survey on their opinions of the two different tests. The expectation of the survey was that it would give a clearer picture of test takers' comfort levels with the two different test styles and whether they are more comfortable typing or handwriting essays.

When they arrived, participants were assigned a computer. When all of the participants were present the researcher passed out a packet to each student. The packet contained all of the pages needed for the experiment. The researcher then explained the consent form. Once participants had signed the consent form and it was collected the researcher asked the participants to fill out the background questionnaire. The background questionnaire was completed before the start of the writing task.

Once all of the students had completed the background questionnaire the researcher collected the questionnaires and then explained the directions for the next part of task. After all questions had been answered and the researcher had ensured that the directions were clear the participants began the first writing task.

The first writing task was hidden under a piece of white paper in the packet. Participants were instructed not to flip past the white paper until they were instructed to do so. After the instructions were explained the researcher told participants to flip over the white paper and the timer started. After the 35 minutes had passed, participants were instructed to stop and the researcher saved each document for the students. Participants

were then given a 10-minute break and the opportunity to have drinks and snacks. This was done in order to prevent any test fatigue that may occur.

After the 10-minute break participants began the second writing task. This task was conducted in the same way as the first, the only difference being that students who had typed the first essay would now be handwriting and vice versa.

Finally, after the second task participants were asked to complete the exit questionnaire. After they had finished answering the questions on the exit questionnaire the researcher checked to be sure all materials had been collected and then participants were allowed to leave.

### **Essay Rating**

I asked two trained and experienced raters of the English language proficiency test used in this study to rate each essay. The raters scored the essays based on the essay rubric from this specific proficiency test. The raters were the same as those who rated the official test at this university and they volunteered to assist with the study. The raters' scores were averaged to determine each participant's final score. Rater one was asked to rate all of the typed essays first, followed by the handwritten essays. Rater two was asked to do the opposite and rate the handwritten essays followed by the typed essays. Because Lee (2004) found that transcribed essays were rated lower than their handwritten counterparts, the handwritten essays were not transcribed. Since the raters are used to rating handwritten essays and are both experienced raters and because Powers, Fowles, Farnum, and Ramsey (1994) and Russell and Tao (2004) stated that raters tend to rate handwritten essays higher than their typed counterparts, I believed that there would not be a negative rater bias due to the appearance of the handwritten essays compared to the

typed essays. In addition, I wanted to replicate how real testing programs have handwritten essays rated: They do not transcribe them. Raters rate the original or a photocopy of the handwritten text.

### **Analysis**

I used quantitative methodology to address the three research questions. Quantitative data consists of the raw scores collected from scoring the essays with the analytic rubric as well as the averages of scores found on the Likert scale items on the questionnaire exit questionnaire. I used IBM SPSS 20 software to perform statistical analyses of the quantitative data. Trends in the qualitative data will be investigated and analyzed.

### **RQ 1 and 2:**

I addressed the first and second research questions by using paired samples  $t$  tests to investigate the raw scores given to each category on the rubric and to compare those scores between typing and handwriting.

### **RQ3:**

I addressed the second research question using a multiple logistical regression to see if the difference in the scores was correlated to the proficiency of the students. I also performed a correlation to see if there was a true effect of writing medium for the students overall and in each proficiency group.

### **RQ 4:**

I addressed the fourth research question using Likert-scale items to see which writing medium students preferred. I also used a logistical regression to see if any other

factors (students preference for typing versus handwriting or L1) could predict a student's success on handwriting compared to typing essays.

## CHAPTER 4: RESULTS

The purpose of chapter four is to present the results of the data analysis for each of the four research questions. In this chapter I will discuss each research question and present the results for each question.

Before presenting the results, I first display the reliability of the scores on the tests used in this study. In Table 2, the inter-rater reliability of the subsections of the analytic rubric are displayed, along with the overall reliability of each type of test.

Table 2: *Rater Reliability*

	Handwritten	Typed
Content	0.68	0.73
Organization	0.64	0.58
Vocabulary	0.71	0.56
Language Use	0.71	0.75
Mechanics	0.50	0.45
Total	0.74	0.78

Note: Values refer to Pearson correlation coefficients

Table 2 shows that the rater-reliability overall is fairly high. For the handwritten condition the raters agree 74% of the time while for the typed condition they agree 78% of the time. The rater-reliability among some of the individual categories, particularly the mechanics category, are low. Mechanics has the lowest rater reliability with 50% for the handwritten condition and only 45% for the typed condition.

### Research Question 1 and 2

The first research question investigated whether there were score differences between the typed and handwritten essays among the entire population. To investigate this I conducted a paired samples  $t$  test on the values of each rubric category and

compared them between the typed and handwritten condition. Table 3 shows the descriptive statistics of the data.

Table 3: *Descriptive Statistics for Writing Medium*

	Typed		Handwritten	
	Mean	SD	Mean	SD
Content	21.46	2.62	22.22	2.40
Organization	16.07	1.95	10.8	1.20
Vocabulary	12.81	1.74	13.3	1.83
Language Use	21.54	2.40	22.37	2.43
Mechanics	3.97	0.60	3.81	0.59
Total	75.69	8.76	72.50	8.11

Note: n=61

In Table 3 it can be seen that for the overall score, the typed condition is scored slightly higher than the handwritten condition. The content, vocabulary, and language use subcategories are lower for the typed condition, and the organization subcategory is scored higher than on the typed condition compared to the handwritten condition.

In order to investigate whether these means are significantly different, I preformed paired samples *t* tests for the overall scores as well as for each of the subcategories. These results can be found in Table 4.

Table 4: *Paired-Samples t-test*

Category	t-value	df	p	SE	Effect Size (r)
Content	-3.42	60	0.00 *	0.22	0.40
Organization	5.63	60	0.00 *	0.18	0.97
Vocabulary	-0.11	60	0.01 *	0.19	0.31
Language Use	-0.42	60	0.00 *	0.20	0.47
Mechanics	-0.17	60	0.84	0.08	0.03
Total	1.77	60	0.00 *	0.71	0.50

Note: \* = Significance at the .05 level

Participants overall received a higher score in the typed ( $M=75.69$ ,  $SD=8.76$ ) condition compared to the handwritten ( $M=72.50$ ,  $SD=8.11$ ) condition,  $t(60)=-1.77$ ,  $p=0.00$ ,  $r=0.50$ . The organization category was also scored significantly higher between the typed ( $M=16.07$ ,  $SD=1.95$ ) condition and the handwritten ( $M=10.80$ ,  $SD=1.20$ ) condition,  $t(60)=5.63$ ,  $p=0.00$ ,  $r=0.97$ , and this effect size was extremely large. The effect sizes are, in general, medium to large, with organization scores being largely impacted depending on the medium of essay composition: when typed, the scores were significantly and meaningfully higher (Cohen 1988, 1992).

However, participants received lower scores in the areas of content in the typing condition ( $M=21.46$ ,  $SD=2.62$ ) compared to content in the handwritten ( $M=22.22$ ,  $SD=2.40$ ) condition,  $t(60)=3.42$ ,  $p=0.00$ ,  $r=.40$ . The same occurred with vocabulary in the typing condition ( $M=12.81$ ,  $SD=1.74$ ) compared to the handwritten ( $M=13.3$ ,  $SD=1.83$ ) condition,  $t(60)=-0.11$ ,  $p=0.01$ ,  $r=0.31$ . Likewise, language use in the typed ( $M=21.54$ ,  $SD=2.40$ ) condition compared to handwritten ( $M=22.37$ ,  $SD=2.43$ ) condition was significantly lower,  $t(60)=-0.42$ ,  $p=0.00$ ,  $r=0.47$ .

The mechanics category did not show any significant difference in the typed ( $M=3.97$ ,  $SD=0.60$ ) condition compared to the handwritten ( $M=3.81$ ,  $SD=0.59$ ) condition.

In order to help understand why many of the categorical scores were lower for the typed condition compared to the handwritten condition I did word counts on each essay and compared them between the two conditions. This can be seen in Table 5.



Table 5: *Word Counts by Level*

	Typed		Handwritten	
	<b>Mean</b>	<b>SD</b>	<b>Mean</b>	<b>SD</b>
Low	244.75	79.37	249.46	64.55
Intermediate	392.00	83.93	358.59	69.17
Advanced	454.50	76.30	403.50	73.70
Total	342.30	114.28	321.50	90.00

Table 5 shows that for the Intermediate and advanced students the word counts were higher on the typed essay compared to the handwritten essay. For the low students, however, the handwritten essays had higher word counts than the typed essays.

### **Research Question 3**

The second research question investigated whether the differences in scores varied depending on the proficiency of the participants. To investigate this I used a multiple logistic regression. In this multiple logistic regression, I predicted students' benefits in typing: in other words, in this regression analysis, I used gain scores (each student's typed essay score minus his or her written-essay score) as the dependent variable. I wanted to see which of the following independent variables would be associated the gain scores between written and typed essays:

- L1 background (Chinese or Arabic only—I did not investigate other backgrounds because there were not as many participants with those backgrounds)
- The test-takers' preference for typing over handwriting (with a preference for typing scored as a one, and a preference for writing scored as zero)
- L2 proficiency (intermediate or high)

The results of this regression can be seen in Table 6.

Table 6: *Regression Factors and Their Effect*

Factor	B	Std. Error	Beta	Sig
L1 Chinese	0.89	3.15	-0.06	0.78
L1 Arabic	4.13	3.74	0.25	0.23
Preference	0.08	0.37	0.03	0.84
Intermediate	1.44	1.67	0.13	0.40
High	7.28	2.33	0.45	0.00*

Note: \* = Significance is at the .05 level

To recap, L1 Chinese, L1 Arabic, Essay Preference, Intermediate Level, and High level were used in a stepwise multiple regression analysis to predict students essay-score increase in a typed condition compared to a handwritten condition. The correlations of the variables are shown in Table 4. As can be seen, all correlations except for the one that involves being a high level student were not significant. Table 6 shows that there is no significant difference between the increase of scores by typing and the intermediate group,  $B=1.44$ ,  $SE=1.67$ ,  $P=.40$ . However, for the high level students there is a significant difference between the two test conditions and a significant gain in typing,  $B=7.28$ ,  $SE=2.33$ ,  $P=0.00$ . Students with a high level of proficiency gain around 7 points on the typed essay compared to the handwritten one.

In order to further understand these differences and to get a clearer picture of them I preformed a correlation on the scores by level as well as overall to see if there were high or low correlations. These results can be seen in Table 7.

Table 7: *Correlation of Scores*

	<b>Correlation</b>	<b>Significance</b>
Low	0.67	0.00
Intermediate	0.35	0.07
High	0.35	0.39
Total	0.79*	0.00

As would be predicted, the overall correlation is high, however when broken down by level the correlations are slightly different. For the low students the correlation remains fairly high at .67. So, for these students the writing medium makes very little difference in terms of how the students are ordered by rank. For both the intermediate and high students, however, the correlation is fairly low at .35 and .39 respectively. This shows that for both of these groups of students the writing mediums are not comparable. The low correlations indicate that test takers are ranked differently depending on the writing medium. At the two upper levels, some students receive higher scores when typing, while others receive higher scores when handwriting. The low correlations indicate these differences at the individual level. The low correlations also indicate that the two writing modes may be measuring different writing constructs. (I will revisit this notion in the discussion section.)

Figure 2: *Test Scores vs. Proficiency*

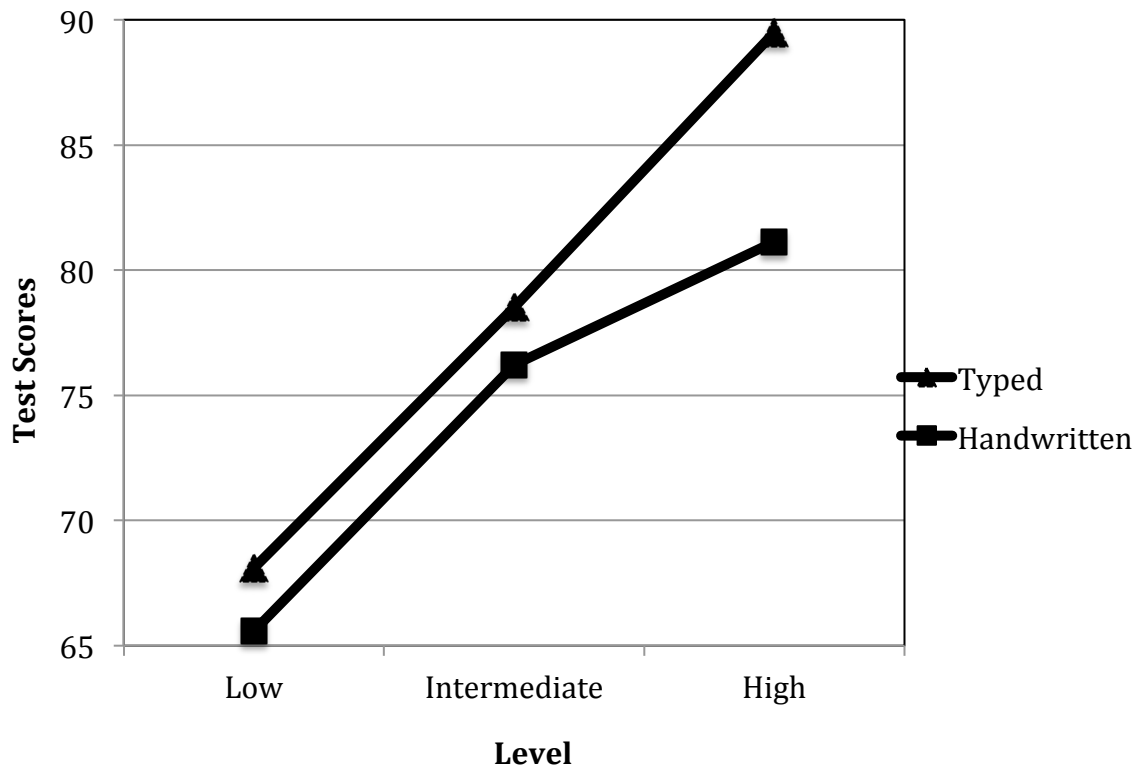


Figure 2 shows a visual representation of the mean scores by students in the two different test conditions. As was found in the regression, students' scores were not significantly different until they were at the advanced level of proficiency; that is at a higher level of proficiency, the students' scores on the handwritten test ended up being dramatically (and significantly) lower than when they typed.

#### **Research Question 4**

With the third research question I investigated how students felt about typing essays compared with how they felt about handwriting them. The means and standard deviations for the students' preferences are reported in Table 5 below.

Table 8: *Student Perceptions: Typing versus Handwriting*

	Typed		Handwritten	
	Mean	SD	Mean	SD
Low	5.67*	0.991	5.22*	1.25
Intermediate	6.19*	2.11	4.67*	1.92
High	6.88*	1.84	3.88*	1.98

Note: \*=  $p=.000$ , Measured on a Scale from 1-8

Table shows that students overall preferred to type essays rather than to handwrite them. The size of this difference becomes greater as the proficiency of the students goes up. Lower proficiency students only have a slight preference for typing while High proficiency students have a more dramatic preference for typing.

Table 8 shows the percent of students who thought their handwritten score would be scored higher than their typed score.

Table 9: *Student Perceptions: Which Essay Will Be Scored Higher*

	Handwritten	Typed
Low	40	57
Intermediate	39	61
High	43	57

As is demonstrated in table 8 more than half of the participants believed that their typed essays would be scored higher than the handwritten essays. This is true of all three of the proficiencies.

## **CHAPTER 5: DISCUSSION AND CONCLUSION**

In chapter five I examine the results presented in chapter 4 with reference to each of the research questions. I then continue with the general and pedagogical implications of this study. Finally, I conclude with a discussion of the limitations and suggestions for future research.

With the increased use of computers in the United States and around the world, it has become increasingly important to understand the effect that computer-based assessment has on writing tests. In education programs, computer processing has become the ideal way for students to compose their writing assignments, and because of this, handwritten essays have become nearly obsolete in mainstream, academic courses on college and university campuses. Although this is the case, many proficiency and placement tests, even those offered at universities, still employ the use of paper-and-pencil writing tasks to assess students' academic writing proficiency. This is because many test developers believe that having students handwrite essays eliminates a possible bias for students who are used to typing on English keyboards over students who are not. The differences in the way proficiency tests assess writing and the way writing is assessed in university classes has created the need to understand if students of varying proficiency in the language being assessed are able to write equally well with and without a computer.

### **Categorical Changes**

In this study I had 61 English-language learners compose two academic English essays: one that they typed using the computer-processing program Microsoft Word (with all Word processing features available to the writers), and one that they handwrote (with

no dictionary or outside help). In response to the first research question, which was “Do the results of a placement test differ between a paper-and-pencil based test and a computer-processed test”, I looked at whether there were significant differences in the overall test scores and in the analytic rating categories between the two different conditions. Paired sample *t* tests showed that there is an overall improvement in the typed condition compared to the handwritten condition. These results agree with Lee (2004) and Lam and Pennington (1995) who also found an increase in scores on typed essays. As Lee (2004) found, the differences in this study seem to occur mostly in relation to the organization category (as defined on the rubric), which includes the sequencing of the essay, use of main ideas and supporting ideas, and cohesion devices. This category was the only positive categorical change found in the typing condition over the handwritten condition. This finding is also supported by Whithouse, Harrison and Midyette’s (2008) study where raters commented on the poor organization in handwritten essays compared to typed essays. In sum, the higher scores on organization for the typed essays resulted in (or contributed most to) the overall higher total score on typed versus handwritten essays.

This difference could also be due to the fact that, overall, students write more in the typed condition than in the handwritten condition. Because they are writing more they may also be writing faster, and that could lead to this score drop in some of the categories, particularly language use.

This improvement in organization in the typed essays could be argued to be the result of a few different factors. First, it is easier to move text around on a computer than it is on a handwritten essay. The ability to easily move text using a computer processor may influence a student’s willingness to change his or her mind and actually move the

text. Shaw (2005) suggested this was the case when he wrote that the ability for a student to use word-processing tools might aid in the development of their essay. The second factor contributing to this increase could be the neatness of the text on a computer: it might be easier for students to visualize the organization of the essay. Also, the fact that more words fit on one typed page than on one handwritten page may also contribute to test takers' ability to *see* the organization of the essay better. This computer-aided, bird's-eye view of the essay might enable students to be able to make corrections to it more easily. Essay organization in writing is a difficult task especially for second language learners and typically requires one to write and move and rewrite before finding an acceptable organization (Whithouse, Harrison & Midyette). It seems fair to argue that the computer helps to solve this problem.

Three of the categories (content, language use, and vocabulary), had moderate, but significant, negative effects in the typed-essay condition. In other words, when the students handwrote their essays, they got moderately higher scores in the categories of content, language use, and vocabulary. This could possibly be due to rater bias as discussed by Fowels, Franum, and Ramsey (1994) and Shaw (2005). In both of these studies the researchers suggested that raters feel more sympathy for students who are handwriting essays over typing; that is, the raters feel that test takers put forth more effort in handwriting. Thus, it may be the case that that the raters were allowing for more errors in the handwritten essays than were allowing in the typed essays. The raters in this study were tasked with rating both the handwritten and typed essays (though the rating was counterbalanced; one rater rated all typed essays first, while the other rated all handwritten essays first), but these score differences may reflect a general trend of the



raters to, when possible, use a higher range of the rating scale when rating the handwritten essays. This will be discussed more in the limitations section. This is also consistent with the findings of several researchers who noted in their studies that test takers seemed to be paying less attention to content in their word-processed essays and instead were making changes at the word and sentence level. (Bridwell, Sirc, & Brooke 1985; Bridwell-Bowels, Johnson, & Brehe 1987; Collier 1983).

Out of the five categories on the rating rubric, only the mechanics category was not significantly different between the two writing mediums. Looking at Table 2, one can see that the two groups' scores on mechanics are very similar. This finding may suggest that a computer really cannot help a student with the technical aspect of essay mechanics. The student either knows where punctuation belongs, knows how to create different sentence structures, spell, and capitalize or doesn't. This finding is slightly contrary to, Whithouse Harrison, and Midyette's (2008) results. They wrote that mechanical errors seemed to jump off the page in the typed essays and were not as noticeable in handwritten essays, thus the mechanics category was thought to be scored lower when the essays were typed in their study. This finding also contradicts the claim by Whithouse, Harrison, and Midyette (2004) that students seemed to lose the ability to proofread when essays were typed. In their study, raters commented that the mechanics of students' essays were worse on the typed essays compared to the handwritten ones. However, in this study, I found that these categories were similar between the two different writing mediums, and thus this claim does not hold when considering the present study. Another interpretation of the results could be that in fact the students' mechanics were worse in the handwritten condition, but the raters were biased toward giving the handwritten

essays a higher score on mechanics. This too will be discussed more in the limitations section.

### **The Effects of Proficiency, L1, and Preference**

Using simple regression, I looked at whether English language proficiency, L1 background, and test takers' preference for handwriting or typing essay-test exams were associated with the overall scores obtained on the essay tests. In particular, for each individual, I calculated his or her gain score in typing over handwriting: that is, I subtracted each test taker's score on the typed essay minus his or her score on the handwritten essay (and all the subcategories represented by the different categories on the analytic rubric) to see if proficiency, L1, or their preference predicted *gains in typing*. The test takers' native language and essay preference did not have a significant effect on test score gains. L1 Chinese test takers did not do better than any other L1 background when typing essays, and L1 Arabic speakers did not do any better on typed essays than any other L1 background. I did not investigate the effect of the other language backgrounds on gains in typing over handwriting because of the low number of participants with those language backgrounds. But overall L1 background did not contribute to ability to type over handwrite. Likewise, test takers' preference for typing was not associated with higher scores on the typed versus handwritten essay tests, which I believe is a very significant finding in this research.

Past research has shown that when students are allowed to pick, before writing, whether to handwrite or type, lower proficiency students do better on the handwriting and higher students do better on typing (Wolfe, Bolton, Feltovich, & Niday 1996). As discussed in the literature review, this finding is interesting but problematic, in that

having students *chose* their writing medium results in a study with non-randomized group assignments. Wolfe et al. were not able to attribute the differences in the test scores within the groups to (a) the test-taking condition or (b) the choice that was made by the test takers. In this study I found that student preference for one type of essay over the other type of essay does not predict their ability to type or handwrite an essay, and even the lower level students did better on the typed test. This finding may suggest that students cannot accurately pick the writing medium in which they will best perform. This could be important for testing programs that offer test takers the option of handwriting or typing an essay test. If a student cannot accurately chose whether they are better at typing or handwriting, then this could be a problem if they are offered an option in a writing test and they chose, for comfort reasons, the option that results in lower test scores. It could be that a student *feels* they will write better by hand when this is not the case, or they may believe that raters prefer typed essays, even if this is not true. Such decisions could lead to a student not scoring as well on a test than he or she might otherwise.

Finally, in this study I found that the proficiency of a student does have an effect on the scores the students receive on typed essays compared to handwritten essays. For the low and intermediate students there is no significant difference between scores on the typed essay compared to the handwritten essay. In both cases, the students performed slightly better on the typed essays. However, advanced students received a significantly higher score on the typed test than the handwritten test. They also preferred typing over handwriting giving it on average a score of 6.88 out of 8 on the Likert scale whereas they gave handwriting a score of 3.88 out of 8. Advanced students scored an average of seven

points higher on the typed essay than on the handwritten essay. This is a significant and large score difference that can cause a student to be placed into an English language class merely because that student was required to handwrite rather than type the placement test essay.

There are several possible reasons for this difference in scores for advanced students. The first is that advanced students, those with bachelor's degrees, may be used to composing academic essays on a computer. Many universities around the world require students to compose essays on computers, and these students may have much more experience typing over handwriting. Also, because of general college and university typing requirements, students may have stopped handwriting assignments long ago, and thus might be very uncomfortable handwriting an essay.

Second, students, especially at this level, who may be quite comfortable with composing essays on computers, may also know how to use the functions available in word processors such as cut, copy, and paste. They may take advantage of spelling and grammatical checking features. They may rely on these functions for organizing and planning their essays and for checking their syntax, spelling, and grammar. Taking these functions away could make it much more difficult for students to compose essays.

Writing by hand has been said to make it difficult to move and change things in an essay (Lee, 2004). Research has shown that when students have to hand write essays, they have to actually plan the organization before writing the essay. If students are used to computer-processing an essay and not worrying about the essay's organization until after the essay is composed, then may be difficult for the students to pre-plan when handwriting. Any lack of knowing how to strategy shift in organizing the essay

depending on the writing medium could result in organization-score differences, and as this study suggests, a poorly organized handwritten essay. Powers, Fowles, Farnum and Ramsey (1994) and Russell and Tao (2004) noted that handwritten organizational edits do not have an effect on raters, but studies also have found that students think that they will (Lee 2004). Thus, when handwriting, students might be reluctant to make organizational changes after the entirety of the essay is composed. But when computer processing, writers may be more apt to make organizational changes after the initial essay is drafted.

The correlations that I preformed further support these findings. The correlations showed that for students in the intermediate and high groups, the writing medium does make a difference. For many of the upper-proficiency-level students, their writing score changed depending on the writing medium they used. The low correlations within the upper-level groups (between their typed and handwritten essay scores) indicate that even though on average students receive the same scores, individuals within the groups perform differently depending on the writing medium. This is further proof that typing may be better for these students than handwriting because typing matches the academic mode of writing being assessed. For the low students, there is no difference in scores and the scores correlate, so it does not matter which writing medium they use as they will score about the same.

### **Student Perceptions**

Regarding the third research question, none of the three proficiency groups (low, intermediate, or advanced) preferred handwriting over typing. Of course, this is a generalization as individual differences among students were not considered in the

current study. On average, students at all levels have a preference for typing over handwriting. However, at the low level, this preference is slight, whereas at the advanced level, this preference is much larger. As was expected, the higher the level of the student, the larger the preference for typing over handwriting.

When asked which essay will be scored higher, students' responses showed that for all three levels, over half of the participants believed that the typed essay would be scored higher than the handwritten essay. This is most likely due to the fact that, as Withause, Harission and Midyette (2008) found, students tend to believe that typed essays are more legible than handwritten essays and thus will receive a higher score than their handwritten counterparts. This sheds light on what may happen when programs allow students to choose to handwrite or type on academic essay tests: some test takers may choose based on their true preference (as the test developers intend), while others may choose based on what they think raters expect or will like better (*not* as the test developers intend). Thus, in some cases, the *choice* may introduce underestimations of student performance. In the worst-case scenarios, students may perform poorer than expected (get lower scores than they should) because they chose one medium over the other, with each student making that choice based on different (apparently random) reasons.

### **Implications**

Concerning research on score differences between handwritten and typed essays, this study has several implications. Proficiency seems to play the biggest role in determining whether typing an essay can help or harm students. For the low and intermediate levels there is no significant difference in the final essays scores. However,

for the most advanced students, handwriting actually has a harmful effect. This means that a student who is very advanced and forced to handwrite an essay may end up with a lower score than they would if they typed the essay. The score could be so much lower that they may have to, in certain places, take English language classes because of it and could lose time, scholarships, and admittance to a university. It seems that switching from a handwritten proficiency test to a typed test would help students more than hurt students. This is even true at the lower level. This study does not look at novice or true beginners of the language being assessed; however, it is important to keep in mind that those levels typically have separate writing tests because a general test of academic-essay-writing can typically not tease apart students at the lowest levels of language proficiency. Indeed, when considering very low-level language learners, giving them a test to measure their ability to perform *academically* in the language would seem absurd; for example, they should be given a general English-writing test, not an academic-English-essay writing test.

Also of importance are the results found on the preferences of the students. English language programs everywhere tend to have some problems with students believing they were unfairly tested and placed into the incorrect level or class. In this study I found that the majority of students believed that a typed essay test will place them correctly. Thus, it may be beneficial for such programs to switch to a typed test and thereby eliminate students' blame on handwriting as the cause of any misplacement.

The results of this paper point to the benefits of having test takers type their essays for academic-essay tests. This finding is especially helpful because having students type in such testing situations would better match the actual, real-world tasks of

academic writing. Bachman and Palmer (2010) and others (Douglas, 2000; Lewkowicz, 2000) stressed that tests must represent (as well as possible) how the skills being tested are utilized in the real world. In academic settings all over the world, individuals type their written work and computer-processes their academic essays.

### **Limitations**

Although the data in this study is very informative, there are still a few limitations to this study in both the population used and the study itself.

The first limitation to this study was the final number of the most advanced students used in the study. Because this is where the largest difference was found, it would be useful to have more students of this level in order to better understand the effect of the test condition and the test-taking behaviors of these students.

The second limitation is the participants themselves. Because the program in which I conducted this study was heavily populated by Chinese and Arabic speaking students, it was difficult to get students of varying backgrounds and L1s. Not only was there a lack of cultural diversity, there was also a lack in age diversity. Most students were around 18 or 19 years old. While this is not likely to have a large impact on the results, it would still be an interesting area to look for an effect in both age and native language.

Another limitation could be the effect of spell checkers on the essays of the students who are familiar with computers. Due to the use of Microsoft word, there was no way that this could be controlled. This could offer a slight advantage to those students over the ones who are not as familiar with composing essays on computers. I believe that this would not actually affect the overall scores, but in hindsight I could have better



controlled this factor by asking questions that tapped into students' knowledge of MS Word computer-processing tools.

Finally, a limitation is that I did not interview raters about the way they rated the essays. This is a limitation because it would have offered many insights into the differences in the way raters rated the essays and exactly what they thought about while rating each type of essay. In a future study this could be done through think alouds with each of the raters.

### **Directions for Future Research**

Finally, there is additional research in this area that could be done to help further understand this area of research. It could be helpful to expand this study to include the lowest-level students offered in the program. Adding this population would show whether it would be better to offer a dual option test or if it would be appropriate to completely switch to a typed format.

Second, it would be helpful to have a larger number of the highest-level students to get a clearer picture of what is happening with students when they type compared to when they handwrite. This would provide further support for the current study and help to further inform university testing centers on which type of test is the best to use.

Third, it would also potentially be interesting to look at students with other, different native languages (other than Arabic and Chinese) and see if there is any difference in their scores depending on the language background they come from. Data from such populations may help to inform whether an optional typing and computer-processing class should be offered to language learners in programs that promote academic-language and academic-skills development.

A fourth possibility would be to look more closely at rater differences in rating typed compared to handwritten essays. While Powers, Fowles, Farnum, and Ramsey (1994) and Russell and Tao (2004) wrote that raters tend to rate handwritten essays higher than their typed counterparts, this may not be the case in a real rating situation. In both of these studies the raters were trained to rate the specific essays they were given. Lee (2004) discussed the possibility that transcribed essays are usually scored lower than their handwritten counterparts. It would be interesting to look at how raters actually perceive these different types of essays, possibly through using think alouds to understand further what raters are thinking and to see how different the ratings actually are when essays are typed compared to handwritten.

Another possibility for future study would be to look at the specific differences that occur for students when they are typing compared to writing. It would be interesting to see the changes that they make while they are typing on a computer as well as the changes they make when handwriting an essay. This would show further what the qualitative differences are for students. Such a study would allow researchers to understand more concerning any differences between the handwritten and typed conditions.

It would also be interesting to, in a future study, to give an objective measure to mechanics. This could be done through coding, grading, and counting the mechanical error in the essays in order to see how well the raters were doing. It could also help researchers to know more about rater bias and where it comes from.

Finally, it would be interesting to look more closely at students' preferences and the motivations behind any choices they make when they have the ability to decide

whether to type or handwrite academic essays. This would allow teachers and researchers to better understand how students feel about these tests and to further understand how they view a typed test compared to a handwritten test. Understanding the students better may help programs not only to explain to students why tests are done the way they are, but it could also help pedagogically. If students are making assumptions that are not true, teachers could help them to understand better how they should approach, not just tests, but also all academic writing tasks.

## APPENDICES

## **Appendix A:**

### **Essay Prompts**

#### **TEST A** **English Language Test** **Timed Writing Exam**

**Name** \_\_\_\_\_ **Test Number** \_\_\_\_\_

Write as much as you can, as well as you can, in an original, 35-minute composition on the topics below.

Some people go to college directly after high school, while others take a job after high school and attend college a few years later. Which do you think is better, and why? Be sure to support your ideas with specific explanations and details.

#### **TEST B** **English Language Test** **Timed Writing Exam**

**Name** \_\_\_\_\_ **Test Number** \_\_\_\_\_

Write as much as you can, as well as you can, in an original, 35-minute composition on the topics below.

Some people get married directly after high school or in their early 20s, while others get married later, in their 30s, 40s, or even later. Which do you think is better (early marriage or late marriage), and why? Be sure to support your ideas with specific explanations and details.

## Appendix B:

### Background Questionnaire

#### BACKGROUND QUESTIONNAIRE

##### *ELT Essay Writing Project*

*PLEASE FILL OUT THE FOLLOWING BACKGROUND INFORMATION. PLEASE PRINT CLEARLY.*

1. Name: a. First name: \_\_\_\_\_

c. Middle initial: \_\_\_\_\_

b. Last name: \_\_\_\_\_

2. Age: \_\_\_\_\_

3. Gender: ☐ Male ☐ Female

4. Phone number: ( ) \_\_\_\_\_ - \_\_\_\_\_

5. Email address: \_\_\_\_\_

7. Native language (first fluent language, also known as your “mother tongue”):

\_\_\_\_\_

a. How did you learn English?

\_\_\_\_\_  
\_\_\_\_\_

b. How old were you when you started learning English?

\_\_\_\_\_

8. How long have you studied at the ELC?

\_\_\_\_\_

9. Did you attend University in your home country?

\_\_\_\_\_

10. What is your current ELC level?

\_\_\_\_\_

## Appendix C

## Exit Questionnaire

Exit Questionnaire: Please answer the following questions to the best of your ability based on your test-taking experience.

1. How much did you like **typing** your essay on a computer?

1 2 3 4 5 6 7 8

I didn't like it I liked it very much

Please Explain.

2. How much did you like **handwriting** your essay on paper?

1 2 3 4 5 6 7 8

I didn't like it I liked it very much

Please Explain.

3. 2. How much did you like answering the question about **marriage**?

1	2	3	4	5	6	7	8
I didn't like it				I liked it very much			

Please Explain.

4. How much did you like answering the question about **College**?

1	2	3	4	5	6	7	8
I didn't like it				I liked it very much			

Please Explain.

1      2      3      4      5      6      7      8

Not familiar at all      Very Familiar

- The one I handwrote on paper
- The one I typed on a computer

- The one I handwrote on paper
- The one I typed on a computer

8. Do you feel that you were placed into the correct level at the ELC when you first came?

Please Explain



## Appendix D:

Table 10: *Rubric*

Content		Organization		Vocabulary		Language Use		Mechanics	
<b><i>Clear Competence for Academic Study</i></b>									
30 to 27	<ul style="list-style-type: none"><li>• Main ideas and support are clear, precise, and relevant</li><li>• Thorough development of thesis</li><li>• Addresses the prompt</li></ul>	15 to 13	<ul style="list-style-type: none"><li>• Succinct, logical sequencing</li><li>• Clear differentiation between main ideas and support</li><li>• Excellent internal cohesion through sophisticated cohesive devices</li></ul>	20 to 17	<ul style="list-style-type: none"><li>• Sophisticated range</li><li>• Precise word/idiom choice and usage, word form mastery</li><li>• Appropriate register</li></ul>	30 to 27	<ul style="list-style-type: none"><li>• Mastery of simple and complex constructions</li><li>• Virtually no global errors</li><li>• Few minor grammatical errors</li><li>• Meaning is clear and precise</li></ul>	5	<ul style="list-style-type: none"><li>• Demonstrates mastery of conventions</li><li>• Few errors of spelling, punctuation, capitalization, or paragraphing</li><li>• Meaning is clear</li></ul>
<b><i>Developing to Sufficient Competence for Academic Study (shaded area and above meet and exceed MSU minimum requirements)</i></b>									
26 to 25	<ul style="list-style-type: none"><li>• Main ideas and support are generally clear and relevant</li><li>• Sufficient development of thesis</li><li>• Addresses the prompt</li></ul>	12 to 11	<ul style="list-style-type: none"><li>• Generally clear organizational structure</li><li>• Main ideas stand out</li><li>• Somewhat limited or superficial internal cohesion;</li></ul>	16 to 15	<ul style="list-style-type: none"><li>• Good range of higher level vocabulary</li><li>• Generally effective word / idiom choice and usage, despite the occasional error in word choice and word form</li><li>• Meaning is generally clear and requires no reader compensation</li></ul>	26 to 25	<ul style="list-style-type: none"><li>• Strong and consistent control of simple constructions</li><li>• Generally effective control of complex constructions</li><li>• Few global errors</li><li>• Occasional local errors</li><li>• Meaning is generally clear and requires no reader compensation</li></ul>	4	<ul style="list-style-type: none"><li>• Demonstrates strong control of conventions</li><li>• Occasional errors of spelling, punctuation, capitalization, or paragraphing</li><li>• Meaning is clear</li></ul>

Table 10 (cont'd)

24 to 22	<ul style="list-style-type: none"> <li>• Main ideas are generally clear</li> <li>• Support ideas are mostly clear and relevant</li> <li>• Generally adequate development of thesis, but support may be somewhat limited, superficial, or repetitive at times</li> <li>• Addresses the prompt</li> </ul>		possibly repetitious or awkward use of cohesive devices, over-reliance on simplistic transitions; somewhat choppy	14 to 13	<ul style="list-style-type: none"> <li>• Adequate range of higher level vocabulary</li> <li>• Occasional errors of word/idiom form/choice</li> <li>• Meaning is generally not obscured or may require only slight reader compensation</li> </ul>	24 to 22	<ul style="list-style-type: none"> <li>• Strong control of simple constructions</li> <li>• Inconsistent control of complex constructions</li> <li>• Global and local errors not infrequent</li> <li>• Meaning is generally not obscured or may require only slight reader compensation</li> </ul>		
<b><i>Suggests Insufficient Competence for Academic Study</i></b>									
21 to 19	<ul style="list-style-type: none"> <li>• Main ideas generally clear</li> <li>• Supporting ideas may be somewhat obscured</li> <li>• Development is generally limited, superficial, or repetitive</li> <li>• Related to the prompt, but may be slightly off-topic</li> </ul>	10 to 8	<ul style="list-style-type: none"> <li>• <b>Somewhat unclear organizational structure</b></li> <li>• Ideas seem disconnected</li> <li>• <b>Very limited or ineffective use of cohesive devices</b></li> <li>• Lacks logical sequencing</li> <li>• Limited sample;</li> </ul>	12 to 10	<ul style="list-style-type: none"> <li>• Limited range: (i.e., repetition of a small number of commonly used words, rare use of words from the AWL)</li> <li>• Frequent or distracting errors of word/idiom form/choice</li> <li>• Meaning confused or obscured and requires significant</li> </ul>	21 to 19	<ul style="list-style-type: none"> <li>• Inconsistent control of simple constructions</li> <li>• Lack of control or void of a variety of complex constructions</li> <li>• Frequent global &amp; local errors</li> <li>• Meaning may be somewhat obscured but not unintelligible, requires some reader compensation</li> </ul>	3	<ul style="list-style-type: none"> <li>• Demonstrates inconsistent control of conventions</li> <li>• Frequent or distracting errors of spelling, punctuation, capitalization, or paragraphing</li> <li>• Meaning may be confused or obscured</li> </ul>

Table 10 (cont'd)

18 to 17	<ul style="list-style-type: none"><li>• Main ideas and/or supporting ideas somewhat obscured</li><li>• Development is very limited, superficial, or repetitive</li><li>• Relationship to the prompt may be vague but discernable.</li></ul>		does not demonstrate significant organizational features		reader compensation	18 to 17	<ul style="list-style-type: none"><li>• Weak control of simple constructions</li><li>• Generally ineffective complex constructions or repetition of only a few formulaic complex constructions</li><li>• Frequent global and local errors</li><li>• Meaning is often obscured; requires significant reader compensation</li></ul>		
<b><i>Clear Lack of Competence for Academic Study</i></b>									
16 to 13	<ul style="list-style-type: none"><li>• Main ideas and/or supporting ideas generally obscured/confusing</li><li>• Minimal development of thesis</li><li>• May be off-topic</li><li>AND/OR</li><li>• Not enough to evaluate</li></ul>	7 to 6	<ul style="list-style-type: none"><li>• organizational structure very unclear and/or confusing</li><li>AND/OR</li><li>• not enough to evaluate</li></ul>	9 to 7	<ul style="list-style-type: none"><li>• Very limited range; repetition of a small number of words</li><li>• Frequent errors of word / idiom / form /choice</li><li>• Meaning may be <b>unintelligible</b></li><li>AND/OR</li><li>• <b>not enough to evaluate</b></li></ul>	16 to 13	<ul style="list-style-type: none"><li>• No control over basic sentence construction</li><li>• Dominated by global and local errors</li><li>• Meaning is often unintelligible</li><li>AND/OR</li><li>• not enough to evaluate</li></ul>	2	<ul style="list-style-type: none"><li>• Demonstrates lack of control of conventions</li><li>• Dominated by errors of spelling, punctuation, capitalization, &amp;/or paragraphing</li><li>• Meaning is confused or obscured</li><li>AND/OR</li><li>• not enough to evaluate</li></ul>

## REFERENCES

## REFERENCES

- Bachman, L. F. (1990). *Fundamental considerations in language testing*. Oxford: Oxford University Press.
- Bachman, L. F., & Palmer, A. S. (2010). *Language assessment in practice, revised edition* (2nd ed.). Oxford: Oxford University Press.
- Benesch, S. (1987). *Word processing in English as a second language: A case study of three non-native college students*. (Available ERIC: ED281383.)
- Bridwell-Bowles, L., Johnson, P., & Brehe, S. (1987). Composing and computers: Case studies of experienced writers. In: A. Matsushashi (Ed.), *Writing in real time: Modeling composing processes* (pp. 81–107). Norwood, NJ: Ablex.
- Bridwell, L.S., Sirc, G., & Brooke, R. (1985). Revising and computing: Case studies of student writers. In S. Freedman (Ed.), *The acquisition of written language: Revision and response*. Norwood, NJ: Ablex.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). New York: Academic Press.
- Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112(1), 155-159.
- Collier, R. (1983). The word processor and revision strategies. *College Composition and Communication*, 34, 149–155.
- Haas, C. (1989). *How the writing medium shapes the writing process: Effects of word processing on planning*. (Available ERIC: EJ388 596.)
- Daiute, C. (1985). *Writing & computers*. Addison-Wesley.
- Douglas, D. (2000). *Assessing languages for specific purposes*. Cambridge: Cambridge University Press.
- Gebril, A., (2009) Score generalizability of academic writing tasks: Does one test method fit it all? *Language Testing*, 26, 507-531
- Gerbril, A., & Plankans, L. (2009) Investigating source use, discourse features, and process in integrated writing tests. *Spain Fellow Working Papers in Second or Foreign Language Assessment*, 7, 47-84.
- Hamp-Lyons, L., & Kroll, B. (1996). Issues in ESL writing assessment: An overview. *College ESL*, 6 (1), 52–72.
- Harrington, S. (2000). The influence of word processing on English placement test. *Computers and Compositions*, 17, 197–210.

- Hays J.R. (1996). A new framework for understanding cognition and affect in writing, in C.M. Levy et S. Ransdell (Edit.), *The science of writing*, Mahwah (NJ), Erlbaum, 1-27
- King, F. J., Rohani, F., Sanfilippo, C., White, N. (2008). Effects of handwritten versus computer-written modes of communication on the quality of student essays. *CALA Report*, 208. Available at [http://www.cala.fsu.edu/files/writing\\_modes.pdf](http://www.cala.fsu.edu/files/writing_modes.pdf)
- Knoch, U. (2009). Diagnostic assessment of writing: A comparison of two rating scales. *Language Testing*, 26(2), 275-304.
- Lam, F. S., Pennington, M. C. (1995). The computer vs. the pen: A comparative study of word processing in a Hone Kong secondary classroom. *Computer-Assisted Language Learning*, 7, 75–92.
- 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.
- Lewkowicz, J. A. (2000). Authenticity in language testing: Some outstanding questions. *Language Testing*, 17, 43–64.
- Li, J., (2005). The Mediation of Technology in ESL writings and its implications for writing assessment. *Assessing Writing*, 11, 5–21.
- Plankans, L. (2010) Independent vs. integrated writing tasks: A comparison of task representation. *TESOL Quarterly*, 44, 185-194
- Powers, D. E., Fowles, M. E., Farnum, M., & Ramsey, P. (1994). Will they think less of my handwritten essay if others word process theirs? Effects on essay scores of intermingling handwritten and word-processed essays. *Journal of Educational Measurement*, 31 (3), 220-233.
- Russell, M. & Tao, W. (2004). The influence of computer-print on rater scores. *Practical Assessment, Research and Evaluation*, 9 (10), 1-14.
- Schwartz, H., Fitzpatrik, C., & Huot, B. (1994). The computer medium in writing for discovery. *Computers and Composition*, 11, 137–149.
- Schoonen, R. (2005). Generalizability of writing scores: An application of structural equation modeling. *Language Testing*, 22 (1), 1–30.
- Shaw, S. (2005). Evaluating the impact of word processed text on writing quality and rater behaviour. *Research Notes*, 22, 13-19.
- Susser, B. (1994). Process approaches in ESL/EFL writing instruction. *Journal of Second*

Language Writing, 3, 31–47.

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

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

Whithaus, C., Harrison, S.B., Midyette, J. (2008). Keyboarding compared with handwriting on a high-stakes writing assessment: Student choice of composing medium, raters' perceptions, and text quality. *Assessing Writing*, 13, 4-25.

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. *Assessing*

Wolfe, E. W., & Manalo, J. R. (2004). Composition medium comparability in a direct writing assessment of non-native English speakers. *Language Learning & Technology*, 8(1), 53–65.