A Thesis

entitled

Analysis of the Composition I-ESL Course

by

Tedrick R. Klingbeil

Submitted to the Graduate Faculty as partial fulfillment of the
requirements for the Master of Arts Degree in English

Dr. Melinda Reichelt, Committee Chair

Dr. Anthony Edgington, Committee Member

Dr. Barbara Schneider, Committee Member

Dr. Patricia R. Komuniecki, Dean
College of Graduate Studies

The University of Toledo

May, 2012
An Abstract of

Analysis of the Composition I-ESL Course

by

Tedrick R. Klingbeil

Submitted to the Graduate Faculty in partial fulfillment of the Requirements for the Master of Arts Degree in English

The University of Toledo

March 2012

This thesis investigates the Composition I-ESL writing course that is required for most non-native English speaking international students enrolled at The University of Toledo. It sought to determine if any progress had been made through the course of a semester in students’ writing, specifically in the areas of summarizing, paraphrasing and quoting. To evaluate this, a pre-test was given at the beginning of the fall, 2011 semester and a post-test was given at the end. Both tests were the same, and students were asked to write an essay in response to five written excerpts. Fifty-eight students participated in both tests. These tests were rated by trained raters and a statistical measurement of the results showed a measurable improvement in the area of summarizing, paraphrasing and quoting.
For all those who helped me in this program, I am truly grateful. There were many instances where colleagues and faculty went out of their way to aid me in this project and many others. Most of all, I would like to thank God, for his marvelous ways and especially to my parents who were most supportive throughout the process.
Acknowledgements

I would like to extend my most heartfelt gratitude to Dr. Melinda Reichelt, my thesis chairperson as well as my wonderful committee Dr. Barbara Schneider and Dr. Anthony Edgington, who graciously took on this task and gave me very timely response to my drafts. I would also like to thank all those who served as raters for this project. They gave up their Friday afternoon to score the pre and post tests upon which this thesis is based.
# Table of Contents

Abstract iii

Acknowledgments v

Table of Contents vi

I. Introduction and review of the literature 1
   A. Introduction 1
      a. The Current Project

   B. Review of the Literature 4
      b. Test-retest Design studies
      c. Evaluating some Psychological Features Related to the

   C. The Testing Process 12
      a. Pitfalls in Testing 19
      b. Placement Tests With Writing Components 23

II. Methodology 25
   A. Research Questions 25
   B. Context 25
   C. Participants 26
   D. Procedures 27
   E. Qualitative and Quantitative Analysis of Summarizing, Paraphrasing and Quoting 31

III. Results of the Study 38
   A. Findings 40
   B. Discussion 42
   C. Implications of Research 43
D. Limitations of Research  

E. Suggestions for Future Research  

F. Conclusion  

G. References
Chapter I

Introduction and review of the literature

Introduction

The University of Toledo (UT) has long cultivated relationships with universities overseas, predominantly in Asia, and has formed many cooperative ventures with them. This has resulted in numerous exchange programs and has bolstered the international student population attending the university. In 2003, there were 908 international students studying at UT in undergraduate and postgraduate programs. This number has risen steadily since then and as of the writing of this thesis, there are approximately 1,750 international students at UT (Thomas, P, personal communication, January, 5, 2012). Although this number accounts for only about 7% of total student population, it is a sizeable number and has led the university to structure programs to accommodate and encourage them in their linguistic development. The most prominent program for international students on campus is the American Language Institute (ALI), which is an intensive language course to help students enter the mainstream university if their TOEFL scores are below 450. When the students can obtain a score of 450 or higher, they are admitted into the mainstream university. Additionally, and of concern to this study, the university operates sheltered composition courses for international students. Baine (1996) found that international students who were placed in sheltered ESL composition programs were less likely to fail or withdraw and reported feeling more comfortable participating in class activities as well as asking questions.

There are two sheltered composition courses for ESL students at UT. They consist of ENGL 1020, an introductory course for students whose composition skills, as tested by
a written exam, have enabled them to enter the mainstream university, but are not high enough for ESL ENGL 1110, the ESL equivalent of Composition 1. Students must remain in ENGL 1020 until their proficiency reaches a high enough level, then they may register for English composition 1110 for international students, hereafter called ESL Comp I.

English Composition 1110, hereafter called ESL Comp I, will be the focus of this study. It is taught by first and second year graduate student teaching assistants and is exclusively for ESL students. Although it is not formally listed as such, it functions as an English for academic purposes (EAP) course. As such, it strives for authenticity, and seeks to give students the tools they will need in pursuing their future studies. The director of the course described it as “a forced march towards the completion of a research paper” (Reichelt, 2010). The guiding theory is that students will need to have strong WID (writing in the disciplines) skills, and proficiency in writing research papers is one of the most important WID skills (Hill et al., 1982). This research paper project, upon which the course is based, consists of several steps the students must undertake. They are as follows:

- A personal interview with a professor.
  - Students should ask the professor what s/he considers to be a controversial topic in their field of study. This helps students select a topic for their research paper.
- Research Proposal paper.
- Response essay.
  - This is to teach students how summarize and respond to a scholarly article.
At least four casebooks entries.

- Students read a scholarly source and summarize it in about half a page. Then they reflect on it for half a page.

- One or more preliminary drafts of the research paper.

- A final draft of the research paper.

This is the primary content of the course, but teachers have some discretion to bring in other materials as time permits. Time usually doesn’t permit very much divergence from this curriculum however. This is because the course entails two 75 minute in-class contact sessions per week, over a 16 week semester, so time management is essential. Additionally, teachers keep office hours for teacher-student conferences. Twice a semester classes are cancelled and special conferences with students are scheduled during critical periods of the project. A writing center also supports the program, with dedicated ESL tutors.

**The Current Project**

This thesis has two purposes. First, it seeks to evaluate the effectiveness of the 1110 course, given to undergraduate students, and to extend knowledge on the subject of how to analyze writing courses within a university, in the United States. The ESL Composition I course under study will be evaluated based on the ability of the students enrolled in it to summarize, paraphrase and quote.

This study will loosely replicate one done by Storch (2009), which analyzed the improvement in writing proficiency of 25 international students who had studied at the University of Melbourne for a semester. Storch utilized a test-retest design and analyzed results qualitatively (structure and content) and quantitatively (fluency, accuracy,
grammatical complexity and lexical complexity). Like Storch’s study, this one will the student’s ability to summarize, paraphrase and quote. A sample of about 60 subjects will be studied in this project. All of them will be undergraduates in the 1110 course, who have been in the United States for differing lengths of time. Both genders will be represented and students will be studying multiple disciplines.

**Review of the Literature**

Systematic analysis of results from ESL/EFL programs, and language programs in general is a relatively new practice and there is not a great deal of empirical research, or assessment of effectiveness, on the subject. It seems that most schools and institutions simply assume that their programs are effective and leave it at that. Some programs, such as the University of Toledo’s ALI program use the TOEFL test as a benchmark of their program’s success. Students may graduate from the ALI and move into regular university classes if they can gain a TOEFL score of 450 (Sayers, B. personal communication, March 14, 2011). Beginning in the late 1990s, researchers have begun to take up the topic and have done some empirical studies, collecting primary data from students enrolled in various language programs at different institutions. The reader should note that all of the literature reviewed here is related in some way to testing and evaluation of second language (L2) student writing with emphasis given to ways of tracking student improvement over time.

**Test/Retest Design Studies**

As previously noted, this thesis seeks to extend knowledge about how writing courses for non-native speakers can be analyzed and improved. For the UT Composition I evaluation tool under consideration in this thesis, a test, re-test design was selected
because it was determined this type of test would provide the best tool for accurate
analysis. The test-retest design is ideal for this project because it allows the progress of
the students to be tracked throughout the semester. All of the studies reviewed in this
section are relevant because they use a test-retest design, and serve to inform the
methodologies employed in this thesis.

Storch (2009), the first study under consideration, is most important because this
thesis loosely uses it as a replication model. It is by Neomy Storch, one of the leading
researchers in the area of ESL writing program analysis, and it measures English writing
improvement, fostered in an English medium environment. It differs from the UT
environment in that students involved were not taking a formalized writing course. It was
chosen as a replication model because it made a logical, authoritative study of a writing
program. Of most interest in this study, and an additional reason it was chosen as a
replication model, is that it included, among other things, an analysis of the student’s
ability to summarize, paraphrase and quote. To do so, it looked at the results of a
semester’s study at the University of Melbourne on the writing of some ESL students
from Asia. Storch used a test, retest design administered at the beginning and end of the
semester. To select the subjects, she used results from the non-mandatory English
diagnostic test that was administered to incoming international students to determine their
English ability. This test, called the Diagnostic Language Assessment, was simply used
to identify students who could benefit from further language formation. It consisted of
three subsets: reading, writing and listening. For each subset students were given a score
of 1-9, in which 9 represented the highest proficiency level.
Based on this test, 25 participants were selected for the study. They were selected because, despite their need for it, they had sought only minimal formal ESL writing support. Most of them were post-graduate students, (17 female, 8 male, all from Asia) who had not sought help with English, nor attended ESL courses (on this campus). They also had a rather narrow range of proficiency, all being classed as intermediate to upper intermediate, with IELTS scores of 6.5-7. These students were asked to write an argumentative essay of at least three hundred words on the topic of animal rights. They were given five minutes to read six short excerpts, five minutes to plan and 45 minutes to write. Analysis focused on communicative fluency, content, form, organization of ideas, number of developed ideas, language used (grammar and lexis), fluency, accuracy, lexical complexity, coherence and cohesion. Since the students in this study are not post-graduate students and most likely have lower-level writing skills, the focus of this study will be narrowed and only evaluate the global aspects of the writing of students involved.

Keeping in mind that this literature review is focused on the testing of L2 student writing using pre and post tests, we now turn to the another study that uses this methodology. This study Storch and Hill (2008), is one which analyzed the effect integration by international students into the mainstream, English speaking Australian community had on student English proficiency. This study used a pre and post-test at the beginning and end of a semester to measure the writing, listening, speaking skills as well as social confidence developed by non-native-speaking students entering the University of Melbourne, Only the part about writing will be discussed here, however.

Students typically scored higher on the diagnostic exam, including the written part, the second time around in a manner that was statistically significant. The higher the
initial score the lower the second score, showing some plateauing. Analysis of the results showed that graduate students did better than undergraduates. Although this study was only partially about writing, it was included because it extends knowledge about a testing scenario in which a test-retest design was employed. The way survey data was included in this study will also help to serve as a guide for the use of survey data in this project because it includes Likert scale-type data as well as general questions and data about student attitudes.

Storch (2007) is another relevant study to continue the discussion of the analysis of ESL student writing ability gains, by having them take a pre and post test. It is similar to the last one, but it also extends knowledge of test-retests design testing models by including information about the use of surveys. This study represents a further opportunity for Storch to evaluate the effectiveness of being in an English-medium university on non-native speaking students at the University of Melbourne. To this end, she designed a study to measure its results using a sample of 25 students. These were students who the researchers suggest should perhaps have gotten additional support on their writing in previous semesters, but mostly didn’t do so. A few of the subjects did get limited support, but it amounted to less than 5 hours. All of these students were from Asia and most were graduate students.

These study subjects, from various disciplines within the university, completed a questionnaire about their first language learning background and type of ESL writing support they used. They were then asked to write an argumentative essay of 300 words on two occasions; at the beginning of the semester and at the end. The essay prompts were the same on both occasions. They were given 55 minutes to write and instructed to spend
5 minutes reading 6 excerpts, 5 minutes planning their essay and 45 minutes writing it. They were also asked to paraphrase items from the excerpts in their essays. Citations were also required. The essays were assessed quantitatively and qualitatively. Content was assessed qualitatively based on idea formation. Form was assessed by how students used grammar and range of vocabulary (by consulting the Coxhead (Academic Word List) and ability to paraphrase and quote. The students’ paraphrases of material were rated on this scale of magnitude: near copies (NC) (students borrowed 50% or more of the original text) moderate revisions 10-49% (MR) and substantial revisions (SR) 10% borrowing. Quotations were quantified separately. Emphasis was given to the introductions to see if they contained background information and were cohesive. Form looked at ideal development and structure.

To insure inter-rater reliability, a random sample of the exams was scored by a second scorer. The percentages of accuracy were considered adequate for this research. Comparisons of the different rater’s data which assessed the essays for Time 1 and Time 2 showed an improvement in the qualitative measures of their writing. They had gained proficiency in the areas of structure and content development. They also showed an improvement by decreasing their use of informal language. The quantitative measures of fluency, accuracy and grammatical/lexical complexity showed little improvement. It was reasoned, however, that the limitations of length and time, as well as use of more complex grammatical structures, may have influenced this outcome.

This study further extends existing knowledge of test/re-test design studies, offering additional information about how to analyze students’ ability to summarize
paraphrase and quote by introducing a percentage-based means of scoring. Other than that, it is fairly similar to the 2009 study being replicated here.

To further enhance knowledge of the test/retest model of evaluating the progress of ESL students in composition programs a further study by Storch and Tapper (2009) is relevant. In it they set to measure the effectiveness of an English for academic purposes course, offered by the University of Melbourne for graduate students. The study sought to measure the impact of the course on ESL student writing in structure, accuracy and academic vocabulary. To this end, a study was organized, in which 69 students participated. Most of them were graduate students from China who were in their first semester at University of Melbourne. All had achieved a score of at least 6 on the IELTS and 18 had 6.5. The course itself was structured to achieve authenticity. Students were required to write a 500 word summary of a text, then a 1000 word critical summary comparing 2 texts, then a 2500-3000 word research proposal, and present an oral seminar. At the beginning of the 12 week course, they wrote a timed thirty minute essay on a topic of their choice relevant to their major, which served as a pretest, and at the end of the course they wrote the same essay in the same amount of time and that served as a post test. The two essays were analyzed for the factors listed above. Additionally, a survey was conducted week 10 about the student’s perception of the course, their previous work and English related experiences.

When the final test was completed, it was coded by independent coders, with first and second test labels removed. For fluency, number of words per T-unit was accounted for. Errors were coded and the Academic Word List was used to measure academic vocabulary use. Finally for text structure, they made up their own guide.
An ANOVA (analysis of variance, which is a statistical way of determining if the means of several groups are equal) was used to calculate the results of the quantitative part of the test; in test two, essays were shorter, but not significantly so; fluency did not show a significant difference either. Accuracy improved in 17 categories, but the top 4 mistakes persisted. Use of Academic Word List words increased in the second test. Qualitative results showed 33 students improving in structure, 23 improving in content, 7 did worse (3 in coherence and 2 in content and 2 in structure). Some strong writers showed no improvement. The answers to the survey questions were mostly positive and students credited the course with improving their academic writing ability. The article ended with the caution that this is a study based on a rather short course and one text.

The preceding study extends knowledge of ESL writing course testing which is of concern to this thesis. This is not only because it is by Storch, the author of the primary replication study, but also because it introduces a few variations to the other studies. The subject sample is larger and the essay-test they write is longer. It also introduces the element of text comparison to the process, which previous studies had not done.

Citing scant research information regarding the efficacy of L2 writing programs other than English, Storch and Benvento (2011) conducted a study of a French L2 writing program in Canada. It was comprised of 15 secondary students who were native speakers of English. The data collected were three essays taken from a standardized test, upon which the students’ grade was based. The essays were written approximately one month apart allowing time for possible improvement. These fifty minute essays; an essay about the black market during WWII, a film review, and a magazine article about the homeless, were collected by the researchers and coded for T-unit complexity. T-units are defined as
clauses, dependent clauses and independent clauses. They were also analyzed with grammar and global concerns in mind.

Results found some improvement in student writing over time. The students showed evidence of more creative use and less reliance on language chunks. They also showed improvement in the vocabulary use and also in complexity of structure. They didn’t however show any statistically significant improvement in grammar. The author finished the study by stating that the study was only based on only a few writing samples, which served loosely as pre and post tests. These samples were of different genres as well, which could perhaps be a mitigating factor in measuring student performance. It would have been interesting if this study had been more formalized using an official test that was inter-rater reviewed.

Even though it has some limitations, this is still a useful study to examine in light of this project because it introduces another element for consideration, namely, the use of different genres in the pre-test/post-test study design. It helps to give clarity to test construction because it underscores some problems related to ambiguity of results when multiple genres are introduced in the testing process. Therefore, multiple genres will not be used in this study.

To give more background about the test/re-test design and to get a perspective other than Storch’s, a study by Icy (2002) will now be examined. In it, she presents a systematic and practical example of how to teach writing coherence. She began the article by defining coherence as proper connection of text using cohesive devices, information structure which contributes to topical development, connecting content by relations between propositions, macrostructure, context, situation and signaling the reader with
metadiscoursal features. She went on to describe the problem many teachers face when trying to teach cohesion. This problem lies in the fact that cohesion is a kind of fuzzy concept and is somewhat difficult to get a handle on. The 16 students in the study, Cantonese-speaking low-level English writers with 16-17 years of exposure to English, were introduced to each topic by a handout containing definitions and were then asked to engage in text analysis. Then they were given a text and asked to read and discuss it in the light of one of the six concepts, which they were asked to discuss with their peers. Next, they were asked to revise it to make it more cohesive. In total, students were assigned to write 10 drafts, two revised twice each and two revised three times each with these goals in mind. They were also asked to engage in think aloud protocol while writing. Only 4 were willing to do the think aloud protocol however.

The pre and post drafts of the student’s writing served as pre and posttests. They were analyzed by Icy, a second rater, with whom she conferenced. They were also rated by a group of three independent raters, with whom she didn’t conference. They analyzed the student’s work in-depth to see if it had improved in cohesion. Results showed students were more likely to elaborate on their post-revision texts. They were also less likely to rely on sequential progressions. Think aloud protocol was also shown to be helpful in developing cohesion for those who did it. In a post-study interview, students were generally positive about the value of the instruction they received about cohesion, although some had a rather narrow view that good writing was simply following the 6 areas outlined in this particular course.

As a means of giving context to this thesis, this article shows how pre and post testing works. But most importantly, and of interest here is that it introduced the concept
of the impact of conferences in the process. The ESL Composition I class includes conferences it is hoped the test administered as a part of this project will show improvement in the student’s writing ability as a result.

Evaluating some Psychological Features Related to the Evaluation of ESL Writing

Although the concept of evaluating a writing course with a test and retest design is instructive, the attention of this study will now turn to other factors that come into play when evaluating a writing course. One of the most important aspects of a course is how it impacts the students psychologically. If the teacher of the course can gain insights into how students mentally approach the literacy problems they are given, methods which promote good learning attitudes can perhaps be designed to help them in their quest for linguistic knowledge. This can further knowledge into proper test design construction and administration.

Cumming endeavored to do just this in a study he published in 1990 where he forayed into an exploration of the minds of second language learners. His goal was to understand some of the thought process and psychology that happens when people write in their L2. To do this, he analyzed metalinguistic data provided by students through think aloud protocols while they composed. The aim of this study was to describe frequencies of occurrences of metalinguistic ideational thinking as the students wrote; to determine if the frequency of idea generating occurrences varied according to learners writing proficiency and nature of the writing task. This in turn would generate insight into how the nature of the reported thinking relates to second language acquisition.
To this end, Cumming designed a study where 23 Canadian francophone students, representing three levels of writing proficiency were participants. These students were given two writing tasks. These were an informal letter and an expository piece on women’s role in contemporary society. Participants were asked to speak their thoughts as they composed and record them on audio tape. These recordings were subsequently transcribed. Units of speech were segmented into separate units if they followed a pause of 3 or more seconds. These units were determined to be single episodes of thinking and were coded as such. These aspects were gist, language use, discourse organization and procedures in the writing process. An inter-coder agreement of 87% and an intra-coder agreement of 95% were achieved. To analyze the relationship of the learner’s expertise and second language ability to their reported thinking language use and gist, a MANOVA analysis was done. It was found that 30% of reported thinking decisions were applied to gist and language use concurrently. This was not uniform however. The students with the greatest second-language writing proficiency did this more often than those demonstrating lower-second-language-writing proficiency. Additionally, it was found that most students applied a greater percentage of their decisions to thinking about language use and gist when working on the argumentative assignment. The students demonstrated two general goals when engaged in metalinguistic thinking: to find appropriate words and phrase and to determine cross linguistic comparisons. This activity was so common its frequency wasn’t even calculated, thus eliminating the initial question a.). Analysis also indicated that the more proficient writers used metalinguistic thinking and idea generating thinking much more than their counterparts with lower skills.
Although my study doesn’t use audio recordings, nor does it use think aloud protocol, this study by Cummings is useful in aiding our understanding of how students think as they take the tests. Since so much mental effort is put forth to translate from one language to another, students are allowed to use dictionaries during the taking of the tests administered during this project. This will make the process more efficient and allow the students to spend more time thinking about writing, not translation.

Continuing this inquiry, Ojima (2003) also looked at a psychological feature of testing in her study where she examined the effects of pre-task planning on the performance of ESL writers as well as the ways in which Japanese ESL writers utilize it. She was especially interested in how complexity, fluency and accuracy of student writing were affected by it. This was a small study, with only three Japanese students who were taking classes for the TOEIC and TOEFL test in a language school in Canada. The students in this study were taking a writing class in which the instructor was teaching them how to do concept mapping, in which students arrange concepts in tree-like formations on a page as a prewriting strategy.

While conducting this research, the author observed the class where the instruction was taking place and the three students Miho, Yuri, and Chie and took field notes while observing. They were taught how to use concept mapping and asked to use it during the planning process for their writing. An interview was conducted after this to ascertain how the students used concept mapping, whether they perceived a benefit from it, if it presented any trouble to them, and what kind of problems they encountered. All were fairly positive about it. Yuri had some reservations however.
Because of the evidence presented here that pre-test planning is important and helps the students to generate writing with more fluency and accuracy, the test given during this project will give students time to prepare for the writing of the test. They will not be required to simply start writing. No instruction will be given to them on what kind of planning to do however.

Next, Ong and Zhang (2010) continue the discussion of psychological factors affecting student performance by examining how time and complexity of the prompt given to students affects the level of complexity of their writing. Secondarily, they also looked at the effect revising has on written complexity. They sought to determine if student fluency and lexical complexity would be affected by making an assignment more or less complex. There were 108 Chinese EFL writers who participated in this study, ranging in ages from 16-19. They were given some evaluative writing exercises to determine their writing level and then groups were made, stratifying the different writers, so that each group would have a similar level. There was a control group and three other groups. Differing levels of complexity were achieved in the writing assignments by providing a more comprehensive prompt to one group, less so to another group and finally, the other group was asked to free-write. The different groups were also given different amounts of time to plan in the pre-writing process and then the amount of time varied that they were allowed to write. Also one group was allowed to revise drafts before submitting and other groups were not. Results showed that a higher task complexity resulted in higher fluency II (average number of words produced per minute of total time spent writing on the assignment) and lexical complexity, but there was no significant effect on lexical complexity. Greater task complexity also produced
significantly higher lexical complexity, but did not affect fluency I (average number of words produced per minute) or fluency II. Also the increase of task complexity through having students produce more than one draft didn’t result in greater fluency or lexical complexity.

For reasons given by Ong and Zhang in the previous study, prompts for this pre and post test will be simplified but not oversimplified. Care will be given to make prompts that are appropriate length, and have the proper level of complexity.

To further knowledge on the subject a paper related to the psychology of testing by Reid and Kroll (1995) will be examined next. In this paper they set out some important criteria that they believe professors should use when designing writing prompts for papers and essays assigned in their classes. The focal point of the study was that writing assignments are a form of preference-based testing and therefore similar care and attention to guidelines should be exercised when designing them. To this end, they outlined six variables that should be taken into account when making up a writing assignment or prompt. These are contextual variables, content variables, linguistic variables, task variables, rhetorical variables, and evaluation variables. It was also underscored that care should also be given to create prompts and assignments that take the needs of ESL students into consideration. Teacher should be able to answer certain questions about their writing assignments such as: What is being tested? How will it be tested? Are guidelines clear and appropriate? Is sound pedagogy being applied? Is the assignment too broad or too narrow? Does the prompt use understandable language and culturally appropriate topics? The study went on to give many examples of appropriate and inappropriate prompts.
This Reid and Kroll study served to re-emphasize the point made by the previous study that tasks should be made clear to students, otherwise confusion and frustration can cause poor outcomes. During test design, this concept will be taken seriously and adequate explanations given to clarify the students’ task. It is hoped that this will help to extend the knowledge of proper prompt construction when tests are given for similar studies.

In designing a test it is important to apply some standards to the process. Way et al. (2000) conducted the next study under consideration. In it they take issue with the standards of the ACTFL (American Council for Teaching Foreign Languages) proficiency guidelines for foreign language classrooms. To test students’ proficiency and in doing so, endeavor to disprove the ACTFL guidelines they gave tests with different kinds of prompts. These writing prompts included a bare prompt (simple statement of task), a vocabulary prompt (a list of French words and expressions with English definitions to be used in the writing task), prose (requiring reflection and analysis), and expository narrative (students were required to write a letter about teenagers’ role on American society). Each student was randomly assigned 3 of nine possible task prompt combinations. Three writing samples were collected from each student over a three month period. They were also given a questionnaire about each assignment, the results of which weren’t reported in this study. The net result of the study, after testing was that the prose model produced the highest level of accuracy.

Since this study reported the highest degree of accuracy using the prose model, the test prompts given to students in this thesis study will require them to use the prose
model. Content and development is one of the things this study seeks to measure, so it makes sense to use a test prompt that fosters it.

Raters are an important component of the course evaluation testing process. It is therefore useful to look at research related to their role in the process. Rater psychology comes into play when they evaluate tests and this affects outcomes because certain aspects of the test are more important to them than others. To assess the effects of how different components of predetermined grading criteria for writing tests affect scores given by raters, Ahour and Mukundan (2009) conducted a study. They found evidence to recommend analytic scoring of L2 writing over holistic scoring. To reach this conclusion, they conducted a study based on a writing test. They gave 128 Asian students a 45 minute timed writing test. Two prompts were given and the students were asked to write a five paragraph essay. These essays were rated by two qualified raters, who achieved a high degree of inter-rater reliability. It was found the grammar component played the most significant role in what grade the essay was given. It was further suggested that that rubrics can be useful to direct teachers’ emphasis as they instruct the students. They can also make the grading process fair and uniform so that considerations such as grammar don’t overwhelm the grading process. For this reason, among others, the tests given in this project will not assess grammar. The primary goal of this project is to analyze students’ global writing ability. Previous studies reported no improvement in grammar. So, for this reason it was determined to forgo a grammar component in the test analysis scale.

Pitfalls in Testing
Heretofore, this literature review has looked mostly at tests made by teachers, given to evaluate various courses of study. This section has been included for several reasons. First of all, it will demonstrate why an original test will be used to evaluate the 1110 program, instead of a standardized test. Secondly, and most importantly, these studies contain important information about the various pitfalls that can develop when rating tests. These are placed within the context of standardized tests, but many of the fundamentals are the same when it comes to rating of written work.

First of all, let’s look at Green’s (2004) study in which he took issue with a notion put forth by the IELTS establishment. He did an investigation of the recommendation proposed by the IELTS partners, that 200 hours of study will result in an increase of one band on the IELTS exam. He explicated several studies on this topic and found that the idea that 200 hours of study can make students increase one band level on the IELTS was not supportable. What the studies found was that low level students made the greatest gains on a test given to them, which was similar to the IELTS writing section, while the higher levels students made smaller gains. According to this study, two hundred hours of classroom English preparation, is unlikely to improve scores very much but the study stopped short of saying that it was unhelpful. It was pointed out that test-prep courses as well as English for academic purposes courses can aid students in learning about local study habits and expectations.

This study will attempt to answer the question that was unanswered by this article. It will try to say what is helpful for learning in an ESL course.

To continue the discussion of pitfalls in testing in a slightly different vein, the next study presents information on how language proficiency is determined. Iwashita,
Brown, McNamara, and O’Hagan (2008) is a useful study to look at, even though it doesn’t analyze a written test. It focuses on how spoken language proficiency is determined and seeks to answer two questions about how levels are determined when ESL test takers take the TOEFL iBT test. These two questions are as follows: How does performance on English for academic purposes (EAP) speaking tasks vary by level and what features of test performance distinguish each of the five levels of proficiency on the iBT? The features they felt were important in influencing raters and therefore worthy of study were accent, comprehension, vocabulary, and fluency, and grammar. The section of the TOEFL studied was the speaking section. Levels were assigned to the test takers by experienced raters. During the course of this pilot test project, the five subjects involved completed 200 speaking tasks. Some had prompts; others had what the study called “stimuli.” Many measuring tools were used, including statistical sampling and rater feedback. Results showed more influence by production features such as pronunciation, fluency, and vocabulary awareness than by grammatical features. They found that levels could be determined by range of features used throughout the process that could be identified by the raters.

The relevance of this article to this thesis research is to show that raters need clear instructions as they rate the tests. Care will be given to make sure that every feature of the writing being evaluated by them will be weighted appropriately and results won’t be skewed by raters focusing on one aspect of the students’ writing.

For further analysis of the pitfalls in testing Jinyan (2008), takes issue with the results ESL students routinely receive on a standardized, government issued exam in Canada. The test in question is the provincial English exam that assesses high school
seniors’ reading and writing abilities. It is intended for students who are going on to university/college and tests their ability to analyze and interpret short stories, poems and non-fiction essays. It has four parts 1) informational text; 2) interpretation of poetry; 3) interpretation of literary prose; and 4) an original composition. Each of these tasks is rated holistically on a 6-point scale and rating is done by two scorers. A bundle of 30 tests is split between these two raters and each pair of raters only rates one section, for which they have been trained, such as poetry, prose or composition. After scores have been assigned, papers are compared and only one point variance is allowed. If the raters can’t come to an agreement, the section head arbitrates. This methodology is of interest to my study because I will use a similar rating methodology.

Jinyan compared three years of exam results (2002-2004) and did some statistical analysis using measures they felt were more accurate than the ones used by the Canadian government. The basic question was this: are there any significant differences in rating variability and reliability between ESL students and native English speaking students on this provincial exam when measured across three years?

Statistical methods used were quite complex, and therefore won’t be discussed here. Only the conclusions will be presented. First of all, there were differences in the variation of scores between ESL and NES students, therefore these two groups had unequal performance across tasks and the ESL students were consistently lower. Secondly, the residual variance component of the analysis was significantly larger for ESL students. In other words, there was a lot of unexplained difference in the scores received by ESL students as compared to NES students. This led the author to conclude
that these results indicate there may be some unfairness in this testing system because
there should be no difference in the variability and reliability of scores.

According to this study, raters should be carefully trained and fully monitored to
make sure they are giving proper scores to the tests. In this study, raters will be given
clear guidelines to insure accuracy in the rating of the tests.

**Placement Tests with Writing Components**

Many institutions who accept international students use placement tests to
evaluate the level of students who enter. These are often in addition to the standardized
language proficiency tests because results of standardized tests such as TOEFL, IELTs
and similar ones for other languages may have been taken some time in the past and may
not provide accurate linguistic data for the admissions process. The examination of these
exams is relevant to this study because they shed light on how academic language tests
are conceived and executed for purposes of evaluating a course of study.

Crusan (2002) authored a study that raised some concerns on how ESL students
are placed when entering American universities. She found that many universities use
indirect methods (multiple choice questions) to assess student writing during the
placement process. She proposed that this is not an accurate method of placement,
especially for ESL students, stating that writing assessment methods that don’t use actual
text are problematic, resulting in guesswork. She proposed that direct assessment, (essay
tests) are the best method to assess student writing. She also pointed out that there has not
been an increase in attention given to ESL students in writing programs even though their
numbers have increased. Assessment of writing is complex and requires attention to more
than the mechanics of writing. She recommended that teachers should do some out-of-
classroom assessment of students. This is one of the many reasons the test for this course analysis will be an essay test, not a multiple choice test.

Finally, Lee (2006) proposes some changes to how placement tests should be constructed and administered. The primary focus of the study is the placement test given by the University of Illinois Urbana-Champaign to incoming international students with TOEFL scores under 610 or iBT scores under 253. This test, called the English Placement Test (EPT) was initially a paper test, which students took only one time. However, the university came up with the innovation of letting students do this test in a seminar environment and get feedback from peers about their compositions. This became known as the Enhanced English Placement Test (EEPT). Later, citing the benefits of using a computer during the writing process, they instituted the Computer Enhanced English Placement Test. This test was done on a computer in a seminar environment, with peer feedback and multi-drafting opportunities.

The writing was analyzed on two scales, a holistic scale based on the Computerized Enhanced ESL Placement Test (CEEPT) and a global rating based on flow of ideas, effective elaboration, linguistic expression, synthesis of ideas and paraphrasing. They additionally were given scores from 1-4; 1 being too low, 2 being ESL 500 (a beginners class), 3 being ESL 501 (higher level), and 4 being exempt. Additionally, an analytic score was also used which looked at five areas: organization, content, grammar lexical choice, use of sources and plagiarism. They were also analyzed quantitatively based on number of T-units, T-unit length and number of words (number of words/number of T-units). There was no distinction made about whether the T-units were
error free or not because this can be controversial. Inter-rater reliabilities of 90% were achieved.

As for results, average scores were 0.293 higher on second drafts than on first ones for 58 of the 100 essays. This suggests that writers were able to benefit from peer feedback and opportunities to reflect and revise. Twenty students received scores that resulted in higher class placement. Others received higher scores, but not high enough to move them to the next level class. Five students got worse scores.

Although this test will not be administered in multiple drafts, or collaboratively, the idea that planning time is important will be taken into consideration. Students will be given time to plan their writing before they are asked to write.

It is the intent of this study to extend existing research. It will do this by having a larger sample than most of them. It also will study subjects who have been in an English medium environment (the USA) for differing amounts of time. The focus will primarily be on global writing concerns, not so much on local ones.

The next chapter will utilize the findings of the preceding literature review and will detail the test conducted at the University of Toledo. Like many of the studies in the literature review, it was a test/retest design. In it, students were presented with five writing prompts about a controversial topic and asked to write a 300 word argumentative essay in response to these writing prompts. The next chapter also shows how the tests were rated by multiple raters to determine if there was any improvement in the students’ ability to write effectively with good content, structure and development, as well as their ability to summarize, paraphrase and quote.
Chapter Two

Methodology

Design

As was previously discussed, this study is a loose replication of Storch’s (2009) research. This study diverges from Storch’s (2009) study because it involves only undergraduates who are taking a writing course. Conversely, her study involved mostly graduate students who were not taking a writing course. In her study, she sought to measure the results that studying in a L2 medium had upon the writing proficiency of L2 students. To accomplish this, she evaluated the writing of some ESL students at the University of Melbourne, using a pre-test and post-test methodology. This study follows suit and uses Composition I-ESL students studying at the University of Toledo to undertake a similar analysis of writing improvement over a semester. At the beginning of the semester, before any meaningful instruction had taken place, students were given a pre-test, consisting of five writing excerpts. These are listed in the following procedures section. At the end of the semester the same prompt and excerpts were given as a post-test to the same students to determine if any improvement had been made over the semester in the areas of summarizing, paraphrasing and quoting.

Research Questions

The following questions were selected because they are of great interest to the director of the Composition I-ESL course:

1. Are Composition I-ESL students better at summarizing after they took the course than they were at the beginning of the course?

2. Are Composition I-ESL students better able to paraphrase after completing the course?
3. Are Composition I-ESL students better able to quote at the end of the course?

**Context**

This study was conducted at The University of Toledo, a large state-supported university in the American Midwest, which has approximately twenty-four thousand students on three campuses. The University of Toledo has been aggressively recruiting international students in recent years and this has resulted in a student body that is seven percent international (P. Thomas, personal communication, January, 5, 2012). Most of these students are not native-speakers of English and as such, the academic writing process can be challenging for them. In response to this situation, the university has implemented a sheltered international-students-only writing program called English Composition I-ESL. There are generally five to six sections of this course each semester and they are almost exclusively taught by graduate teaching assistants. Occasionally, a tenured professor will teach one of them. I have had the privilege of teaching a section of this course twice and was one of its teachers when this study was conducted.

Students do not automatically place into Composition I-ESL, but must take a written exam to prove that it is a level-appropriate course for them. These tests are scored by two trained raters, each of which is usually a graduate teaching assistant who teaches a Composition I-ESL course. The director of ESL composition at The University of Toledo scores some of them as well. If both raters agree that the written test reflects writing competence commensurate with the demands of the class, the student will be placed into Composition I-ESL. When both raters do not agree, a third rater scores the test to determine placement of the student. If students do not score high enough to place into Composition I-ESL, they are placed into a preparatory English composition course called

English 1020, where they will remain until they can pass the written entrance exam for Composition I ESL.

**Participants**

A total of fifty-eight students from all sections of the course completed both the pre and post tests. There were ninety students taking the course, but not all of them took both tests due to the fact that absenteeism can be high towards the end of the semester. They were all international undergraduate students at the University of Toledo, who were enrolled in Composition 1110- ESL. These students were asked to participate in the pre and post tests, which were given as in-class writing assignments. The teachers of all six sections of the course administered the test in their classes. Of the students who took the post-test and there were thirty-one females, twenty-eight males and two undeclared. Thirty-two of the students who took the pre test were not present for the post test. Consequently the results of their pre and post test were omitted from this study. There was no survey accompanying the pre-test. Their ages ranged from 18-23, with one 31 year old and one 49 year old and they represented eleven different majors.

Also, according to available responses class standing was reported as follows:

- 22 unspecified class standing.
- 16 sophomores
- 15 freshmen
- 2 juniors 3 seniors

**Procedures**

Pre and post-tests are the primary source of data for this study. The pre-test was administered in the second week of classes and the post-test was administered in the last
or next to last week of classes. Both tests were administered by the teachers of the respective sections of the Composition I-ESL class. Students were allowed to use dictionaries. The two tests were identical and consisted of writing prompts composed of five quotes. Students were asked to write about their ideas regarding issues raised in these quotes. They were also told to summarize, paraphrase and quote and it was hoped that they would do so on the test. The test was fifty minutes long. The instructions suggested that they could take five minutes to read the excerpts, five minutes to plan and forty minutes to write. This study sought to investigate the difference in their ability to do this from one test to the next. The following instructions were given to the students before the tests:

Using animals for food and clothing is controversial for some people. You have been given a piece of paper with five written excerpts. Some of them are in support of the use of animals for food and clothing, others are not. Please write an essay supporting the side of the argument that you believe. Use the information in the five excerpts to support your ideas. Make sure to summarize, paraphrase and quote the information. You will be given five minutes to read five excerpts about this controversial topic, five minutes to prepare for the essay, then forty minutes to write it. The essay should be at least three hundred words and five paragraphs in length. It should follow the standard five-paragraph essay argumentative format. If you don’t understand what this means, that’s okay, just do your best to write an informative essay. You may use scratch paper and a dictionary. Please put your name on your paper.

Below are the five excerpts supplied to the students to be used as prompts on their tests:


Many years ago, I was fishing, and as I was reeling in the poor fish, I realized, “I am killing him—all for the passing pleasure it brings me.” And something inside me clicked. I realized as I watched him fight for breath, that his life was as important to him as mine is to me.

“I’m Casey Affleck for PETA. When people ask me why I don’t eat meat or any other animal products, I say it is because they are unhealthy and are the product of a violent
and inhumane industry. Chickens, cows and pigs on factory farms spend their whole lives in filthy, cramped conditions, only to die prolonged, and painful deaths. They bodies are then turned into food products that have been proven to contribute to heart disease and cancer. To eat that is to eat poison. Please don’t contribute to an industry that makes unhealthy products by torturing animals. I’m Casey Affleck and I’m a vegetarian.


…Animals that are killed for their flesh lead miserable lives in disgusting conditions. The simplest little thing you could do to help animals is simply not to eat them…


Animals don’t have rights because rights are conferred by participation in the democratic process. Since animals can’t participate in the democratic process, they don’t have rights.


Former vegetarians are some of the most outspoken proponents of eating meat. "I was vegan for 16 years, and I truly believed I was doing the right thing for my health," says the actress and model Mariel Hemingway, who is the author of Healthy Living from the Inside Out. "But when I was vegan, I was super-weak. I love animals, and we should not support anything but ethical ranching, but when I eat meat, I feel more grounded. I have more energy."

Additionally, students completed a survey after the post test about their attitudes regarding the class and their writing abilities. It is as follows:

As part of this research, I would also like to see how this class has influenced your perceptions of your writing skills. I invite you to complete the following questions. Do not put your name on the paper so responses can be confidential. Your decision to complete these questions is completely up to you and there are no penalties or benefits to completing them. If you wish to complete this survey, please answer the questions by marking the blank next to the statement that is most true for you. Please use your first thought about the subject in marking the blank.

I am good at summarizing a passage.

1_____ strongly agree

2_____ agree

3_____ somewhat agree
4 _____ disagree
5 _____ strongly disagree

I am good at quoting a passage.
1 _____ strongly agree
2 _____ agree
3 _____ somewhat agree
4 _____ disagree
5 _____ strongly disagree

I am good at paraphrasing a passage.
1 _____ strongly agree
2 _____ agree
3 _____ somewhat agree
4 _____ disagree
5 _____ strongly disagree

I am good at structuring a research paper.
1 _____ strongly agree
2 _____ agree
3 _____ somewhat agree
4 _____ disagree
5 _____ strongly disagree

I am good at structuring an argumentative essay.
1 _____ strongly agree
2 _____ agree
3 _____ somewhat agree
4 _____ disagree
5_____ strongly disagree

Please estimate how many (typed, double-spaced) pages of writing you have done this semester for courses besides this one. In other words, how many typed double-spaced pages have you written this semester that were not for this English class?

It should be noted at this point that lessons about summary, paraphrase and quotation were included in the curriculum of the course as well as the other subjects listed on the survey. These lessons were outlined on the syllabus and the director of the ESL writing program provided the instructors with resources to be used in teaching them. Additionally, a textbook was used which had many good lessons on these topics as well.

A certain amount of discretion was given to the instructors about how instruction on these topics was to be approached.

Qualitative and quantitative analysis of citation and summary/paraphrasing

Initially, the items under consideration by this study, summary, paraphrase and quotation were going to be scored individually. After the data was reviewed by myself and the advisor to this thesis, the determination was made to rate all three items under consideration holistically. The reason is that these three items are somewhat subjectively measured, and also rather imbedded in the text. Additionally, there did not seem to be a clear distinction between summary and paraphrase. The quotations, while present in the data didn’t necessarily conform to standard conventions for citation as well. Therefore, a holistic score was determined to be most appropriate when rating these tests.

This study used 10 raters who were teachers currently teaching Composition I – ESL, or had taught it in the past. They were tasked with reading the exams and determining their scores. These raters were trained by presenting them with three previously-selected essays, that were determined by this thesis-committee chairperson, as
well as the author, to be accurate representatives of each of the 4 score levels. To clarify, the score of zero on a test means there was no attempt to summarize, paraphrase or quote on the part of the student. A score of one means there was at least one attempt at summary/paraphrase. A score of two was assigned if the effort of the student to summarize, paraphrase and quote was in between one and three. A score of three means there was a satisfactory effort made to summarize/paraphrase and quote where the student provided at least three summary/paraphrases and quotations. Here are some examples of each category:

Below is an example of an exam that was given a score of zero given its absence of an attempt to summarize, paraphrase or quote. It is essentially a student’s editorial opinion on the writing prompt.
Here is an example of an essay that was given a score of one because there was only one attempt to use quotations. The summarizing and paraphrasing was also not a prominent feature of the student’s passage.
Human beings have been living in the nature with other living organisms since the origin of mankind. It is no exaggeration to say that human beings have progressed their culture and industry by using animals for food and clothing. It is still the same today and people breed many kinds of animals for many different reasons. It seems like it is essential for humans to breed animals for their uses.

First of all, humans cannot make a good living without animals. There are so many industries in the world that use animals for food and clothing. It takes a vital influence on economy and many people work for these companies for their living. If the food and clothing industries collapse, one may get to experience the most disastrous great depression in the history.

Human beings have adapted to eating meats. Meats contain many nutrients that people need and it is very important for humans to have nourishing meals to keep themselves in the healthy bodies. For example, Christine Lennon, the author of "Why Vegetarians are eating meat," said, "But when I was vegan, I was super-weak... I love animals, and we should not support anything but ethical ranching, but when I eat meat, I feel more grounded." It obviously shows that people need, at least, some amount of meats for their health.

It is almost impossible for humans to have a stable life without using animals for their needs. This industry has been growing with the history of human progress.

Next, is an example of a text that was given a score of two. It has two quotations and has some summary/paraphrase in it.

Have you ever questioned yourself: 'Why do you eat meat? What would have happened with chickens, cows and pigs you? Whom the meat has been obtained? What would you have done if you were in place of those poor animals?'

Me, being a former vegetarian, firmly believe that using animals for food and clothing is not good and should be considered as a crime killing others living being for our selfish
needs us a 'meat' ounce. I totally agree with the sociologist, Casey Affleck, who gave a vegetarian testimonial on April 1 stating that eating meat is unhealthy and a product of violence, of inhuman nature and industry. We are unaware of the meat source ie animal was ever suffering from contagious disease and if the animal was suffering from any such disease, was he ever treated or vaccinated against that disease? Upon our consumption of such infectious meat may lead to several intestinal and mental disfunctionalities and may lead to hormonal imbalance in the body.

As Sadie Frost said on 7th August 2007 in a vegetarian testimonial that animals which are killed for their flesh lead miserable lives in distinguished conditions. This holds absolutely true! As mentioned in the above paragraph, when we question ourselves, what must have happened with the animals, who were killed for flesh and for our selfish needs? We will get an answer from Casey Affleck's a vegetarian testimonial which states that animals spend their whole lives in filthy, cramped conditions, only to die prolonged, painful deaths just to get our own hunger. Animals have their own life which is the animals' own life to be lived their own way; not ours. Products obtained from animals are the unhealthy food we by simply not eating or using the products obtained from animals. Try avoiding them and if possible, make least use of them daily.

As stated in the above paragraphs, using animal products for eating or clothing is not a
Here is an example of an example of a paper deserving of a score of 3. There are two summary paraphrases and the one quote. Additionally there are three in-text citations.

Whether using animals for food and clothing is fine or not has been discussing for a long period. Some people say that it is fine because animals are not human beings and they do not have rights, whilst I strongly feel that animals should not be used for food or clothing, and their lives should be respected. The following paragraphs will explain the both merits and demerits for animals using for food and on clothing.

Using animals for food and clothing is fine to some people. They say that the people who have rights are supposed to participate in the democratic process, while animals are not able to do that. (The Rush Limbaugh Show, 1993).

However, a lot of people say that using animals for food or clothing is too cruel to them. First of all, Paul said that once he was fishing, he realized, "I am killing him—all for the passing pleasure it brings to me" (2007). Fish animals' lives are as important to them as people's lives to us.

Also, some vegetarians say that the reason why they do not eat meat is because the animals are unhealthy, killing animals for vision. Above all, although animals do not participate in the democratic process, they still have the rights to protect themselves from injury or death, because their lives are as important to them as human's to us. People should no longer use animals for food or clothing; instead, we should protect them. Otherwise, we will pay for what we have done.

Before the rating session began, the 10 raters were asked to rate these 4 essays to determine if they could properly identify the rank of each essay, and therefore establish
inter-rater reliability. An inter-rater reliability of 90% was established as a result of the training, insuring uniformity of results.

After the training was completed all the tests were subsequently rated. For this reason, each test was scored by two raters and some by a third rater if the first two could not agree. Both the pre and post tests were rated together, and the raters were not aware which was which. Additionally, the first of the two raters was instructed to place a Post-It note on the score they wrote on the test, so as not to influence the second rater. If the score of the second rater was the same as the first, the rating process was considered completed. If the scores disagreed, it was rated by a third rater.

After all of the tests were rated, the next phase of the study was to conduct quantitative analysis on the data that was collected from the tests. The goal of the quantitative analysis applied in the rating of the tests was to determine if the scores of the post-tests were higher than the scores of the pre-tests. To accomplish this, data was coded into SPSS (a statistical analysis program) and the mean of the pre-test was compared to the mean of the post-test using a Wilcoxon W statistical test. This is a statistical test which compares the means of two ranked samples. The null hypothesis (H₀) was that the two tests had the same means, or the same results. The alternative hypothesis (Hₐ) is that the post test was higher. The results of the statistical analysis will be presented in chapter three.

I summary, this chapter presented the methodology that was used to implement this study. A pre and post test was administered to Composition I-ESL students at the beginning and end of the semester, subsequently rated by trained raters and then analyzed statistically to determine if there was an improvement in the areas of summary,
paraphrase and quotation. In the next chapter the findings and their implications will be presented. Suggestions will be made about the direction the Composition I-ESL course should take in the future. Additionally, limitations related to the study and suggestions for further research will also be presented.
Chapter 3

Results of the study

In this first chapter of this thesis, one can see information presented that is related to the rationale and structure of this study. Methodology can be seen in chapter two and now in chapter three, results of the statistical analysis as well as its implications will be discussed.

Findings

After the scores of the tests were tallied and input into SPSS, two analyses were done using a Wilcoxon W. This is a statistical test to determine if there is any difference in the means of two ranked, related samples, which in this case were scores for the pre and post-test administered at the beginning and end of the semester. The null hypothesis (Ho) for this analysis assumed there was no difference between the results of the two tests. The alternative hypothesis (HA) assumed that the second test results would be higher than the first. The first test, using the Wilcoxon W analyzed all of the data and resulted in a p value of .000. The second analysis was done after all the test score results of students who only took one test were removed. Test results were removed from both the pre and post tests. A second Wilcoxon W analysis with the modified data yielded the same results with a p value of .000. This strongly indicates that the null hypothesis can be rejected and that the results of the second test were higher than the first.

Additionally, when individual test results were tallied, of the students who took both tests, there were twenty-nine students who went up one level in their ability to summarize, paraphrase and quote. Fourteen went up two levels and one went up three
levels. Four students went down one level, and ten showed no change. On the pre-test, after results of those who didn’t take both tests were excluded, there were twenty-four with a score of zero, twenty-six with a score of one, eight with a score of two and three with a score of three. On the post-test, after the scores of those who had not taken both of the tests were removed, there were four scores of zero, eighteen scores of one, eighteen scores of two and twelve scores of three. The results of both tests can be seen in the graph below.

![Graph showing pre-test and post-test results](image)

**Figure 1** Results of the pre-tests and post tests.

Finally, the results of the attitudes survey that was referenced in chapter two are included in figure 2. There were so few responses about the number of pages written per semester, its results were not included.

<table>
<thead>
<tr>
<th></th>
<th>I am good at summarizing</th>
<th>I am good at quoting</th>
<th>I am good at paraphrasing</th>
<th>I am good at structuring a research paper</th>
<th>I am good at structuring an argumentative essay</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 strongly agree</td>
<td>11</td>
<td>20</td>
<td>16</td>
<td>10</td>
<td>12</td>
</tr>
<tr>
<td>2 agree</td>
<td>35</td>
<td>23</td>
<td>22</td>
<td>36</td>
<td>30</td>
</tr>
<tr>
<td>3 somewhat</td>
<td>16</td>
<td>15</td>
<td>19</td>
<td>14</td>
<td>16</td>
</tr>
<tr>
<td>agree</td>
<td>0</td>
<td>1</td>
<td>4</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>-------</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>4 disagree</td>
<td>0</td>
<td>1</td>
<td>4</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>5 strongly disagree</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

*Figure 2* Results of the attitude survey.

**Discussion**

These results are, of course good news for those who teach this particular writing course. The teachers in this program are quite diligent in helping the students understand the importance of summary/paraphrase and quoting. The course has this as one of its foci because plagiarism is often a problem in undergraduate populations on university campuses, especially among international students. Often international students may not have a strong grasp of the conventions of English writing and are often tempted to include direct quotations in their written work without adequate citation. Furthermore, they may come from cultures where this is acceptable. It is therefore one of the primary strategies of this course to confront this problem and direct international students to utilize acceptable English writing conventions, especially in the areas of summarizing, paraphrasing and quotation. Results of this study indicate that the Composition I- ESL course at The University of Toledo is having a measure of success in attaining that goal. This determination can be made because of the results of the statistical test, showing an improvement in the students’ scores on the pretest and post test. The students seemed to agree that they had improved if one takes the results of the attitude survey into
consideration. Most of the responses were skewed toward the positive, as can be seen in figure II.

**Implications of research**

The results of this research can be characterized as good news for the students involved as well as the Composition I- ESL director and teaching staff. Some of the main goals of the course are being met to a certain extent. This does not mean, however, that there is not room for improvement. When individual student scores were compared from the beginning of the semester to the end of the semester, a good number went up, but only by one level. The fact that only fourteen participants had a score increase of two and one participant had a score increase of three shows room for improvement. Results showing more students increasing by two or three levels would have been desirable. The strategy to upgrade the course, so as to have more dramatic improvement among students’ L2 writing could be viewed as task without clear guidelines. Although there is always room for improvement, to say that teachers should work harder or teach better misses the mark by many measures. The teachers in this program are dedicated and hard working. The director makes sure that they are well supported and have the materials they need. Additionally it also cannot be assumed, based on this research that students need to work harder or be more dedicated. They tend to be hard-working students, who have their family’s expectations resting upon them. Many are applying to professional programs that require a certain grade point average, so they are quite attentive to the progress they make in class. Therefore, I think the primary implication of this research is that this course should be continued largely as it is currently conceptualized with emphasis given to the caution that its teachers should stay focused on the students’
progress in summarizing, paraphrasing and quoting. They should also make sure that students understand them at all times. Additionally any logical improvements to the course should be implemented.

**Limitations of Research**

In studies of this nature, the limitations can be numerous. First of all, it is always desirable to have a larger sample than what is available. Sixty participants is a good number, but more is better. Additionally, this study was only conducted on the progress of the students over one semester. To get a more accurate picture of student performance, administering the same study over numerous semesters would also be desirable. Another mitigating factor is that some of the students only took the pre-test and some only took the posttest. Therefore their results had to be excluded. Reflecting one hundred percent of the first and second results would have been preferable. Student attendance throughout the semester is also an important factor affecting the outcome of the post-test. If a student missed numerous classes during the course of instruction their outcome on the final test could possibly be compromised. This would especially be true if the student had missed the specific classes where summarizing, paraphrasing and quoting were discussed. A final limitation that may affect the outcome of the final test is student motivation. At the end of the semester, students’ motivation is often at low level and consequently, they may not care as much about their output.

**Suggestions for Future Research**

The original study, Storch (2009) upon which this one is based, measured student writing performance by looking at many other factors in addition to summarizing,
paraphrasing and quoting. This study did not replicate very many of them, but if further research were to be done at The University of Toledo or another institution, her study could serve as a guide for further research. Using it as a guide one could evaluate the students’ global writing ability by qualitatively measuring structure, development and content. In addition to this, one could also look at fluency (number of words used), complexity of clauses and use of academic language. These could be measured both qualitatively and quantitatively.

Conclusion

The results of this study were positive for the Composition I- ESL program at The University of Toledo. After a pre-tests and post tests were administered at the beginning and end of the fall, 2011 semester to all sections of the class, approximately sixty students, a statistical analysis of the results strongly indicated an increase in the students’ ability to summarize, paraphrase and quote. Although these results were positive, there is always room for improvement. After conducting the study and examining the results, this research suggests that there does not need to be a radical overhaul of the course, but constant vigilance should be maintained to implement every incremental improvement to the course that can be made.
References


Reichelt, M. (2010, month unknown). Title of lecture not specified. *Issues in ESL Writing.* Lecture conducted from The University of Toledo, Toledo.


