

2013

# An extensive reading approach to teaching English second language reading comprehension with the American Language Institute at the University of Toledo

Neil Edward Sampson  
*The University of Toledo*

Follow this and additional works at: <http://utdr.utoledo.edu/theses-dissertations>

---

## Recommended Citation

Sampson, Neil Edward, "An extensive reading approach to teaching English second language reading comprehension with the American Language Institute at the University of Toledo" (2013). *Theses and Dissertations*. 194.  
<http://utdr.utoledo.edu/theses-dissertations/194>

This Thesis is brought to you for free and open access by The University of Toledo Digital Repository. It has been accepted for inclusion in Theses and Dissertations by an authorized administrator of The University of Toledo Digital Repository. For more information, please see the repository's [About page](#).

A Thesis

entitled

An Extensive Reading Approach to Teaching English Second Language Reading Comprehension

with the American Language Institute at the University of Toledo

by

Neil Edward Sampson

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

December 2013

Copyright 2013, Neil Edward Sampson

This document is copyrighted material. Under copyright law, no parts of this document may be reproduced without the expressed permission of the author.

An Abstract of

An Extensive Reading Approach to Teaching English Second Language Reading Comprehension  
with the American Language Institute at the University of Toledo

by

Neil Edward Sampson

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

The University of Toledo  
December 2013

Seven English second language students enrolled in informal reading comprehension and discussion classes through the American Language Institute at the University of Toledo were compared to determine whether the students who read extensively would outperform the students that did not through pretest and posttest measures. Follow-up interviews were also conducted to investigate the affective influence of extensive reading on the participants, the test-taking strategies they use while taking the Test of English Foreign Language (TOEFL), and their opinions of the reading comprehension portion of the TOEFL. The researcher found that three of the participants who were to have read extensively did not achieve any considerable gain to the scores of their posttest measures, while one participant not required to read extensively did make a considerable gain. Qualitative data from the interviews suggests that whether or not a participant possessed an intrinsic motivation for reading influenced the outcome of these measures.

## Table of Contents

Abstract	iii
Table of Contents	v
List of Tables	vii
List of Abbreviations	vii
I. Literature Review	1
A. Introduction	1
B. Correlation between reading and academic performance in second language study	3
C. A case study of L2 language gains through a self-administered extensive reading regimen	10
D. Comparison studies of the benefits of extensive reading for reading comprehension	17
E. The affective benefit of extensive reading pedagogies	34
F. Summary	39
II. Context and Design of the Present Study	41
A. Introduction	41
B. Theoretical Assumptions for Research	41
C. Description of Context in which Data was Gathered	43
D. Research Questions	46
E. Background and Role of the Researcher	46
F. Background and Role of the Participants	46
G. Methodology	47

a. Revised Methodology	48
H. Data Gathering Procedures	50
III. The Results of the Study	53
A. Introduction	53
B. Findings	53
a. Research Question 1: Will pretest and posttest measures of the English L2 reading comprehension of the participants of this study indicated a differential gain for the participants in the treatment group who read more than the comparison group, whom read less?	56
b. Research Question 2: Will the participants respond favorably to the extensive reading treatment?	57
i. Belief in the benefit of extensive reading	57
ii. Engagement with the extensive reading treatment	57
c. Research Question 3: What test-taking strategies do the participants use when taking reading comprehension tests like the TOEFL?	61
d. Research Question 4: What are the participants' opinions of the reading comprehension section of the TOEFL?	66
i. Three of the participants generally questioned the validity and reliability of the TOEFL	67
ii. Some of the participants would have preferred writing a response paper as a possible alternative to section 3 of the TOEFL, for testing English L2 reading comprehension	71
e. Research Question 5: What are the participants' opinions of the	

timed reading exercises that take place at the beginning of the ALI's reading labs?	72
f. The case of Saudi Arabian male 2	73
C. Discussion	75
a. Implications of Research	75
b. Limitations of Research	78
c. Suggestions for Future Research	79
d. Conclusion	81
References	83
Appendices	
A. Interview Questions	86

## List of Tables

Table 1	Cumulative pretest and posttest reading comprehension scores for both groups .....	54
Table 2	Pretest and posttest reading comprehension scores for comparison group. ....	55
Table 3	Pretest and posttest reading comprehension scores for extensive reading group. ....	55
Table 4	Responses to interview questions 9 – 19 . ....	62
Table 5	Pretest and posttest reading comprehension scores for extensive reading group. ....	73



## List of Abbreviations

ALI.....	American Language Institute
ANCOVA .....	Analysis of Covariance
EFL .....	English as a Foreign Language
ESL .....	English as a Second Language
L2 .....	Second or Foreign Language
TOEFL .....	Test of English Foreign Language
UT .....	University of Toledo

## Chapter 1

### Literature Review

#### Introduction

Extensive reading is an approach to second language (L2) teaching that expects learners to read massive amounts of authentic target language texts for fluency and general comprehension, supposing that this massive and varied amount of authentic target language input will catalyze the learners' acquisition of the L2. This thesis reports on a pilot study of an extensive reading program that the researcher attempted to implement through The American Language Institute (ALI) at the University of Toledo. The problem that led to this study was brought to the researcher's attention during the course of his interactions with the staff of the ALI. The problem was that a majority of students taking the institutional version of the Test Of English Foreign Language (TOEFL) exam, at the end of each session, performed poorly on the timed reading comprehension portions of the exam. Many of the students were complaining about not having enough time to complete the timed reading section. The ALI staff explained to the students that college-level academic pursuits require one to be able to read massive amounts of text efficiently and effectively, and the timed reading section is designed to measure each of their abilities to do this. Based on the researcher's review of the literature produced on extensive reading pedagogies in second and foreign language learning contexts, the researcher was eager to implement a pilot study of an extensive reading approach with the aim of improving the English L2 reading comprehension measure from the TOEFL scores of those ALI students who would participate in the study. The design of the pilot study will be explained in the methodology section of Chapter 2.

This, the first chapter, reviews the research literature that has been produced about extensive reading pedagogies. The purpose of this literature review is to highlight the evidence that indicates the efficacy of extensive reading for language learning, in order to justify the researcher's approach to solving the stated problem. The present study will extend the body of research pertaining to extensive reading pedagogies in an English Second Language (ESL) context.

The following literature review is subdivided into four categories. The first three categories consider studies that present quantitative data demonstrating the efficacy of extensive reading. They are ordered in the following way: category one includes studies that found a positive correlation between the personal reading habits of the participants, and their academic performance in second language study; category two is just a case study of a self-administered extensive reading regimen that documents the subject's progress in L2 reading comprehension through a series of tests; category three includes comparison studies examining, through pretest and posttest measures, the performance of extensive reading treatment groups against controlled groups receiving a traditional curriculum of instruction. The fourth subsection of this literature review presents studies where qualitative data demonstrates the affective benefit of extensive reading pedagogies from participants who received extensive reading treatments.

It should be noted that, although this thesis only investigates how extensive reading will improve measures of its participants' reading comprehension abilities, many of the following studies provide evidence that extensive reading has a positive effect on other areas of language acquisition, as well. Therefore, all of the benefits of extensive

reading reported on by these studies will be reviewed here, in the interest of advocating for extensive reading as a pedagogical method in general.

### **Correlation between reading and academic performance in second language study**

Some studies have demonstrated the existence of a positive correlation between the amount of time students spend reading, and general academic performance. The researcher offers a look at three in particular: Gradman and Hanania (1991); Constantino, et al. (1997); and Pichette (2005).

Gradman and Hanania's (1991) interviews with 101 English Second Language (ESL) students from a diverse range of nationalities, and first language backgrounds, who were enrolled at the Center for English Language Training at Indiana University, is among the most frequently cited studies in the body of research relating to extensive reading. This is likely due to the strong correlation that Gradman and Hanania found between greater amounts of time spent in extracurricular reading and better performance on the TOEFL. It is important to note that extracurricular reading was only one of many factors that Gradman and Hanania anticipated could affect second language learning, because the correlation between extensive reading and TOEFL performance becomes all the more salient in contrast to the other variables they investigated.

Gradman and Hanania distilled a total of forty-four independent variables from the factors that they anticipated would affect the dependent variable of TOEFL scores through personal interviews with their participants (1991). These independent variables fell into four categories, including formal learning, exposure to English and the use of it in class, extracurricular exposure and use, and attitudes and motivation. The researchers found that, among these four categories, extracurricular reading demonstrated the most

significant correlation with superior TOEFL performance, having a Pearson correlation coefficient of .53 (1991). Five other independent variables also demonstrated a statistically significant correlation with TOEFL scores: exposure to native-speaking instructors (.39); English as the medium of instruction (.36); months of intensive English instruction (.26); Effective instructors (based on student reporting; .21); and anticipated future need for English proficiency (.21) (1991).

The remainder of Gradman and Hanania's (1991) independent variables did not show any correlation with TOEFL performance; however, all of the independent variables, showed a statistically significant correlation with extracurricular reading except for seven: enrollment in private schools; French or Spanish as the medium of instruction; enrollment in an intensive English program; a recognition of need for English proficiency; effective instructors; family members' encouragement; and family members who earned degrees in English-speaking countries.

The fact that so many of the independent variables that did not correlate with TOEFL performance nevertheless correlated significantly with extracurricular reading – coupled with the findings from a multiple regression analysis that revealed a negative correlation between the two variables of total index of exposure to English and extracurricular speaking, and TOEFL performance – prompted Gradman and Hanania to conduct a path analysis, in order to determine relationships of direct or indirect causality. The researchers categorized their data into five independent variables for path analysis: oral exposure, communicative oral use, extracurricular listening, extracurricular reading, and extracurricular speaking. Further statistical calculation revealed that only

extracurricular reading directly effected TOEFL scores with a Pearson correlation coefficient of .57 (1991).

Finally, Gradman and Hanania (1991) divided their sample into two groups with respect to three of their independent variables (English instruction began before age thirteen; positive attitude towards learning English; extracurricular reading) and compared the TOEFL scores between groups where the variable was expressed, and groups where the variable was not expressed. The researchers subdivided their sample into students who had previous instruction in intensive English programs, in order to account for the effects of such programs, and further isolate these variables. As expected, in each of the three categories, the group where the variable was expressed outperformed the group where the variable was not expressed, however, for participants who had previous exposure to intensive English programs, the difference in TOEFL performance was virtually eliminated for the two variables of instruction beginning before age thirteen, and positive attitude for learning English. As for the variable of extracurricular reading, Gradman and Hanania (1991) found that prior exposure to an intensive English program could not compensate for the difference in TOEFL scores between the two groups where the expression of this variable was examined. Thus, Gradman and Hanania's (1991) study is powerful evidence of the positive effect that extensive reading has on TOEFL scores, and is frequently referenced in the literature. The researcher of the present study considers Gradman and Hanania (1991), alone, as sufficient justification for attempting an extensive reading pedagogy.

Constantino, et al. (1997) extended Gradman and Hanania's (1991) findings by administering a survey to 43 international students representing 14 different countries

from East Asia, the Middle East, or Europe. All of the subjects were attending unspecified universities in the United States at the time that this study was conducted, and had previously taken the TOEFL exam. The survey was a brief, 13-item questionnaire inquiring about the participants' TOEFL scores, years of English study, exposure to English-language media, and English reading habits and history.

The first item on the questionnaire inquired about the participant's age, and this item was omitted as a variable in data analysis. Three questions on this instrument asked how frequently the participant read English language newspapers, how much free time the participant spent reading English, and the frequency that the participant engaged in reading English-language books for pleasure, all prior to taking the TOEFL exam. The responses to each of these questions were measured on a ratio scale that referred to increments of hours per day, number of months (up to two years), or number of times a year, month, or week, etc. The researchers combined these items into one independent variable labeled "free reading." Two other items from the questionnaire that used the same type of ratio scale to measure frequency of pleasure reading in the native language prior to arrival in the United States, and frequency of current pleasure reading in the native language, were reduced to the single independent variable "read in L1" (Constantino, et al., 1997). Thus the researchers extracted a total of eight independent variables from their data for pairwise analysis, with TOEFL score being considered as the dependent variable. As has already been shown, this was essentially the same methodology that Gradman and Hanania (1991) used.

After quantifying the data from this survey, Constantino et al. (1997) found that among their eight independent variables, four showed a statistically significant positive

correlation with better TOEFL performance. Those variables are presented here with their respective Pearson correlation coefficients: number of books read (.447); amount of time spent in free reading (.390); years of English study in the home country (.367); and length of residence in the United States (.232). The variable “number of books read” is the only one significant at the  $p < .01$  level, while the rest are significant at the  $p < .05$  level.

The study conducted by Constantino, et al. (1997) supports Gradman and Hanania (1991) by showing that extracurricular reading is significantly, positively correlated with higher TOEFL scores. However, in Constantino, et al. (1997), multiple regression analysis revealed that the variables “years of English study in the home country,” and “length of residence in the United States,” were slightly better predictors of TOEFL performance, with beta values of .48 and .42, respectively, compared with number of books read, and time spent in free reading. These two variables, which were shown to be the most significantly correlated with TOEFL scores, were combined into one variable for multiple regression analysis (beta = .41) because pairwise correlations showed them to be highly correlated ( $r = .714$ ) (1997). This discrepancy between Gradman and Hanania (1991) and Constantino, et al. (1997), where extracurricular reading was the most significant predictor of TOEFL scores for the former study, but not the latter, can be plausibly explained by the fact that most of the participants of Constantino, et al. (1997) did not read extensively.

Even though extracurricular reading was significantly, positively correlated with TOEFL scores in Constantino, et al. (1997), and was a significant predictor of TOEFL scores, the population sample for that study was unique for the actual dearth of books read in English among the survey respondents. The survey provided five choices for the



question inquiring about the number of English-language books read prior to taking the TOEFL exam: 1 = none; 2 = 1-5 books; 3 = 6-10 books; 4 = 11-50 books; 5 = more than 50 books. The majority of the respondents ( $n = 20$ ) chose the second response, reporting that they had only read between one and five books in English, prior to taking the TOEFL exam. Ten respondents answered that they had not read any English books at all, while seven chose the third option (6-10 books), three chose the fourth option (11-50 books), and three more claimed to have read over fifty books in English. In other words, even though Constantino's, (et al., 1997) participants had not done much extensive reading, it nevertheless helped to improve their TOEFL scores. The implication for the present pilot study is that, even with its brief time frame and minimal level of reading input, extensive reading could still yield some observable effects among the members of the treatment group in comparison to the control group.

The results of Pichette's (2005) research support those of Gradman and Hanania (1991) and Constantino, et al. (1997), demonstrating once again the existence of a positive correlation between a second language learner's proficiency in the target language, and the amount of time spent reading in the target language. Pichette investigated two research questions: the first research question asked whether there was a relationship between time spent reading in English and English L2 proficiency, and the second research question asked whether the correlation was stronger for those with higher levels of English L2 proficiency. The study involved 81 adult ESL students from three different post-secondary institutions of higher learning in Quebec City, Quebec. In order to preserve a sample with a homogenous language profile, the participants were further sorted into two groups, where one group consisted of 68 speakers of French as a first

language (L1), English L2 participants, and the second group consisted of 13 participants where French was not the L1, or English was not the only L2 (2005).

As with the other two studies mentioned so far, Pichette used a questionnaire to yield data on the amount of time each participant spent reading in the target language. The questionnaire asked participants to estimate the number of hours spent per week reading various kinds of English texts, both in class and out of class, to produce the variable of amount of time spent reading in English, based on number of hours per week, for each participant. The two test instruments used to measure English L2 proficiency were a multiple-choice test and a reading comprehension test. The multiple-choice test assessed vocabulary and grammar knowledge by presenting an example sentence where either the participant had to select the correct word missing from the example sentence among the given choices, or the participant had to identify the incorrect word among four underlined words in the example sentence. The reading comprehension test assessed reading comprehension by having the participants read a short passage, then verify a series of sentences corresponding to information within each passage with either a yes or no response (2005).

For the 68 participants whose L1 was French, and only L2 was English, the results from Pichette's study reveal a moderate and significant positive correlation ( $r = 0.35, p < 0.01$ ) between the amount of time spent reading English and English L2 proficiency (2005). This result answers Pichette's first research question, affirming that there is a relationship between extensive reading and language proficiency. To answer the second research question, Pichette further divided those 68 participants into two groups: one of low English L2 proficiency, and one of high English L2 proficiency. The data

shows that there is no significant correlation between extensive reading and English L2 proficiency, for the low proficiency group ( $r = 0.01, p = 0.95$ ), but for the high proficiency group, extensive reading and English L2 proficiency were highly correlated and significant ( $r = 0.55, p < 0.01$ ) (2005).

These results inform the present study in two ways. The first way Pichette's (2005) research informs the present study is that it adds to the body of evidence that extensive reading pedagogies are effective in improving the English L2 proficiency of ESL students. The second way in which Pichette is relevant to the present study has to do with the progressive positive correlation demonstrated to exist between the efficacy of extensive reading pedagogies and the initial level of proficiency of the language learner, where the more proficient the language learner already is in the target language, the more extensive reading will help the language learner to continue to advance. This informs the selection of advance-level ESL students as a convenience sample for the present study.

In spite of the fact that Pichette's (2005) findings suggest that extensive reading is less effective with low-proficiency students, there is a case study of a near-beginner in Portuguese who followed a self-administered extensive reading regimen to increase his general proficiency in that language. This case study is remarkable not only for the substantial gains that the subject made in vocabulary acquisition, reading comprehension, and listening comprehension, but also for the fact that the subject began the extensive reading regimen having very little prior exposure to Portuguese. This literature review will now present an overview of that case study and consider its relevance to the present study.

### **A case study of L2 language gains through a self-administered extensive reading**

## **regimen**

Grabe and Stoller (1997) realized that the prospect of living in Brazil for five months in 1990 was a rare opportunity for a researcher of reading pedagogies to experience and observe, first-hand, the development of L2 reading comprehension from an initial standpoint of having virtually no prior exposure to the L2. This realization prompted the first named researcher, in cooperation with his colleagues at Pontificia Universidade Catolica do Rio de Janeiro, to volunteer himself as the subject of a case study in extensive reading for development of L2 reading comprehension. The actual length of this study was twenty-one weeks.

The subject of this case study was the first named author, identified in this study as “Bill,” who also happens to be one of the foremost authorities in L2 reading research. At the time that this case study took place, Bill was in Brazil to work as an English professor, and to develop the professional skills of Brazilian English professors. Consequently, Bill was generally not exposed to the target language, Portuguese, in his professional life. For his first week in Brazil, Bill stayed with a host family that spoke good English, and then he moved into an apartment where he lived by himself, for the remainder of his time in Brazil. Thus, his natural immersion in Portuguese was minimal, and he had little purpose to be productive in the language. Aside from three consecutive weeks of intensive formal instruction in Portuguese at the beginning of his stay in Brazil, where it is reported that he learned the basic tenses and sentence structures of Portuguese, along with about 170 verbs for a total of over 500 vocabulary items, the input for Bill’s acquisition of Portuguese came largely from daily reading of print sources, and daily exposure to local news broadcasts and a popular drama series on Brazilian television

(Grabe & Stoller, 1997).

Beginning from the fifth week of his stay in Brazil, Bill spent three to four hours a day reading from the daily newspaper *Journal do Brasil*. The procedure for using the newspaper as reading input involved first reading the most interesting articles from the front page and underlining all of the unknown words. Then Bill consulted a large, hard-bound, English-Portuguese dictionary for the definition of each unknown word, and reread the articles for comprehension. Bill kept a record of each word that was looked up in the dictionary on a yellow writing pad. He stopped using the dictionary to look up words after he finished reading the front-page articles that he was most interested in, but he continued to read the rest of the newspaper. The most salient component of this procedure was the reading that the subject did after he stopped looking up new words for the day, since the pedagogical method being attempted in this case study was extensive reading (Grabe & Stoller, 1997).

This method was considered advantageous because it involved Bill's interest in Brazilian and international news, and also because it would naturally provide repeated exposure to the input as the news stories continued over the course of several days, while watching the evening news broadcasts provided further reinforcement of that input. Another advantage of using the newspaper as input was that Bill found it fairly easy to read (Grabe & Stoller, 1997).

This was not the case for the editorials, though, which the subject found too difficult to read until the fourteenth week of his stay, and even then it involved intensive consultation with the dictionary. Other sources of print input included: a weekly Portuguese comic book, which the subject began to read from week five, onward; a

Portuguese grammar textbook, replete with explications in English; a weekly Portuguese newsmagazine titled *Veja*, which the subject started reading from week eight, onward; and a sparse selection of other Portuguese newspapers and magazines. Bill tried to read short stories and novels in Portuguese, but he always found them to be too difficult for his level of reading proficiency. Consequently, he was never able to read Portuguese fiction throughout the course of the study (Grabe & Stoller, 1997).

The data for this case study was collected from the front pages of the *Journal do Brasil* newspaper, the subject's record of words that he looked up in the dictionary, a journal wherein the subject regularly reflected upon his daily reading experiences in Portuguese, and a battery of tests that assessed vocabulary, reading comprehension, listening comprehension, and a cloze test which was considered to test for production of the target language. These tests were administered at the end of each month, starting from the second month of Bill's twenty-one week stay in Brazil. This produced data from four separate administrations of the battery of tests (Grabe & Stoller, 1997). The data that this thesis is most concerned with comes from the results of the battery of tests.

The faculty of Pontificia Universidade Catolica do Rio de Janeiro designed the vocabulary test. The vocabulary tests were made from 800 content-words drawn from a Portuguese junior high school reader. The words were randomly sorted into groups of 200 for each testing session. When taking the test, Bill was given thirty minutes to write an English synonym for each of the 200 words. His answers were scored from 0 – 3, with 0 given for no recognition of a word, or an incorrect response, 1 for having a vague sense of the word, 2 for a good sense of the word, but not the best, and 3 for providing the most semantically and morphologically precise English synonym. The results from the

vocabulary test show that Bill's acquired Portuguese vocabulary nearly doubled from the first test measure to the fourth (Test 1, score of 3:  $N = 51$ ; Test 4, score of 3:  $N = 99$ ), while the number of Portuguese words on the test that Bill could not identify at all dropped by 35 (Test 1, score of 0:  $N = 119$ ; Test 4, score of 0:  $N = 84$ ) (Grabe & Stoller, 1997).

The reading comprehension tests used passages of approximately 400 – 500 words in length that Bill's collaborators took from Portuguese junior high school textbooks. For each administration of the test, Bill had thirty minutes to translate one 400 – 500 word passage into English. Two scorers provided a holistic assessment of Bill's translations with written commentary and a numerical value based on a 100-point scale. Based on this method of measurement, the data shows a radical improvement in Bill's Portuguese reading comprehension from the first administration of the test, to the fourth. The scorers rated the first test at 30% and the fourth test at 90%. In addition, the scorers noted that Bill's translations had not only become more accurate, but also more idiomatically sophisticated, as well, having succeeded in matching the original author's rhetorical style, whereas before, Bill's writing was encumbered with one-to-one translations at the sentence level (Grabe & Stoller, 1997).

The listening comprehension tests involved a three-minute compilation of assorted television and radio news broadcasts, with either male or female voices. Some of the excerpts were reporting on the same news story. The subject listened to the recording three times. The first time, he could not stop the tape but he could take notes. On the second play through, he had to determine the number of unique news events reported on in the recording and write a summary of each news event with as much detail as possible.

He was allowed to stop the tape but he could not rewind it. The summaries were given an overall percentage score based on a ratio of stories that were deemed to have been accurately summarized against the total number of news stories reported on the tape. On the third play through, Bill had to answer several comprehension questions about the content of the recording. Once more, he was allowed to stop the tape but not rewind it. The comprehension questions received a percentage score based on the ratio of correct answers to the total number of questions.

For all four of the test sessions, Bill never had significant trouble identifying the correct number of news reports from the recording, but on the first administration of the test, his written summaries of the news stories were scored at 5 out of 12, or 41.6% accurate, while his comprehension questions were scored at 7 out of 35, or 20% correct. On the third administration of the listening comprehension test, Bill summarized 100% of the news stories accurately, while scoring 80.5% on the comprehension questions (on the fourth administration of the test, Bill only scored a 75% for accurately summarized news stories. The researchers explained this as being due to the overconfidence of the test administrators, which influenced the decision to incorporate a trick in the fourth administration of the test, whereby a very similar – yet still unique – news story was included in the recording. The score for correctly answered comprehension questions on the fourth test was 37 out of 44, or 84.1%) (Grabe & Stoller, 1997).

A cloze test is a type of measure for language proficiency that involves deleting words, at a predetermined interval, from a complete, authentic, target-language passage, thus prompting the one taking the test to produce the words that were deleted. The cloze tests used passages from Portuguese junior high school textbooks. The first and last



sentences of each passage were left intact, but every seventh word in-between was deleted. This created 50, 43, 48, and 50 blanks left to be filled in by the subject for tests 1–4, respectively. The cloze tests were scored on a basis of one point for each semantically and morphologically accurate word that the subject produced, half of a point for a semantically correct word that was morphologically incorrect, or spelled wrong, and no points for a semantically incorrect answer. Throughout the course of the study, Bill never improved his performance on the cloze tests (44% 46.5%, 43.7%, 35%; tests 1–4, respectively). Since the progress of Bill’s performance on the other three tests was quite impressive, as shown, the researchers attributed this to the cloze test being a measure of language production, which they speculate is likely not closely related to reading comprehension.

As a case study, these results cannot be generalized. However, it is clear that the subject made substantial gains in vocabulary acquisition, reading comprehension, and listening comprehension, and by far the most significant source of input from the target language came from reading. These results cannot be ignored and they demand more research into the relationship between extensive reading and foreign and second language instruction. Although this thesis is concerned mainly with participants’ reading comprehension, reports of the peripheral benefits to language acquisition that extensive reading brings are nonetheless encouraging. Studies like Grabe and Stoller (1997) are the reason why the researcher of the present study chose to investigate extensive reading for this thesis.

The pilot study that this thesis reports on attempted to compare extensive reading treatment groups to control groups that did not read extensively. Therefore, this literature

review will now consider previous studies of similar design.

### **Comparison studies of the benefits of extensive reading for reading comprehension**

A critical way of determining whether extensive reading pedagogies are effective in improving various measures of students' overall proficiency in the second language is to conduct comparison studies involving pretest and posttest measures of an extensive reading treatment group and a control group receiving the standard curriculum of instruction. Six such studies are presented in this section in chronological order of publication: Mason & Krashen (1997); Yang (2001); Maxim (2002); Hitosugi & Day (2004); Tanaka & Stapleton (2007); and Lee (2007).

Mason and Krashen (1997) conducted a series of three studies involving the application of an extensive reading pedagogy to one class of Japanese college students, comparing pre- and posttest measures to another class receiving the standard curriculum of instruction using intact English foreign language (EFL) classes at the International Buddhist University in Osaka, Japan. In the first study, the standard curriculum group was comprised of second year students in the general education curriculum, while the extensive reading group was comprised of remedial students in their second, third, or fourth year of college who had previously failed EFL classes (Mason & Krashen, 1997). This study followed both groups over the course of two semesters. The method of instruction for the standard curriculum group involved readings with comprehension questions, vocabulary and grammar exercises, and translation exercises. The participants in the standard curriculum group were accustomed to this style of instruction, and excelled in accordance with the standard curriculum. The extensive reading group received the same style of instruction as the standard curriculum group in the first

semester, but most students exhibited lackluster attendance records and neglected to complete, or turn in many assignments. The extensive reading pedagogy was then applied to the remedial EFL students in the second semester. The researchers made 100 graded readers available to these students, and required each of them to read fifty of the books during that semester. Additionally, the students in the extensive reading group were required to write short summaries of the books they read, and to maintain a journal in Japanese wherein they reflected on the experience and progress with the readers (Mason & Krashen, 1997).

For this first study, the researchers used a 100-item cloze test for both the pre- and posttest measurements. To reiterate, a cloze test is a type of measurement of language proficiency that involves deleting words from an authentic target-language passage at a predetermined interval, thus prompting the one taking the test to produce the word that was deleted. The standard curriculum group had outperformed the extensive reading group in the pretest measure, which was expected given both groups' previous academic performance, but the posttest showed that the extensive group had nearly eliminated the difference between the mean scores of both groups. The gain that the extensive reading group had made in the mean test score from pretest to posttest was significantly greater than the gain achieved by the standard curriculum group ( $p < 0.025$ ) (Mason & Krashen, 1997).

The second study that Mason and Krashen (1997) conducted followed four intact classes over the course of two semesters. Two of these classes were from the English literature department of a prestigious four-year university, and the other two were from a junior college. Each institution had one extensive reading group and one standard

curriculum group. The design of this study was largely the same as the first study, with the extensive reading group at the four-year college moving from graded readers to authentic texts at the start of the second semester. In addition to the same 100-item cloze test as used before, another measurement was applied to both extensive reading groups that involved them in writing summaries of the first and last books that they read (the pretest and posttest, respectively). A panel of three native-speaking judges rated the summaries without knowing whether they had been written at the beginning or end of the semester. The judges were instructed to rate the summaries simply as either “good,” “average,” or “not good,” however they saw fit.

The mean scores of the cloze tests in pretest and posttest measurements reveal that both groups from the four-year university outperformed the groups from the junior college, but the gains that each respective extensive reading group had made against their associated comparison groups were statistically significant at the  $p < 0.001$  level. The scores from the writing summaries were categorized according to each of the three judges, and quantified by counting the number of summaries each judge rated as “good.” For each judge, the difference between the number of pretest summaries rated as “good,” and posttest summaries rated as “good,” was statistically significant (Mason & Krashen, 1997).

The third study that Mason and Krashen (1997) did involved three intact classes from the first author’s institution. There were two extensive reading groups and one comparison group: the first extensive reading group wrote book summaries in English, throughout the academic year; the other extensive reading group wrote these book summaries in Japanese; for the comparison group, a specific class was purposefully

chosen because the students in it had extensive practice with cloze exercises, and they appeared to be the most proficient group of students among their peers. The extensive reading pedagogy also differed from the last two studies in that the extensive reading groups had overwhelmingly more books made available to them.

The measurements in the third study consisted of the same 100-item cloze test as in the first two studies, a reading comprehension test, and a written book summary evaluation. The reading comprehension was designed to be rigorous, with 20 pages of text excerpted from an intermediate level graded reader (1600 word level), which the participants had one hour to read. Then, the participants were given 30 minutes to answer 50 multiple choice comprehension questions. The written summary assessments measurement was of a similar procedure to the second study, with the exceptions that there were only two native speaking judges, and after rating each summary “good,” “average,” and “bad,” three piles were made, accordingly, and then the summaries within each pile were reevaluated as either “good,” or “bad,” and further subdivided for a final total of six classifications of writing quality (Mason & Krashen, 1997).

The results from these measurements reveal that both extensive reading groups made greater gains in the average score of the cloze test, from pretest to posttest, but only the extensive reading group that wrote book summaries in English throughout the course of the year made a statistically significant gain against the comparison group. There was no statistically significant difference between both extensive reading groups for the cloze test measure, however. For the reading comprehension measure, both extensive reading groups outperformed the comparison group and were not significantly different from each other. The statistics produced from the writing assessment measures show that both

judges agreed that the extensive reading group that wrote summaries in Japanese throughout the year made a statistically significant gain against the comparison group, as well as the extensive reading group that wrote summaries in English (Mason & Krashen, 1997). Yet, a statistical analysis of the first judge's ratings reveal that the extensive reading group that wrote summaries in English achieved a lower average gain compared to the standard curriculum group in both pretest and posttest measures, while the second judge's ratings show this extensive reading group made a greater average gain. Neither of these measures was statistically significant, however. The authors provided two possible explanations for this anomaly: one relates to the different standards of assessment held by the judges, and the other relates to the significantly lower mean score of the first extensive reading group on the pretest summaries compared to the other two groups (Mason & Krashen, 1997).

Mason and Krashen's (1997) series of three comparison studies provides strong evidence for the positive effect of extensive reading on L2 reading comprehension as well as on writing proficiency, which is why the researcher chose to investigate extensive reading for this thesis. Once again, it should be noted that the present study is concerned only with how extensive reading will improve the reading comprehension of its participants. However, the improvement in writing performance achieved by the overwhelming majority of Mason and Krashen's (1997) participants provides general justification for extensive reading as an approach to language instruction.

More evidence is available from Yang (2001) that extensive reading has a positive effect on various measures of language acquisition. Yang's (2001) participants were ESL adults in Hong Kong with full-time careers who were enrolled in a fifteen-week English

course at the School of Professional and Continuing Education at the University of Hong Kong, in the Summer of 1999. These participants were enrolled in the course either for simple, personal interest, or for the benefit of their professional lives. There were four groups involved in this study: two groups that each read two particular mystery novels in their personal time, while receiving regular classroom instruction from an English grammar textbook, and two groups that also received such instruction, but did not read the novels. The researchers designated these classes A, B, C, and D. Classes A and B read two English mystery novels: Agatha Christie's *A Murder is Announced* (1950) and *They Do It With Mirrors* (1952). The classes were held in once weekly, three-hour sessions, and class procedures were largely the same for all four classes, with the exception that classes A and B would spend the first hour of class discussing the plot of the novels and how it related to contemporary social issues. Additionally, once every two weeks, classes A and B would write an open-ended response in class about what they had read in the novels, up to that point.

Yang (2001) used a multiple-choice test that assessed English grammar use to measure the language proficiency of the participants. The multiple-choice test had 100 items and was intended to measure the participants' awareness of English grammar and sentence structure. The participants had one hour to complete the test. Pretest scores on the multiple-choice grammar were nearly the same for all four classes, but the posttest scores revealed that the treatment groups significantly outperformed the control groups ( $p < 0.01$ ) (Yang, 2001).

Like Mason and Krashen's (1997) experiments, Yang's (2001) research brings strong statistical evidence to the body of literature supporting the positive effect of

extensive reading on second language acquisition. Studies like these are what motivated the researcher of this thesis to try an extensive reading pedagogy with his research participants because he wanted to improve their reading comprehension scores on the TOEFL. Now this literature review will continue with consideration of two comparative studies that used extensive reading pedagogies in foreign languages other than English: Maxim (2002) and Hitosugi and Day (2004).

Maxim (2002) conducted a comparison of two groups of students in their first semester of German foreign language study at the University of Texas at Austin, in the Spring of 1997. The treatment variable was an authentic German romance novel, 142 pages long. Hence, Maxim's research is a breathlessly bold attempt to apply extensive reading to students possessing a nascent level of proficiency in all areas of the target language, whereas conventional wisdom assumes that authentic texts are too difficult for all but the most advanced of second language learners. The novel that the treatment group read was *Mit dem Sturm kam die Liebe* (With the Storm Came Love) by Marianne Andrau (1981). Although it was originally written in English, and subsequently translated into German, the researcher deemed it to be an authentic text because it could be found on newsstands in Germany.

The class meetings for both the treatment group and the comparison group lasted for one hour. Both groups followed the same syllabus prescribed by the foreign language department, for the first four weeks. The prescribed syllabus involved use of a workbook that provided explicit lessons, with some illustrations, and exercises in German grammar and vocabulary. Class time was spent reading brief textual dialogues from the prescribed textbooks and communicative language practice with the words and grammar



constructions introduced in the lesson. After the fourth week, both groups continued to follow the prescribed syllabus, with the exception that the treatment group spent the first half hour of class reading from the German romance novel (Maxim, 2002).

As mentioned, the application of extensive reading to language beginners is unusual. Indeed Maxim (2002) acknowledges that he violates several conventions of extensive reading approaches. Reading the novel was done out loud, in class, in pairs or groups. The classroom teacher guided language learning from the novel. He would frequently direct the students' attention to major events in the narrative and how the text was used to introduce them and evoke the imagery. This approach also incorporated student reproduction of the textual input orally and in writing (Maxim, 2002). The participants' nearly total lack of proficiency in German likely made these unorthodox methods necessary.

Two methods of pretest and posttest measurement were used. One was the standard test prescribed by the foreign language department, and the other was introduced specifically for Maxim's (2002) experiment. The departmental tests focused on the material introduced in the textbooks. The test developed specifically for this study involved four brief readings from different genres: an advertisement, a newspaper editorial, a sample from a romance novel, and a commentary on culture from a German textbook. After reading each passage, the participants had to write what they could remember from it in English. When finished with that task, the participants were asked to define certain words in the passage and answer grammar questions about the text (Maxim, 2002).

Statistical analysis of the scores taken from these tests reveals that both groups made statistically significant gains on all measures of German language proficiency, but the differences between the gains made by each group were not statistically different from each other. The implication of this is that extensive reading from an authentic text is not detrimental to the language acquisition process when the language learners are at a nascent level of proficiency. On the other hand, Maxim's (2002) research would seem to agree with what has already been presented in this literature review with Pichette's (2005) research, specifically that extensive reading is only more effective than other methods of instruction when students already possess a sufficiently high level of language proficiency. It will be interesting to see whether this holds true for the participants that this thesis reports on. Regardless, Maxim's (2002) research alleviates any concern that trying an extensive reading pedagogy on less proficient students might hinder their progress in the language acquisition process. If nothing else, this clears the way for considering extensive reading purely for the affective benefits of enthusiasm and fondness for reading, which will be considered in the next section. Three more comparison studies are presented now.

Hitosugi and Day (2004) conducted a comparative study investigating the effects of extensive reading for the acquisition of Japanese at the University of Hawaii. The participants of this study were two classes of students, from among multiple sections, who were in their first year and second semester of Japanese language study. These classes met five times a week, for sixteen weeks, although the extensive reading treatment was only applied over ten weeks. The class receiving the extensive reading

treatment was taught by the first researcher, but it is unclear why only one class was chosen for the comparison group, or why it was preferred to any of the other sections.

The extensive reading component of the treatment group's instruction was in addition to the standard prescribed syllabus that the comparison group received. The researchers decided that the extensive reading treatment would be applied with a library of 266 authentic Japanese language children's books, since they did not know of any published series of graded readers for learning Japanese. The instructor of the treatment group made it a course requirement for the students to read up to 40 of the books for a full 10% portion of the students' final grades, but partial credit was awarded for students who read fewer books at various intervals. In order to adhere to the prescribed syllabus, the students were instructed to do their reading as homework outside of class, although 30 minutes of class time was reserved once a week for students to share their reading experiences through simple class presentations (Hitosugi & Day, 2004).

To measure the Japanese reading comprehension abilities of the participants in both groups, the researchers synthesized a three-part test, culled together from exams used in previous semesters. For part A, students had to read a memorandum and write a short summary of the content in English. Part B presented an edited excerpt of a travel brochure for a hotel in Hawaii, and gave students the task of citing five of the hotel's amenities in English. Part C was the most challenging part of the test, according to the researchers, requiring the participants to read a letter and answer five comprehension questions in English. The total number of points possible for each of these three parts was 5, 5, and 10 (Hitosugi & Day, 2004).

The researchers did not conduct a thorough statistical analysis on the data from the pretest and posttest measures because their sample sizes were too small (extensive reading group,  $N=14$ ; comparison group,  $N=10$ ), and also because “the purpose of administering the two measures was for the enrichment of learning and the evaluation of the ER program, not to be research per se” (Hitosugi & Day, 2004). Still, the researchers do provide the mean scores and the gains that the participant groups made on each section of the pretest and posttest. The gains, from pretest to posttest, that the extensive reading group made for the average score on each of the three sections of the test are 0.20,  $-0.01$ , and 0.88, respectively. For the comparison group, those numbers are 0.46, 0.04, and  $-0.13$ . This data shows that while the extensive reading group gained nearly one point (out of a possible 10) on average for Part C of the test, the comparison group’s average score decreased. In the opinion of Hitosugi and Day (2004), since Part C was the most difficult part of the test, this indicates that the extensive reading treatment enhanced the reading comprehension proficiency of those in the treatment group.

The participants involved in the pilot study for this thesis are also too few to warrant a thorough statistical analysis. However, if the researcher finds results similar to those presented by Hitosugi and Day (2004), it would extend the call for more robust investigations in both contexts. In the meantime, if the reader is still not convinced of the promise that an extensive reading approach holds for the participants of the present study, this review of the literature will present two more comparison studies in the EFL context.

Tanaka and Stapleton (2007) initiated a study of 226 Japanese high school students, aged 15–16, in six different classrooms to see whether an extensive reading program could improve the EFL reading proficiency of Japanese students. This study

took place between October 2003 and March 2004. Three classes served as the control group, while the other three represented the treatment group. Both groups received in-class instruction in EFL reading through the use of an authorized textbook in coordination with in-class exercises designed by the researchers. In addition to this standard instruction, the treatment group was also given a series of graded readers designed by two different publishers and some materials designed by Tanaka himself. In designing these materials, Tanaka attempted to adhere to previously established guidelines that students should be able to recognize 95% of the words contained in any texts used for extensive reading. Due to the emphasis of official school policy on grammar-translation methods and textbook usage, most of the extensive reading had to be conducted as homework. Consequently, the researchers report that only eighteen participants from the treatment group ended up reading both the graded readers and the researcher-designed materials. The rest of the treatment group only read the researcher-designed materials outside of the classroom (Tanaka & Stapleton, 2007).

For the pretest and posttest measurements, Tanaka and Stapleton (2007) used the reading comprehension component of the STEP test, developed by the Society for Testing English Proficiency, which is the official test instrument authorized by the Japanese Ministry of Education, Culture, Sports, Science, and Technology. The STEP test assesses English ability at seven levels of proficiency. For this study, the researchers created one test from parts of the third and fourth level tests for their pretest and posttest measures. Since some of the participants in the treatment group did not make use of all of the reading input provided for them, the researchers decided to apply a filter to the test data from the treatment group according to those that had read both the graded readers

and the researcher-designed materials, and those that only read the researcher-designed materials. Data analysis reveals that the treatment group significantly outperformed the control group on the reading comprehension test ( $p = 0.01$ ). However, with the filter applied, the treatment group still outperforms the control group, though the difference is not statistically significant ( $p = 0.19$ ). Therefore, the participants who had read both the graded readers and the researcher-designed materials were responsible for bringing the greatest gains among the treatment group (Tanaka & Stapleton, 2007).

The research of Tanaka and Stapleton (2007) reinforces the strong statistical evidence for the beneficial effects of extensive reading already demonstrated to exist by Mason and Krashen (1997) and Yang (2001). Accordingly, the author of this thesis feels that it further justifies his attempt to apply an extensive reading treatment to his own participants. This review of the literature will now present one more comparison study before moving on to present the evidence for the affective benefit of extensive reading.

Lee (2007) conducted a series of three studies of extensive reading with EFL students in regularly scheduled classes at the National Taipei University in Taiwan. For the first study, the researcher was brought in to teach in the classes in the middle of the academic year, due to the poor academic performance of the students in the EFL class that was to become the extensive reading class. Thus the first study only took place over the course of one semester. The study involved one extensive reading group and two comparison groups. The extensive reading treatment, itself, lasted twelve weeks. The participants in the extensive reading group had access to a library of 215 graded readers. Once a week they spent a portion of the class reading silently and then writing summaries of what they had read. For the rest of the class, and for every other meeting, theories

about second language acquisition were presented and discussed in order to help the participants understand the purpose of reading extensively. The first comparison group used a textbook of reading and writing exercises, and the second comparison group followed the same syllabus, but was encouraged to engage in extracurricular reading activities, though no attempt to account for such reading was made (Lee, 2007).

The researcher used a vocabulary test and a cloze test to take pretest and posttest measures. The vocabulary test measured gains at the 2000, 3000, and 5000 word levels. The posttest scores were adjusted by ANCOVA on account of the pretest proficiency deficit of the extensive reading group. The results show that the extensive reading group had made greater gains, on average compared to the group whose instruction focused solely on the textbook, while the group that was encouraged to read independently, outside of class time made slightly greater average gains at the 3000 and 5000 word level, compared to the extensive reading group. There were no statistically significant differences between these groups, though. As for the cloze test, the extensive reading group achieved a statistically significant greater average gain from pretest to posttest than the first comparison group. The second comparison group's average gain on the cloze test was slightly more than the extensive reading group's average gain on the cloze test, though it was not statistically significant (Lee, 2007).

The researcher offered three explanations for these results. One was that the treatment group was not at the same level of language proficiency as either of the comparison groups, and another was that the graded readers used in the extensive reading treatment were written at the 2000 and 3000 word levels. The third explanation was that

this experiment only lasted one semester, and so was not enough time for the treatment group to fully flourish from the effects of extensive reading (Lee, 2007).

In the second study, the researcher compared one freshman English class receiving the extensive reading pedagogy to three other freshman English classes representing the comparison group. While the first study was flawed because the extensive reading class was not equal to the comparison classes in pretest measures of EFL proficiency, the pretest measure for the second study showed that all of the participants generally had the same level of English language proficiency and there was no statistically significant difference between them. The three comparison groups were thereby amalgamated into one comparison group. The second study was also an improvement over the first study in that it would proceed for the entire academic year, as well (Lee, 2007).

The extensive reading program was improved to include a total of 570 graded readers in the first semester. In the second semester, the extensive reading group moved to using authentic texts for extensive reading. The researcher specifically chose five such texts as required reading for the course. Each participant also had to choose two more books to read from a list of suggestions (Lee, 2007).

Classes met once a week for three hours. The students in the extensive reading group spent half of the class in sustained silent reading, with the remainder of the time being spent interacting with their fellow classmates in paired reading activities, giving presentations of what they read, and writing reflections on their readings in either English or Chinese. The comparison group followed the standard curriculum of instruction, which was centered on the use of a textbook (Lee, 2007).



The posttest measure of vocabulary for the second study differed from the first in that it tested for vocabulary at the 2000, 3000, 5000, 10000, and “academic” word levels. Results from this test show that the extensive reading group made greater gains for the average score at each vocabulary level than the comparison group with statistical significance at the 3000 ( $p < 0.007$ ), 5000 ( $p < 0.012$ ) and 10000 ( $p < 0.00013$ ) word levels (Lee, 2007). Results from the cloze test, however, are much less impressive showing that both groups achieved almost the exact same measure of gain from pretest to posttest (Lee, 2007).

The design of Lee’s (2007) third study was largely the same as the second study with the exception that the extensive reading treatment was much more liberal. The graded readers were not used at all and neither was there any number of authentic texts that the researcher specifically required the participants to read. Instead, for the entire academic year, the participants simply chose self-selected texts from a library of approximately 1200 titles. Lee (2007) makes no mention of a minimum requirement for the number of books that the participants should have read. The same instruments were used to take measurements, but the data from the extensive reading group in the second study was included in the statistical analysis.

Posttest data of vocabulary levels from the third experiment reveals the same pattern of superiority and significance of the extensive reading group over the comparison group. The treatment group for the third experiment also made greater average gains from pretest to posttest at every vocabulary level, except for the 2000 word level, compared to the treatment group from the second experiment, but the differences were not significant. As for the cloze test, the treatment group from the third experiment

made a significantly greater gain in test performance than both the comparison group and the treatment group from the previous year. Lee does not offer an exact statistic, but the data provided shows that the gain that the third year treatment group made over the second year treatment group on the cloze test is significant at the  $p < 0.00$  level (2007).

With these results, Lee asserts how essential it is to give students the freedom to direct their own reading endeavors, in order to get the most benefit from extensive reading pedagogies. Doing so will allow them to harness intrinsic motivation (2007). This was an important point for the researcher of this thesis to consider when designing his own attempt at implementing an extensive reading program.

The comparison studies conducted by Mason and Krashen (1997), Yang (2001), Tanaka and Stapleton (2007), and Lee (2007) offer powerfully convincing statistical evidence of the effect extensive reading on many areas of second language acquisition, especially measures of reading comprehension. Studies such as these are what provided the impetus for the researcher of this thesis to attempt an extensive reading pedagogy with his research participants because he wanted to improve his participants' reading comprehension scores from the TOEFL. However, the comparison studies conducted by Maxim (2002) and Hitosugi and Day (2004) revealed that their participants who underwent extensive reading pedagogies did not perform better than comparison groups on posttest measures. These researchers point out, though, that the extensive reading treatment groups performed no worse than the comparison groups. Even if statistical analyses did show no difference between the quantitative performance of extensive reading treatment groups and groups receiving the standard curriculum of instruction, there is still a reason to prefer the extensive reading approach, and that reason is the

affective benefit of extensive reading. Qualitative research shows that many students develop positive attitudes about reading through extensive reading programs, which, as Lee (2007) points out, can help teachers to tap the intrinsic motivation of their students. This thesis will now present findings from the research literature that demonstrates the affective benefit of extensive reading for language learning.

### **The affective benefit of extensive reading pedagogies**

Extensive reading not only has the potential to improve the reading comprehension skills of the second or foreign language learner, it can also foster favorable dispositions toward the task of reading. This section will review studies in the body of reading research literature that report on this potential advantage. Due to the singular character of this consideration, and the relative brevity of this section, each study will first be reviewed in sequence, and then their implications for the present study that this thesis reports on will be considered in aggregation. These studies are presented in chronological order of publication: Mason & Krashen, (1997); Yang, (2001); Hitosugi & Day, (2004); and Greenburg, (2006).

Mason and Krashen's (1997) three experiments involving EFL students at a university in Osaka, Japan has already been mentioned in this chapter for the quantitative evidence it offers showing the benefit of extensive reading toward the improvement of reading comprehension skills. The extensive reading treatment for the first experiment involved students who had previously failed EFL reading classes. In the first semester, this group followed the same standard curriculum of instruction as the comparison group, but their academic performance under this style of instruction was generally very poor. However, the researchers report that after a semester of experiencing an extensive

reading approach, many of these students became enthusiastic readers. One of the ways in which qualitative data was collected from this experiment was from the journal writing that the participants in the treatment group were required to do. The researchers report from these journals that some participants expressed amazement at how much their reading comprehension ability had improved (Mason & Krashen, 1997).

The participants in Yang's (2001) study were full-time career professionals living in Hong Kong. The extensive reading treatment involved the participants in reading authentic English-language mystery novels. Yang extracted qualitative data from the participants who read the novels through a questionnaire and personal interviews. The questionnaire had twenty items which participants responded to using a four-point Likert scale: I = strongly agree; II = agree; III = disagree; IV = strongly disagree. (2001) Yang considers the top four items receiving the most affirmative answers to be of particular interest: #15, "In general, I understand the plot of the story"; #2, "On average, I read at least 15 minutes of English for different reasons every day"; #12, "The novel discussion promotes more English learning opportunities"; and #20, "I benefit from the reading experience in this class." (2001) In Yang's words, "the fact that items 15, 2, 12, and 20 ranked high shows that students were able to read an authentic text designed for native speakers and that they felt positive about the learning opportunities on offer" (2001).

Yang reports from the interviews that were conducted with some of the participants in the extensive reading group that they greatly appreciated the in-class discussions of the mystery novels because it felt more like authentic conversation in English to them, and they also commented that "answering grammar questions or making sentences in class was boring," as an alternative (2001). Yang further states the interview

participants believed that reading an authentic text helped to improve their grammar and that they felt pride in being able to completely read two English-language novels within three months' time (2001).

The pilot study of Hitosugi and Day (2004) was also previously mentioned in this thesis. In that study, the authors report on a Japanese-language extensive reading program that they implemented at the University of Hawaii. In addition to comparing the average test scores of the treatment group to a control group being taught by a traditional method of instruction, Hitosugi and Day also administered a 22-item questionnaire with a four-point Likert scale with an orientation much like the one mentioned in Yang (2001). The questionnaire was administered twice as pre-treatment and post-treatment measures. In order to quantify the responses for affective measurement increases, each response was assigned a point value according to the degree to which it expressed a favorable disposition toward the Japanese language. So, for example, where item #1 asks, "So far, I am enjoying Japanese 102," a response of "A – Strongly Agree" would garner four points, "B – Agree" three points, "C – Disagree" two, and "D – Strongly Disagree" one point. Point values for question items that were negatively oriented, such as #15, "I do not enjoy reading Japanese," were inverted along the same lines (Hitosugi & Day, 2004).

The results from Hitosugi and Day's questionnaire seem contradictory in some respects. For example, while the gain in the average affective measurement of the extensive reading class over the comparison class for item #1 was 0.21, the comparison class made an average gain of 0.09 in affirmative responses to item #15, while the extensive reading class showed no gain at all. Similarly, question items #3, "It is easy for me to read Japanese," and #8, "I have confidence in my ability to read Japanese," show

the extensive reading class registering substantial losses in their average affirmative affective measures versus the comparison class ( $-0.55$  and  $-0.50$ , respectively). Hitosugi and Day explain these responses as being due to the fact that students in the extensive reading classroom faced a greater academic challenge in reading authentic Japanese children's books than the comparison class, which only used the textbook throughout the course of the study. In spite of this, question item #5, "I read Japanese books, comics, newspapers, etc., outside of class," shows the treatment class making an average gain of  $0.84$  over the comparison class, while question item #13, "I would like to read more Japanese," shows a  $0.15$  average gain. Hitosugi and Day conclude that, overall, the extensive reading class "developed more positive attitudes toward their study of Japanese than did the students in the regular class by the end of the semester." (2004)

Greenberg, Rodrigo, Berry, Brinck, & Joseph (2006) conducted a study of twenty-seven adults who had been measured to possess a reading proficiency between the third grade and fifth grade level. Twenty-two of these adults were native English speakers; the rest spoke English as a second language. The average age of the participants was 39 with the age range being 17 to 63. The instructors held scheduled class meetings with their participants once a day for two hours, four days a week, for thirteen weeks. There were about five students per class. The classroom had a mini-library stocked with books according to the preferences of the participants, which they indicated in a survey prior to the start of the study.

The instruction for Greenberg, Rodrigo, Berry, Brinck, & Joseph's (2006) study consisted entirely of extensive reading activities; there was no textbook or attempt to teach grammar, and there were no comparison groups that did not engage in extensive

reading. The method of Greenberg, et al. proceeded in the following way: at the beginning of class everyone would speak for about one minute about what they had read since the last class meeting. Then, everyone would engage in sustained silent reading with the book of his or her choice. The goal was to eventually spend eighty minutes of class time in sustained silent reading; however, the researchers reasoned that this would, at first, be too demanding for the participants, so initially, sustained silent reading lasted about twenty minutes. After the silent reading session, everyone would talk about what they had read for about a minute. The next activity was a fifteen-minute read-aloud led by the teacher and each participant silently following along with their own copy of the book. Sometimes the teacher would provide the students with a vocabulary list as some of the books used in this study went up to an eighth grade reading level. After the read-aloud, the teacher and the students would casually discuss the reading for a few minutes, and return again to sustained silent reading. If, after once more sharing with each other what they had read, there was still time left for the class, they would choose either to play some type of vocabulary word game, or discuss background information about the story being read aloud, such as the author, setting, or the historical context of the narrative (Greenberg, et al., 2006).

Greenberg, Rodrigo, Berry, Brinck, & Joseph (2006) extracted qualitative data with a short survey and interviews. Like Hitosugi and Day (2004), the survey was administered pre-treatment and post-treatment. The researchers' data shows that for the survey item "I understand most of what I read," the pre-treatment rate of response was 53% in the affirmative; the post-treatment rate of response was 78% affirmative. For the survey item "I read slowly," the pre-treatment rate of response was 38% affirmative; the

post-treatment rate of response was 8% affirmative. Greenberg, et al. describe this data as being indicative of her participants' increasing confidence in their reading abilities (2006).

Greenberg, Rodrigo, Berry, Brinck, & Joseph (2006) also provide a list of quotations from their interviewees where they expressed positive attitudes about reading: "I liked the reading time and then telling what you read because it gave me a chance to let everyone know what I was reading;" "For the first time, I went to the library, checked out the books, and got so caught up in the book that I missed my train stop;" "The read-aloud showed me how to read stories better and make them more interesting;" "At first I didn't like to read but when I started to read books that were interesting, I read longer than I thought," (Greenberg, et al., 2006) and so on.

What the researchers report from these four studies should make one thing very clear: extensive reading fosters positive attitudes about reading in general. This is an important point to consider even in the absence of significant statistical evidence indicating increased reading comprehension abilities because teachers, of course, want their students to be motivated and enthusiastic for their learning efforts. It is with this assurance that the author of this thesis attempted to devise and implement his own extensive reading pedagogy.

## **Summary**

Education professionals can observe a positive correlation between the amount of time their students spend reading and the level of their academic performance. Statistical analyses strongly indicate a causal relationship between these two variables. Moreover, comparison studies that implemented extensive reading treatments under ideal conditions



have shown that extensive reading groups consistently and significantly outperform control groups. Some comparison studies attempted to implement extensive reading treatments under less than ideal conditions and consequently show no statistically significant differences between extensive reading groups and comparison groups. Nevertheless, the researchers of such studies point out that the treatment groups performed no worse than comparison groups did on posttest measures. This means that as long as students respond favorably to extensive reading in terms of their motivation and enthusiasm for academic pursuits, there is still sufficient reason to take an extensive reading approach. Qualitative research bears this purpose out as well. The researcher of this thesis hopes that with this review of the scholarly literature, the case for trying an extensive reading treatment to improve the reading comprehension abilities of the University of Toledo's ALI students has been soundly made.

## Chapter 2

### Context and Design of the Present Study

#### Introduction

In Chapter 1, the review of the literature produced about the positive effects of extensive reading sought to make the case for trying an extensive reading pedagogy with the University of Toledo's ALI reading labs. This chapter will introduce the researcher and the participants of this study and the setting in which this study took place. This chapter also provides a detailed methodology so that the reader will know precisely how the study proceeded and how the data was obtained. Before getting into such specifics, however, this chapter will couch the present study within a conceptual framework about the relationship between the pure act of reading in a second language and the skill of L2 reading comprehension.

#### Theoretical Assumptions for Research

In launching this investigation into the potential benefits of an extensive reading approach for the University of Toledo's ALI students, the author of this thesis references ideas about second language reading and acquisition, as expatiated on in two landmark texts, written by reputable researchers in the field of second language acquisition.

Palmer's *The Scientific Study and Teaching of Languages* was originally published in 1917. It was born of his desire to provide foreign language teachers with a formal credo and disquisition (Harper, 1968). Palmer's treatise seems to be the first mention of extensive reading in the literature of foreign and second language pedagogy. It occurs in the text where Palmer begins to discuss the curriculum of the advanced language learner. As Palmer explains, this is "the point at which the scholastic study

should be replaced by those types of study which consist of using the language in social intercourse, and applying the language to intellectual ends” (1968). In other words, once the language learner has progressed sufficiently far in formal classroom instruction, thereby having the ability to successfully navigate in an authentic target language environment, she may assume greater autonomy over the direction of her own language learning. Palmer continues, “What is solely required during the final period is quantity. To this effect, the student must read books...” (1968). To this day, professionals in the field of L2 reading research still agree with Palmer. Krashen, for example, has found the notion of extensive reading for second language acquisition to be quite compatible with his own theories.

Krashen is well known in the field of Second Language Acquisition for his hypotheses, which he expounded on in *Principles and Practice in Second Language Acquisition* (1982). Of particular relevance to the present study is his “Input Hypothesis” (1982). With the Input Hypothesis, Krashen uses a simple symbolic formula to illustrate the incremental progress that the language learner makes throughout the learning endeavor, where her present level of language proficiency is represented as  $i$ , and her next immediate level of language proficiency is  $i + 1$ . Therefore, the language learning enterprise is a process of moving from  $i$  to  $i + 1$ . According to Krashen, the way this process happens is “by understanding language that contains structure a bit beyond our current level of competence ( $i + 1$ ). This is done with the help of context...” (1982). Extensive reading can provide the language learner with large amounts of input from the target language, which is made comprehensible through the context of the surrounding text, previously incorporated into the L2 reader’s lexicon.

Thus the researcher of this thesis suspects that there is a point in the foreign or second language learner's progress where her L2 lexicon begins to expand exponentially, and it becomes increasingly impractical for the instructor to anticipate what novel lexical items and rhetorical constructions may be encountered, and to contrive scripted events in order to introduce them in formal classroom instruction. It is at this point that new L2 input is less likely to be formally introduced by the teacher, than it is to be introduced in authentic contexts, such as reading in the target language. Accordingly, the more the language learner reads, the more proficient she will become in her new language. Therefore, the purpose of this study is to compare two groups, where one was given more English L2 text to read than the other, to see whether this application of the idea of extensive reading will lead to better scores on an English L2 reading comprehension test. This study was conducted at the ALI on the campus of the University of Toledo.

### **Description of Context in which Data was Gathered**

The ALI is an intensive English language program located on the campus of the University of Toledo. It was established in 1977, for the purpose of improving the English language skills of foreign nationals in order to successfully engage with the university and greater Toledo community (<http://www.utoledo.edu/cisp/ali/mission.html>). A session of classes at the ALI typically lasts about seven weeks, or half of a regular semester of classes at the University of Toledo. Therefore, the ALI conducts two sessions of classes within each regular semester at the University of Toledo: Session I and Session II.

All foreign nationals (henceforth referred to as "the students") wishing to attend the University of Toledo (UT) must take the institutional version of the Test of English

Foreign Language (TOEFL) exam to determine their English language skill levels. The TOEFL is divided into three main parts. Section 1 of the TOEFL tests for English listening comprehension by playing a recording of two English speakers having a conversation for a few minutes and then having the student answer multiple-choice content questions about what was heard. Section 2 of the TOEFL tests for knowledge of English sentence structure and idiomatic expression in writing through a multiple-choice, fill-in-the-blank testing format. Each question item in Section 2 is one sentence with an omission of one or a couple of words. Section 3 of the TOEFL is an English reading comprehension test involving the reading of a passage of text about five hundred words long and then answering ten comprehension questions about the passage of text. The passages are suitable for a general reading audience and require no specialized knowledge. There are five passages of text for the institutional version of the TOEFL and fifty test questions for section 3.

The students must achieve a minimum score of 500 out of 677 on the TOEFL in order to gain full-time admission to the University of Toledo. Those students who score under 500 will be enrolled at the ALI until they qualify for full-time enrollment at UT. There is no uniform policy for matriculation out of the ALI, and into regular enrollment at UT: a complicated array of sets of criteria governs this process for different students under various circumstances. The reader of this thesis needs only to know that the simplest way for a student at the ALI to qualify for full-time enrollment at UT is to achieve a score of 500 on the institutional version of the TOEFL, and that, generally speaking, exceptions may be made for some students who achieve a score of at least 450,

provided that they meet satisfactory standards of other measures of academic excellence, such as grade point average (American Language Institute, 2007).

The researcher conducted the research for this thesis with students in the ALI's reading labs during the ALI's Session I and Session II of the Fall 2012 semester at the University of Toledo. The reading labs are comprised of students numbering from as few as three, to potentially as many eight, or more, depending on enrollment and attendance. Students meet with the instructor for fifty minutes, twice a week, either on Mondays and Wednesdays, or Tuesdays and Thursdays. The staff of the ALI prepare the reading materials for those reading labs. The reading materials usually consist of news articles printed off from online news media outlets. The reading labs begin with timed reading exercises intended to increase the students' reading speeds. These exercises usually last for about the first ten minutes of the class period. Then the instructor sits with the students and together they silently read the article provided for them. Twenty minutes is allotted for this activity.

For the remainder of the class period, the instructor discusses the content of the articles with the students. The instructor prompts the students to answer questions about the content of the articles in order to assess their comprehension of what they have read. Usually the staff of the ALI has specific comprehension questions prepared for the articles; however, instructors may also improvise and spontaneously pose their own questions during conversation. These exchanges take place using the target language, which is English

### **Research questions**

To investigate the effect of extensive reading on the English L2 reading comprehension performance of the ALI's students, the following research questions were pursued:

1. Will pretest and posttest measures of the English L2 reading comprehension of the participants of this study indicate a differential gain for the participants in a treatment group who read more than a comparison group who read less?
2. Will the participants respond favorably to the extensive reading treatment?
3. What test-taking strategies do the participants use when taking reading comprehension tests like the TOEFL?
4. What are the participants' opinions of the reading comprehension section of the TOEFL?
5. What are the participants' opinions of the timed reading exercises that take place at the beginning of the ALI's reading labs?

### **Background and role of the researcher**

The researcher of the present study is a graduate student of the ESL teaching program of the English department at the University of Toledo. He first conceived of the present study while he was enrolled in a course called *Issues in ESL Reading* during the Spring 2012 semester at UT. This study was conducted as a partial fulfillment of the requirements for obtaining a Master of Arts degree in English, with concentration in ESL. As a graduate teaching assistant, the researcher's duties involve teaching ESL writing classes, Basic ESL classes for communicative tasks, and reading instruction in the ALI's reading labs throughout the calendar year 2012.

### **Background and role of the participants**

The students in these reading labs were the participants of the present study and constituted the convenience sample that the researcher used for this thesis. The participants in this study range in age from 20–26 years old. There were seven participants in total. To preserve their confidentiality, the participants in the extensive reading group will be identified as Saudi Arabian male 1, Chinese male, Chinese female, and Taiwanese female; the participants in the comparison group will be identified as Saudi Arabian male 2, Saudi Arabian male 3, and Japanese female. The researcher coordinated with the administrative staff of the ALI to schedule him with two reading lab classes populated with the most proficient students. These classes are officially designated by the ALI as “Advanced” reading labs. The researcher specifically sought to draw participants from the advanced reading labs because, based on the literature reviewed in this thesis, it followed that only the most advanced students would be able to cope with reading authentic English texts and to receive the most benefit from it.

### **Methodology**

No data was collected from Session I. The researcher intended to arbitrarily designate one of the two reading labs as the treatment group and the other as the comparison group. Both groups were to receive the regular curriculum of instruction as previously described in this chapter. The treatment group would be unique only in the application of the extensive reading treatment. For Session I, the text *George Washington: How the Great Uniter Helped Create the United States* (Lacayo, 2011) was used for the extensive reading treatment. The researcher provided the members of the treatment group with one chapter of this text to be read outside of the scheduled class meetings, over the course of each week. However, the participants in Session I did not



read any substantial amount of this text. They complained either that it was too difficult or too boring to read. This led the researcher to revise his methodology for Session II.

**Revised methodology.** The extensive reading treatment for Session II incorporated a few of Day and Bamford's (2002) top ten principles of extensive reading. Some of these principles are that a wide variety of reading materials should be made available to students, the students choose what texts they want to read, the students should be encouraged to stop reading something they find boring and choose another text, and the teacher should set an example by choosing a text for himself to read, among others.

In accordance with these ideals, the researcher invested his own personal finances into assembling a very modest library of graphic novels, magazines, and collections of short stories for the participants to choose from. The graphic novels included adapted versions of *The Hobbit* (Tolkien, Wenzel, Dixon & Deming, 2012) and *20,000 Leagues Under the Sea* (Verne, Rafter & Ahluwalia, 2010). The magazines were one issue of *National Geographic* magazine (2012) and a magazine about ocean ecosystems called *Wonders of the Deep* (Life, 2012). One of the collections of short stories was *The Stories of Ray Bradbury, Volume 2* (1980). The participants in the treatment group were to choose any text that interested them at the beginning of Session II, and read as much of it as they could for the duration of the study. Once each week, the participants in the treatment group were asked if they wanted to exchange their selected texts for a different one. Three of the four participants in the extensive reading group exchanged their selected texts throughout the course of this five-week study. Saudi Arabian male 1 first selected *Twelve Angry Men* (Rose, 1997), which he later exchanged for *Pride of Baghdad*

(Vaughan & Henrichon, 2006). His reason for exchanging was that he was not enjoying *Twelve Angry Men*. Chinese male first selected the graphic novel version of *The Hobbit* (Tolkien, Wenzel, Dixon & Demming, 2012), which he later exchanged for the graphic novel version of *20,000 Leagues Under the Sea* (Verne, Rafter & Ahluwalia, 2010). His reason was that he had finished reading the graphic novel version of *The Hobbit*. Chinese female first chose the National Geographic (2012) magazine from the researcher's makeshift library, which she later exchanged for *Wonders of the Deep* (Life, 2012). Subsequently, she exchanged *Wonders of the Deep* for *Chicago Lightning: The Collected Nathan Heller Short Stories* (Collins, 2011). Her reason for each exchange was that she was finished with the magazine that she had at the time of each exchange. Taiwanese female selected *The Stories of Ray Bradbury, Volume 2* (Bradbury, 1980) but she kept this text throughout the course of the study.

No coercive measures were taken in order to force reading. The researcher believes that this is in accordance with principle number 3 of Day and Bamford's Top Ten Principles for Teaching Extensive Reading (2002) wherein Day and Bamford stress the importance of student initiative and independence, with minimal micromanagement on the part of the teacher, in an extensive reading program. However, a minimal amount of accountability for continued participation in the extensive reading program was attempted and it is described as follows: all four of the participants in the extensive reading group were expected to report on their reading activity on one of the two scheduled class meetings per week. These reports lasted only a few minutes each at the beginning of class. The chairs of the participants, and the researcher's chair, were arranged in a circle. The reports were given while seated. The researcher would ask the

participants to tell him and the other participants about what they read and they would usually mention the characters, setting, and the plot of the story. The amount of reading that the participants actually did will be reported on, along with all other findings from this study, in Chapter 3.

Most of the participants enrolled in both of the researcher's two reading labs for Session II did not show up for the first day of class. In fact, only one student showed up for each class. The researcher administered the informed consent forms and the pre-test measurement to the student in each class, but decided not to allow either of them to choose a text to read until he could meet with the rest of his would-be participants. Through the help of the administrative staff of the ALI, the researcher immediately took up the task of tracking down as many of the absentee students that were enrolled in his reading labs as possible, and arranged to meet those who were successfully reached by phone during one evening of the first weekend of Session II. The researcher was then able to administer the consent forms and the pretest to the rest of his students, and to initiate the extensive reading treatment with the ones who volunteered to receive it.

The researcher originally intended to maintain homogeneity among both the treatment group and the comparison group in congruence with the two separate reading labs, but as a consequence of this minor snafu, the participants in the extensive reading group and the comparison group ended up being mixed together among both of the two advanced reading lab classes.

### **Data gathering procedures**

For Session II, the pretest instrument was one reading comprehension practice test, taken from the *Longman Preparation Course for the TOEFL Test – The Paper Test*

(Phillips, 2003). The posttest instrument was a different practice test, taken from the same source. Under normal circumstances, when students take the reading comprehension section of the paper-based TOEFL test, they are allowed fifty minutes to read five short passages of text, with ten comprehension questions to answer for each passage, for a total of fifty question items. Since the pretest and posttest were to be administered during the scheduled meeting times for the reading labs, and the reading labs were scheduled to last only fifty minutes, the researcher eliminated the last reading passage and the last ten questions from both practice exams, and gave his participants forty minutes to complete the tests. The tests were scored based on a percentage of correct responses.

Pretest data was obtained from eight participants, but one of those eight participants dropped out from the study. Ultimately, the complete set of data used in this study was obtained from seven participants. The treatment group consisted of four participants and the comparison group consisted of three. Due to the small sample size, the results cannot be generalized, and a thorough statistical analysis will not be conducted to determine the statistical significance of either the treatment group's or the comparison group's scores. Instead, the raw scores of each participant's posttest measure will be considered in contrast to the raw scores of their pretest measures.

After the posttest and the end of Session II classes, the researcher conducted personal interviews with these seven participants. The purpose of these interviews was to inquire of the participants' personal feelings about extensive reading and the reading comprehension section of the TOEFL, the test-taking strategies they use for standardized reading comprehension tests, and the teaching methods that they think work best for

them, in order to help them perform better on these tests. For a list of the interview questions, see Appendix A.

## Chapter 3

### The Results of the Study

#### Introduction

Chapter 1 sought to make the case for trying an extensive reading pedagogy with the University of Toledo's ALI reading labs through a review of the literature produced about the positive effects of extensive reading. Chapter 2 introduced the researcher and the participants of this study and the setting in which this study took place, along with a detailed methodology for how the study proceeded and how the data was obtained. This chapter will present the results of the test instruments used and answer the research questions from Chapter 2 based on that data, along with qualitative data obtained through personal interviews with the research participants.

#### Findings

The researcher used two different reading comprehension practice tests from the *Longman Preparation Course for the TOEFL Test – The Paper Test* (Phillips, 2003) as a pretest and posttest measure of his participants' English L2 reading comprehension skill in order to determine whether the participants receiving the extensive reading treatment would demonstrate a greater improvement of their English L2 reading comprehension scores, after the end of the five week long treatment, vis-à-vis the comparison group. The reading comprehension section of the paper-based TOEFL has fifty questions but since the researcher had a limited amount of class time in which to administer his test instrument to his participants, he eliminated the last reading passage and the last ten questions of the test instrument. Thus the total number of question items on each test was forty.

The results of the pretest and posttest measurements are presented in Table 1, which compares the percentage score of the comparison group to that of the extensive reading treatment group. These scores are cumulative totals of correct answers from each group. Since the comparison group consisted of only three participants, while the extensive reading treatment group consisted of four, the total number of test questions for the comparison group is forty less than the total number of test questions for the extensive reading treatment group.

Table 1  
*Cumulative pretest and posttest reading comprehension scores for both groups.*

	Comparison Group			Treatment Group		
	score	percentage		score	percentage	
Pretest	41/120	34%		56/160	35%	
Posttest	39/120	32.5%		60/160	37.5%	
Gain (Loss)	(2)	(1.5%)		4	2.5%	

Table 1 suggests that, taken as a whole, the treatment group did benefit somewhat from the extensive reading treatment that lasted five weeks, with a differential gain of six more correctly answered questions than the comparison group, for a differential gain of 4% between the overall test scores. With only seven participants involved in this study, a statistical analysis would not be practical. However, the small number of participants yields the advantage of being able to observe and consider each individual participant's personal score in isolation from the others. Considering the scores of each participant individually offers more detailed information about the success of the extensive reading treatment for this particular study. Table 2 presents the pretest and posttest scores for

each member of the comparison group individually, while Table 3 presents the pretest and posttest scores for each member of the extensive reading treatment group individually.

Table 2  
*Pretest and posttest reading comprehension scores for comparison group.*

	Saudi Arabian male 2		Saudi Arabian male 3		Japanese female	
	score	percentage	score	percentage	score	percentage
Pretest	14/40	35%	11/40	27.5%	16/40	40%
Posttest	20/40	50%	6/40	15%	13/40	32.5%
Gain (Loss)	6	15%	(5)	(12.5%)	(3)	(7.5%)

Table 3  
*Pretest and posttest reading comprehension scores for extensive reading group.*

	Saudi Arabian male 1		Chinese male		Chinese female		Taiwanese female	
	score	%	score	%	score	%	score	%
Pre-test	15/40	37.5%	11/40	27.5%	16/40	40%	14/40	35%
Post-test	15/40	37.5%	12/40	30%	14/40	35%	19/40	47.5%
Gain (Loss)	0	0%	1	2.5%	(2)	(5%)	5	12.5%

The greatest gain in the test scores was achieved by Saudi Arabian male 2, who was a member of the comparison group and was not expected to do any extra reading outside of the reading lab for this research. Even so, this individual answered six more items



correctly from pretest to posttest for an increase of 15% on his test score. The next greatest gain was achieved by Taiwanese female, who was part of the extensive reading group. She answered an additional five more items correctly from pretest to posttest for an increase of 12.5% on her test score. Otherwise, the posttest performances of the other three members of the extensive reading group are unimpressive compared to their pretest measures. Chinese male was the only other participant in the extensive reading group to have improved his test score, doing so by only one more correctly answered question. Saudi Arabian male 1 made no gain or loss on his test score, while Chinese female actually answered two fewer reading comprehension questions correctly on the posttest, than she did on the pretest, for a loss of 5% to her test score. However, the other two members of the comparison group, Saudi Arabian male 3 and Japanese female, also suffered a posttest loss of their reading comprehension scores, the former member considerably so, answering five fewer questions correctly for a loss of 12.5% to his score, and the latter member answering three fewer questions correctly for a loss of 7.5% to her score.

**Research Question 1: Will pretest and posttest measures of the English L2 reading comprehension of the participants of this study indicate a differential gain for the participants in the treatment group who read more than the comparison group, whom read less?**

When the pretest and posttest scores of the participants in the treatment group and those of the participants in the comparison group are each considered in aggregation, the answer to this first research question appears to be ‘yes.’ The pretest and posttest measures do indicate a differential gain in favor of the treatment group. However, when

the scores of the participants are considered on an individual basis, the answer that the participants in the treatment group benefitted from the extensive reading treatment becomes doubtful. The improvement that Taiwanese female achieved for her posttest score, over her pretest score, seems to indicate an improvement in her reading comprehension abilities. However, Chinese male achieved only a slight gain, from pretest to posttest, answering just one more test question correctly, and Saudi Arabian male 1 achieved no gain at all, while Chinese female suffered a two-point loss from pretest to posttest.

**Research Question 2: Will the participants respond favorably to the extensive reading treatment?**

Taiwanese female responded favorably to the extensive reading treatment while the other three participants in the treatment group did not. These responses can be described in greater detail through two central themes that emerged from the qualitative data from these four participants: belief in the benefit of extensive reading, and engagement with the extensive reading treatment.

***Belief in the benefit of extensive reading.***

Each of the four members of the extensive reading treatment group expressed their own belief that extensive reading improved their reading comprehension ability. Specifically, they all mentioned that they had learned new English vocabulary words from the extensive reading experience. Chinese female said that she believed it helped to improve her knowledge of English grammar.

***Engagement with the extensive reading treatment.***

Based on the qualitative data gathered, the researcher reports that three of the members of the treatment group did not fully engage with the texts that the researcher made available to them. Those three members are Saudi Arabian male 1, Chinese male and Chinese female. Taiwanese female was engaged with the text she selected and seemed to express a positive attitude for reading, though she admitted to the researcher that she was not able to read much of the book she chose, probably because it was too difficult for her.

Saudi Arabian male 1 was a bit more successful than Chinese male and Chinese female at engaging with the texts that he chose from the collection that the researcher made available to him, but was largely preoccupied with the compulsory reading assignments from his regular classes at the ALI. Saudi Arabian male 1 chose two texts to read during the course of the extensive reading treatment. They were Reginald Rose's *Twelve Angry Men* (1955), and the graphic novel *Pride of Baghdad* (Vaughan & Henrichon, 2006). During class meetings, when the participants were to give a brief oral report of their reading effort, Saudi Arabian male 1 was insouciant and forthcoming about his lack of engagement with *Twelve Angry Men*, so the researcher encouraged him to choose something else to read, and he chose *Pride of Baghdad*. During the follow-up interview, the researcher asked Saudi Arabian male 1 to relate his opinion of *Pride of Baghdad*. He gave a general synopsis of the story, which made the researcher feel confident that, indeed, Saudi Arabian male 1 had read this graphic novel.

Based on the interview data, the researcher doubts that Chinese male managed to read much of the texts he selected. This participant chose two comic book adaptations of J.R.R. Tolkien's *The Hobbit* (Dixon, Deming, & Wenzel, 2012) and Jules Verne's

*20,000 Leagues Under the Sea* (Rafter & Ahluwalia, 2010) to read. Each one was selected on a different occasion, throughout the course of the extensive reading pedagogy. Chinese male claimed to have read the comic book version of *The Hobbit* in its entirety. However, his answers to the first six questions of the interview were circumlocutory at best, and completely incoherent otherwise. The way Chinese male answered the interview questions, especially when asked to cite one new word that he learned, seemed like prevarication to the researcher, leading to the suspicion that Chinese male did not successfully engage with the texts provided to him for reading extensively. On the other hand, perhaps Chinese male did learn some new words, but was unable to access his memory of them in his discomfiture, and preferred to evade the question for the sake of his dignity rather than to admit frankly that he could not remember them. Even if that is the case, it is still unclear from the interview whether Chinese male responded favorably to the extensive reading pedagogy, or whether he was simply eager to comply with what he perceived to be the researcher's expectation that he should.

Chinese female did not respond well to the extensive reading pedagogy. Chinese female chose three different texts to read from throughout the course of this study: the October 2012 issue of *National Geographic*; a special issue of *Life* magazine titled *Wonders of the Deep* (2012), which was about the oceanic ecosystem; and a short story from an anthology called *Chicago Lightning*, about of fictional private investigator named Nathan Heller (Collins, 2011). It must be noted that, during one particular class meeting, when Chinese female was asked to talk about what she had read over the course of the past week, she held up the National Geographic magazine and described the pictures therein, pointing to them and looking them over herself, as she did so. She

seemed to be somewhat discomfited as she did this. Then she displayed what the researcher recognized was actually a full-page advertisement for a pharmaceutical product, though she proceeded to comment on it as if it were a featured article. Based on this observation, the researcher suspects that Chinese female neglected to engage with the *National Geographic* magazine. However, during the interview the researcher asked Chinese female to describe and offer her opinion of *Wonders of the Deep* (2012) and *Chicago Lightning* (Collins, 2011). In contrast to the referenced classroom event, she offered enough detail about both texts that it was convincing to the researcher that she had actually read them.

Yet, in spite of satisfying the researcher's interest in knowing whether she actually read any part of the texts she was given, it still cannot be said that Chinese female responded particularly well to the extensive reading pedagogy. Although she said that she enjoyed reading the magazine about the oceanic ecosystem, she said that she did not enjoy the short stories about the private investigator. Then the researcher was somewhat perplexed when she subsequently told him that she read about thirty to forty pages of the short stories. The researcher asked her why she bothered to read so much if she did not enjoy reading it, to which she replied, "Because, I need to answer your questions!" She was referring to the oral book report that every member of the extensive reading group was expected to make once a week. Even though the researcher tried to design an extensive reading pedagogy oriented around student initiative, it seems that Chinese female still regarded it as somewhat of a hassle. Therefore, it must be said that Chinese female did not respond well to the researcher's extensive reading pedagogy.

As for Taiwanese female, based on the researcher's classroom observations of her while she talked about the reading she did over the course of the previous week, it is his professional opinion that she enjoyed reading a short story by Ray Bradbury, called *Dark They Were, and Golden-Eyed* (1980). The researcher recalls that Taiwanese female always spoke in detail about the unique plot of this science fiction story, and that she did so with confident enthusiasm. Therefore, the researcher doubts that she could have been extemporizing those details, and that she actually did read and enjoy the story. Furthermore, hard evidence of her engagement with the text exists in the form of handwritten translations into her L1, that the researcher later found throughout the text of the story she read. In spite of her earnest engagement with the text, Taiwanese female admitted to the researcher that she was only able to read about twenty pages from this story. While that may have been a lot to read from an authentic English text for her own level of English L2 reading proficiency, it should not count as extensive reading. Regardless, Taiwanese female was engaged with the book that she selected and very enthusiastic about the researcher's extensive reading pedagogy.

**Research Question 3: What test-taking strategies do the participants use when taking reading comprehension tests like the TOEFL?**

All of the participants from whom data was successfully collected answered that they used multiple-choice test-taking strategies such as process of elimination, reading the question first and then looking for the answer in the passage, and blind guessing. There is more variation to the participants' responses about the use of speed-reading strategies that involve skipping parts of the text for general comprehension.

Table 4 organizes the interviewees' responses to these questions, for the convenience of the reader to see where the affirmative responses were given. Unfortunately, the data is incomplete for a few reasons. One reason is due to instances where the interviewees' responses were indirect and obscure. Such responses may have been due to persisting problems with English language proficiency on the part of the interviewee, such as listening comprehension and speech. The researcher made efforts to elicit more clarification in such cases without reaching the point of harassment. Additionally, the haste with which the interview with Chinese male was conducted was due to time and scheduling constraints suffered by both the researcher, and the participant, and explains the missing responses from him. Finally, some data from Chinese female is missing because she did not give the researcher consent to make an audio recording of the interview.

Table 4  
*Responses to interview questions 9–19.*

Interview Question	Saudi Arabian male 2	Saudi Arabian male 3	Japanese female	Saudi Arabian male 1	Chinese male	Chinese female	Taiwanese female
9	yes	yes	yes	no data	no data	yes	yes
10	yes	yes	yes	yes	no data	no data	yes
11	yes	yes	yes	yes	no data	no data	yes
12	yes	yes	yes	yes	no data	no data	yes
13	yes	yes	yes	yes	yes	yes	yes
14	no data	yes	yes	no	no data	yes	yes
15	yes	yes	no	yes	no data	no	yes
16	no	no	no	yes	yes	yes	no
17	no	no	yes	no	yes	yes	yes
18	yes	yes	yes	no	no data	no data	yes
19	no	yes	yes	no	no data	yes	yes

Several things are salient from the data presented in Table 3. First, the most significant feature of the data is that every participant answered ‘yes’ to question 13. Question 13 inquired of the participants’ usage of *process of elimination* – the strategy that involves eliminating given possible answers that are known to be wrong, ultimately to isolate the right answer in spite of having not known it – as a multiple-choice test-taking strategy. The next most significant feature of this data is that every participant for whom data was successfully collected answered ‘yes’ to questions 9, 10, and 11, as well. Questions 9, 10, 11, and 13 were all concerned with test-taking strategies unique to the format of multiple-choice tests like the TOEFL. Questions 9 and 10 asked whether the participants used the strategy of reading the enumerated test questions first, prior to actually reading the passage for which reading comprehension was being tested, in order to direct the reading effort specifically towards finding the answers to those questions. Question 11 asked the participants if they ever used blind guessing as a strategy when taking the TOEFL.

The pattern of unanimously affirmative responses, observed for questions 9, 10, 11, and 13, did not continue with the questions that inquired whether the participants used reading strategies (questions 12, 15, 16, 17, 18, and 19), such as skimming for speed and general comprehension (while taking the reading comprehension portion of the TOEFL exam). Some participants answered ‘no,’ while others answered ‘yes.’ The only exception to this was Question 12, “skipping over unfamiliar words or guessing their meaning based on the context,” where every participant for whom data is available gave an affirmative answer. Question 15 asked whether students used *scanning* or *skimming* as reading comprehension strategies. Scanning or skimming were both defined as alternately



reading a few words in the text, and then skipping over a few words as opposed to reading each and every word. Question 16 asked if the participants tried locating what looked to be the most important information in the text by taking a glance at the text and doing such things as looking for the agent in the syntax of the sentences, or any proper nouns. Question 17 asked whether the participants skipped ahead to the next paragraph once they felt they understood the main idea of the concurrent paragraph that they were reading. Question 18 asked whether the participants tried reading only the first and last sentences of a paragraph, along with one sentence in the middle of that paragraph, for speed and comprehension. Question 19 asked the participants if they tried to build overall comprehension of a text by connecting the information from one paragraph to the information in the next paragraph. The participants who answered 'no' to questions 15, 16, 17 and 18 generally explained that they might use such strategies for classroom assignments, but not for the TOEFL because a more thorough comprehension of the text was required for the TOEFL (for the specific details of each question, please refer to Appendix A).

The responses that Saudi Arabian male 2 provided for these interview questions must be reported in greater detail for their potential to shed more light on the anomaly of his superior performance on the post-test measure, and on the nature of multiple-choice test instruments. His responses offer remarkable revelations about his attitude toward the TOEFL exam and the act of reading in general. This excerpt from a transcript of the interview shows how he responded to question 15:

Researcher: Another reading strategy is something that we call scanning, or skimming,

Saudi Arabian male 2: Yeah.

Researcher: That's where you don't actually read the whole text word-for-word, you might just read a chunk of it here, skip, read a chunk here, and then skip; do you sometimes –

SAM2: Yeah, like the topic sentences, the title,

Researcher: Right, like maybe the topic sentence and then the last sentence, yeah,

SAM2: Some jargon words and some, a lot of things – yeah, I learned that in the reading and writing class.

Researcher: OK. Do you think that it's helpful to use that strategy on the TOEFL?

SAM2: Well, sometimes. If I didn't have time, I'll do it.

Researcher: OK. Alright. Uh, let's see,

SAM2: Because in the first TOEFL test I took with you, I was actually reading the paragraph and I was enjoying reading it, because it was a good paragraph, like information that I could use in my life. Yeah, it was really helpful to read this paragraph.

Researcher: OK, so do you think that when you're reading something that is important to you or relevant to your experience, do you think that helps you comprehend the text better because maybe you can relate to it?

SAM2: Yeah, actually. It does a lot.

As can be seen in this excerpt, the researcher was ready to move on from question 15, but then Saudi Arabian male 2 interrupted the researcher to add that he actually enjoyed reading one of passages from the practice test that the researcher used for the pre-test measure. This surprising fact must be considered along with the answer that Saudi Arabian male 2 gave in response to question 17:

Researcher: Another strategy that you might use is if you are reading a paragraph, and it's kind of long. Let's say you're reading word-for-word, maybe by the time you get halfway through the paragraph you think, 'OK. I understand what this paragraph is about,' and then you just skip to the next paragraph. Do you ever use that strategy?

Saudi Arabian male 2: I only use it with sometimes boring paragraphs, because I couldn't complete it, it's getting me sleepy.

The answers that Saudi Arabian male 2 gave for questions 15 and 17 show that the lack of personal interest in what he is reading – not the effort to read and comprehend more efficiently – governs his use of so-called “reading comprehension” strategies. To be able to exercise the kind of discretion that gives more weight to pleasure rather than exigency – in the context of a timed test, where time and efficiency are of the essence – seems to suggest a higher level of proficiency in English reading comprehension. Indeed the qualitative data elicited from Saudi Arabian male 2 seems to help explain his superior post-test score.

**Research Question 4: What are the participants' opinions of the reading comprehension section of the TOEFL?**

All seven of the research participants expressed some kind of frustration with the TOEFL. For three members of the treatment group – Saudi Arabian male 1, Chinese male, and Chinese female – and one member of the comparison group – Saudi Arabian male 3 – that frustration was owed mainly to the difficulty that those participants had with the vocabulary level and time limit challenges that the TOEFL presented to them. However, the frustrations as expressed by one member of the treatment group – Taiwanese female – and two members of the comparison group – Saudi Arabian male 2

and Japanese female – seemed actually to call into question, generally, the validity and reliability of the TOEFL.

Four of the researcher's interview questions were meant to generate the data that was used to answer research question 4. Those four questions are: #20 – “Do you think that section 3 of the TOEFL accurately reflects your ability to read and comprehend English text;” #21 – “On a scale of 1 to 5, how difficult was section 3 of the TOEFL exam, for you;” #22 – “What was the most difficult aspect of section 3 of the TOEFL exam for you;” and #23 – “Would you prefer writing a response (short answer or essay) to a reading passage or article as a test of your English reading comprehension ability instead of the multiple-choice test,” (see Appendix A).

For Saudi Arabian male 2, and Japanese female, their doubt in the general validity and reliability of the TOEFL seems to have manifested in each of them from their respective lack of personal interests in the content of the reading passages from their previous test experiences. For Taiwanese female, her anecdotal skepticism is related to her observations of fellow ESL students whom, apparently to her, could not communicate with native English speakers in spite of having achieved a superior score on the TOEFL. She speculated that this phenomenon probably resulted from those students using some test-taking strategies begotten by the format of multiple-choice tests, such as the ones considered for research question 3.

***Three of the participants generally questioned the validity and reliability of the TOEFL.***

Question 20 of the interview asked “Do you think that section 3 of the TOEFL accurately reflects your ability to read and comprehend English text? Why or why not?”

When the researcher asked Saudi Arabian male 2 whether he thought the reading comprehension section of the TOEFL accurately reflects English reading comprehension ability, he replied, “I think it doesn’t. I think what reflect my comprehension reading, I think it’s from reading books because in the TOEFL it’s like I’m just reading to get over the test.”

Japanese female did not think that the reading comprehension section of the TOEFL was a fair and accurate measure of her English reading comprehension ability because of what she explained was an element of “luck” involved with the reading passages that she encountered in the test. Here is a partial transcript from the researcher’s interview with her:

Researcher: Do you think that whatever your score is on the reading comprehension, do you think, maybe, your ability to read English is actually better than what your TOEFL score shows, or maybe do you think the TOEFL score is a little higher than you think you deserve, or something? What I’m asking is, if I look at your TOEFL score and – whatever it is – I decide that by looking at your score I have an idea of how well you can read English. Do you think that that judgment is accurate?

Japanese female: [pause] Mmmmm [pause] No?

Researcher: Do you understand the question?

JF: Yeah, yeah, yeah, yeah – maybe no.

Researcher: Maybe not?

JF: Yeah.

Researcher: Why? Why do you think that?

JF: The test is like just one time and it’s not like judging whole ability of the person, so it depends on conditions, or what the topic is, and [pause] so

like – yeah, that’s why I say no, like, yeah it depends on, I think, kind of luck.

Researcher: Luck?

JF: Yeah, TOEFL is like kind of luck. You know, if I know the topic and I have basic ideas I understand everything. I don’t need to read whole articles. If I don’t know everything that’s, yeah, luck.

Researcher: Luck?

JF: Yeah.

Researcher: OK.

JF: [laughs]

Researcher: So, “luck” in the sense that if it’s something that you’re interested in, you’re going to do better...

JF: Yeah...

Researcher:...and then so you’re lucky that it was something that you were interested in reading. But if it’s something that’s very boring, you’re not going to like to read it, and so then you’re unlucky. Is that what you mean?

JF: Mm-hmm.

Researcher: OK.

JF: [laughs]

Taiwanese female expressed doubt that an ESL student’s score on section 3 of the TOEFL would accurately reflect that student’s real English L2 reading comprehension ability. When the researcher asked Taiwanese female why she held this opinion, she related her anecdotal observations of her own test performances and how she felt that she could achieve a higher score simply by using a “skill” while taking the test. She said the

same of her fellow English L2 classmates who scored relatively well, but whom she alleged could not communicate well with native English speakers. An excerpt of the transcript is provided here:

Researcher: Do you think that the reading comprehension portion of the TOEFL, do you think that it is a fair measure of your ability to read and comprehend English?

Taiwanese female: I don't think so.

R: No? You think that when the TOEFL gives you a score that that reflects your English ability pretty well?

TF: I don't think so.

R: Oh, you *don't* think so!

TF: I don't think so.

R: So you don't think that the TOEFL score...

TF: ...equal the English ability?

R: Yeah.

TF: I don't think so.

R: Oh, OK. Why not?

TF: Because I know a lot of people they get very well the score in the TOEFL test, but they cannot communicate with American people.

R: Really?

TF: Yeah.

R: That's interesting.

TF: Yeah. Because, you know, sometimes you taking the TOEFL test, you can use some skill to get a high score in the TOEFL test. Maybe like reading, I read the question first and I find the answer. Actually, if the answer I got right – maybe the ten question I got right six question – actually, I cannot understand the whole article. Actually, I cannot understand. Because when I do a practice in the TOEFL test before I use skill, I read the whole sentence and I cannot understand. So when the before, I don't know this skill, the fifty question, I got right maybe nineteen or twenty. But when I know this skill, I can get right about twenty-nine to thirty-two. But I still cannot understand the article. I don't understand. So I don't think if you have a high TOEFL score, your English ability is equal. I have to say, yes, maybe they have the good English ability, but I don't think it's equal.

Researcher: I see. OK. So maybe they just knew how to...

TF: ...answer,

R: ...answer the question like, maybe by process of elimination, or something like that.

TF: Yeah, but I have to say their English ability, maybe its high enough, but they use some skill to get higher.

This answer provided by Taiwanese female expresses her suspicion that a performance of a sort other than authentic English L2 reading and reading comprehension does sometimes occur among some members of the English L2 student population, while they take section 3 of the TOEFL test.

*Some of the participants would have preferred writing a response paper as a possible alternative to section 3 of the TOEFL, for testing English L2 reading comprehension.*

Question 23 of the interview asked whether the participants would prefer to compose a written response to some kind of reading assignment, such as a news article, rather than to take a multiple-choice test like section 3 of the TOEFL, as a way to test



their reading comprehension ability. The four participants who favored such a method were Saudi Arabian male 2, Saudi Arabian male 3, Chinese female and Taiwanese female. The two participants who favored taking section 3 of the TOEFL as opposed to composing a written response to an assigned reading were Saudi Arabian male 1 and Japanese female. No data is available for the participant identified as Chinese male.

Those who indicated their preference for composing a written response to an assigned reading as a method to test their English L2 reading comprehension generally explained their choice as something that would not only be easier than taking section 3 of the TOEFL, but would also allow them to comprehend the text more thoroughly, thanks to the advantage of having no time limit.

Saudi Arabian male 1 preferred the multiple-choice format of section 3 of the TOEFL for the simplicity and efficiency of such a testing method. Japanese female preferred section 3 of the TOEFL because she felt that the time limit was important for testing reading speed as well as overall English L2 reading comprehension.

**Research Question 5: What are the participants' opinions of the timed reading exercises that take place at the beginning of the ALI's reading labs?**

Beginning with the Summer 2012 session, the ALI had implemented a new policy requiring instructors of the reading labs to administer timed reading exercises. The aim of these exercises was to train the students to read rapidly in order to improve TOEFL scores. This policy was implemented after the researcher had already submitted his research proposal; however, the author of this thesis decided to take into account this novel approach to improving the English L2 reading comprehension scores of the ALI's students and to initiate the inquiry into its effect by appending two more questions to his

list of interview questions. Question #26 asked the research participants for their opinions of these timed reading exercises and question #27 asked the participants to rate how helpful they felt these timed reading exercises were for preparing them to perform well on section 3 of the TOEFL, based on a five-point Likert scale (1=not at all helpful; 3=somewhat helpful; 5=very helpful) (see Appendix A).

The participants expressed unanimous and enthusiastic approval for the ALI's timed reading exercises for the sake of preparing them for the English L2 reading comprehension section of the TOEFL. Saudi Arabian male 1 said that it was “the most helpful thing” for preparing him for section 3 of the TOEFL. In accordance with the five-point Likert scale from question 27, the ratings that each participant ascribed to the degree of helpfulness of the timed reading exercises are provided in Table 5, below:

Table 5  
*Rated helpfulness of timed reading exercises based on a five-point Likert scale.*

Participant	Helpfulness Rating
Saudi Arabian male 1	4
Chinese male	No data
Chinese female	5
Taiwanese female	4
Saudi Arabian male 2	4
Saudi Arabian male 3	5
Japanese female	5

No data is available from Chinese male because the interview had to be concluded prematurely due to time and schedule constraints.

### **The case of Saudi Arabian male 2**

Before this thesis considers the implications of these findings in the discussion section, some crucially important data obtained from the interview with Saudi Arabian male 2 must also be presented for its potential to help explain his unexpectedly superior performance over every member of the treatment group. When asked about how many hours a week he spent reading English (interview question 24), Saudi Arabian male 2 revealed that he had been personally engaged in what this researcher considers to be an extensive reading regimen of his own initiative:

Researcher: How many hours a week would you say you spend reading books or reading texts in English?

Saudi Arabian male 2: Actually I read some of the American news on Voice of America.

R: So how many hours a week do you think you spend reading English?

SAM2: Not a lot, but does the context on some like on IMDb [the Internet Movie Database; it is a website that maintains an archive of the plot synopses and critical reviews of various movies and television series as well as some fictional character bios. ([www.imbd.com](http://www.imbd.com))] count? Like if I read that, does it mean something to read it?

R: Yeah!

SAM2: I read a lot about actors, I read about the movies,

R: stuff on the Internet,

SAM2: yeah, comments, how people think about that movie.

R: So do you do this every night? Like, after school, you get on the Internet and read this stuff?

SAM2: Yeah. Actually, now I'm reading – do you know the TV show Walking Dead?

R: Yeah.

SAM2: Yeah, I read about it, now.

R: Oh, OK, So like the plot, the plot synopsis, you might read that on the internet?

SAM2: Yeah, actually, after I read Life of Pi, I went to the cinemas saw it.

Researcher: So how much time per night would you spend reading about the Walking Dead on the Internet, or something like that, or going to IMDb and reading a plot synopsis?

SAM2: I could say four hours, but it's not,...

R: ...solid, not straight, not four hours straight.

SAM2: Yeah, separate.

Given that the researcher could not plausibly, not to mention ethically, prevent his participants from engaging in any and all English L2 reading beyond the context of his research, this incident should not have been so unexpected. Saudi Arabian male 2 estimates that he spends about four hours a night reading about the plot synopses of television programs and movies, as well as the user comments that are posted on an internet website called IMDb ([www.imdb.com](http://www.imdb.com)). This is a significant amount of English L2 reading by itself, but compared to the answers that the other research participants gave to this interview question (Saudi Arabian male 3, six hours a week; Japanese female, five hours a week; Saudi Arabian male 1, one hour each night; Chinese male, no data; Chinese female, 7 or 8 hours a week; Taiwanese female, one to three hours a week), it perhaps goes a long way to explain his superior performance on the posttest measure.

## **Discussion**

### **Implications of Research**

What could the data obtained from this study suggest about the effect of an extensive reading pedagogy for English L2 students? Based on previous studies as referenced throughout Chapter 1, the researcher expected, at best, some improvement of posttest scores over pretest scores on the test instrument used to take the measurement, for the group of participants receiving the extensive reading treatment. As revealed through the collection of qualitative data through personal interviews, the participant who apparently did the most English L2 reading was Saudi Arabian male 2, a participant from the comparison group. Saudi Arabian male 2 read authentic English L1 text on the Internet for approximately four hours each night, on his own initiative. This participant also achieved the greatest posttest over pretest gain and the highest posttest score overall. Thus, this individual represents a single case study example of a correlation between reading extensively and improved performance on a standardized test of English L2 reading comprehension.

As for the treatment group, Taiwanese female answered five more test questions correctly on her posttest score over her pretest score, for a gain of 12.5%. However, the qualitative data obtained from the interview reveals that this participant freely admits that she did not manage to log many hours of English L2 reading per week. She also revealed that she had only read twenty pages from the book that she selected. Therefore, it does not follow that her improvement could have been due to reading extensively.

The remaining members of the treatment group – Saudi Arabian male 1, Chinese male, and Chinese female – each achieved no gain, a slight gain, or performed slightly worse on the posttest over the pretest measure, respectively. Although the two remaining participants from the comparison group – Saudi Arabian male 3 and Japanese female –

both sustained a loss for their posttest scores, it cannot be reasoned that Saudi Arabian male 1 and Chinese male were spared similar losses because of their participation in extensive reading for two reasons. One reason is that Saudi Arabian male 1 and Chinese male simply did not read very extensively. The other reason is that Chinese female also suffered a loss to her reading comprehension score.

Taking as a whole every participant who was supposed to have read extensively, there is nothing compelling enough about the results comparing the pretest and posttest measures of both the extensive reading group and the comparison group to suggest that extensive reading may significantly improve L2 reading comprehension, unless one considers the only participant to have actually read extensively to be Saudi Arabian male 2, ironically a member of the comparison group.

In consideration for the affective benefit of extensive reading, it seems that the only member of the intended treatment group to have expressed a very favorable opinion of extensive reading was Taiwanese female. As for Saudi Arabian male 2, to say that he found favor with reading extensively is essentially tautological, since he did so entirely on his own initiative and unbeknownst to the researcher. These findings support Day and Bamford's Top Ten Principles for Extensive Reading (2002), which generally hold that extensive reading – when implemented correctly – relies heavily on the initiative of the student and should be free from any coercion and compulsion from the teacher in order to force students to read. What these findings may suggest is that the more intrinsically motivated a given language learner is to read in the target language, the more likely extensive reading is to be successfully implemented. Forcing students to read is unlikely to help. Therefore, as teachers of L2 reading, we must consider each of our students for

the individuals they are, and foster their initiative by encouraging them to pursue their own unique academic interests.

Additionally, the virtual unanimity among the participants in acknowledging the use of test-taking strategies inspired by the format of multiple-choice tests like section 3 of the TOEFL – such as reading the question first and then finding the answer in the passage, using process of elimination, or just making a completely blind guess – especially in light of Taiwanese female’s anecdotal testimony, should remind educators and administrators alike that, while reading for the sake of taking a multiple-choice test like the TOEFL is to achieve a quantified score, reading and writing outside of an academic context is purely for the sake of communication.

### **Limitations of Research**

This research was conducted over a very short period of only five weeks. Realistically, that is not enough time to expect any kind of pedagogy to significantly affect learning, let alone one where an extensive amount of books are supposed to have been read. By contrast, the studies referenced in Chapter 1 that found positive correlations between reading habits and superior TOEFL scores considered those reading habits over longer periods of time. As well, the studies that compared pre-treatment and post-treatment measures were conducted over the course of several months or one or more academic semesters. Given the brevity of the present study, it should come as no surprise that the extensive reading treatment generally did not seem to improve the reading comprehension scores of the participants in the treatment group.

Additionally, none of the participants in the treatment group appear to have truly read extensively. Perhaps the only participant in this study who could have been

considered to have read extensively was Saudi Arabian male 2, who did so on his own initiative, outside of the context of this research, and was supposed to be a member of the comparison group. It is probably inevitable that in any inquiry into the effects of an extensive reading pedagogy on L2 acquisition, some of the participants meant to be exposed to it will falter or drop out of the study entirely. In order to compensate for that likelihood, researchers should be advised to have a larger sample size than was involved in the present study, so that reliable measurements can still be taken even if the treatment does not take on a few of the participants.

Moreover, due to professional obligations and the time and space constraints that characterized the context of this research, the researcher resorted to using a convenience sample, rather than a truly random sample, for his research. This convenience sample consisted of only seven participants, which precluded the conduct of a thorough statistical analysis to determine whether the posttest gains and losses were significant.

### **Suggestions for Future Research**

In this study, the participant who achieved the greatest gain in score from pretest to posttest was Saudi Arabian male 2, who was not part of the extensive reading treatment group. However, during the interview, Saudi Arabian male 2 stated that he spent about four hours a night reading movie and television plot synopses, with user comments, on a popular Internet website. The researcher told Saudi Arabian male 2 during the interview that this counted as extensive reading. Had the researcher not told Saudi Arabian male 2 that it counted, the extent to which Saudi Arabian male 2 had engaged in reading English text on the Internet might not have been revealed. It remains the researcher's opinion that Saudi Arabian male 2's perusal of Internet websites counted



as extensive reading. Is this a valid opinion? The researcher calls for a larger statistical study investigating the possible efficacy of various kinds of digital media in an extensive reading pedagogy.

More work needs to be done to investigate the habits, strategies and attitudes that English L2 students deploy and harbor while taking the TOEFL, as well. The author of this thesis hopes that someone will one day conduct a think aloud protocol, as a method of data gathering with participants as they take the TOEFL test. Such an investigation could even be done with native English speakers to elicit their thoughts and opinions about a test meant to measure proficiency in their own language. Truly, the thoughts and strategies that go through the minds of any and all students from a wide variety of academic and language backgrounds should be investigated, as they take timed multiple-choice tests, not just strictly the TOEFL.

Since the author of this thesis asked his participants what they thought about composing a written response to an assigned reading as an alternative method of testing English L2 reading comprehension, one interesting research endeavor could involve the use of such a test instrument, and then comparing the measure of the raters' assessments of those written responses to the measure of the participants' TOEFL scores to see whether a participant who passed the reading comprehension measure of the TOEFL would also pass the raters' assessments of their reading comprehension through a written response to an assigned reading, or vice versa.

Finally, the ALI's relatively new strategy of improving its students' TOEFL scores through timed reading exercises is worth investigating to see whether it does lead to an improvement of the ALI students' reading comprehension scores on the TOEFL.

Such an inquiry would be to the benefit of the administration of the ALI, if by implementing the timed reading exercises they do indeed hope to improve their students' TOEFL scores. If the timed reading exercises can be shown to have a statistically significant effect in a comparison study, English L2 students everywhere, and the institutions that they are enrolled in, would also benefit from this.

### **Conclusion**

The less compulsory an extensive reading pedagogy is, the more likely it is to be effective. Extensive reading programs are heavily dependent on the intrinsic motivation of students. Forcing students to read will probably only sabotage the intrinsic motivation they might have, otherwise.

The reading comprehension section of the TOEFL was rigorously designed and tested for validity and reliability and it is a convenient way for institutions to quantify ESL students' English L2 reading proficiency. Yet, educators and administrators must acknowledge the clever schemes that multiple-choice test instruments like section 3 of the TOEFL will of course inspire in the students taking such tests, and consider the extent to which such behaviors are in line with authentic reading comprehension, as it would occur in the actual classroom context.

The author of this thesis would like to end with a reflection on the words of Saudi Arabian male 2, the participant who read English text more extensively on his own initiative than any of the participants in the treatment group whom the researcher required to read extensively:

SAM2: I think when people read about something just for learning, I think it's really hard. If you read about something that interests you, just for fun, that's the thing that people need to do. So I think that if the TOEFL test

will be about things that interest people about the paragraphs, maybe that will help people a lot, too.

The TOEFL is not the sole determinant of an ESL student's future academic success.

Some uncontrollable variables may belie a student's true English L2 reading comprehension ability, such as whether the student used multiple-choice test-taking strategies, or whether the student was just not engaged with the test itself at the time that he or she took it. Therefore educators and administrators should remember to look beyond the numbers and remember the quality of the student as a person.

## References

- American Language Institute (2013, April 15). Mission Statement and History. Retrieved from: <http://www.utoledo.edu/cisp/ali/mission.html>
- American Language Institute (2007, October 3). Transitioning Policy for Continuing ALI Students Retrieved from: <http://www.utoledo.edu/cisp/ali/pdfs/TransitionPolicy.pdf>
- Andrau, Marianne. (1981). *Mit dem Sturm kam die Liebe* (With the Storm Came Love). Berlin: Cora Verlag.
- Bradbury, R. (1980). Dark They Were, and Golden-Eyed, in. *The Stories of Ray Bradbury, Volume 2*. 56–73. London: Grafton Books.
- Christie, A., (1950). *A Murder is Announced*. Berkely, N.Y.: HarperCollins.
- Christie, A. (1952). *They do it With Mirrors*. Harper, N.Y.: HarperCollins.
- Collins, M. A. (2011). *Chicago Lightning: The collected Nathan Heller short stories*. Las Vegas, NV: Thomas & Mercer.
- Constantino, R., Lee, S. Y., Cho, K. S., & Krashen, S. (1997). Free voluntary reading as a predictor of TOEFL scores. *Applied Language Learning*, 8(1), 111-118.
- Day, R. & Bamford J. (2002). Top Ten Principles for Teaching Extensive Reading. *Reading in a Foreign Language*, 14(2), Retrieved from <http://nflrc.hawaii.edu/rfl/October2002/day/day.html>
- Grabe, W. & Stoller, F. (1997). *Reading and vocabulary development in a second language: A case study*. In J. Coady & T. Huckin (Eds.), *Second Language*

- Vocabulary Acquisition* (98–122). Cambridge: Cambridge University Press.
- Gradman, H. & Hanania E. (1991). Language learning background factors and ESL proficiency. *Modern Language Journal* 75, 39–51.
- Greenberg, D. (2006). Implementation of an extensive reading program with adult learners. *Adult Basic Education*, 16(2), 81-97.
- Harper, D. Editor's Preface. *The Scientific Study and Teaching of Languages*. By Harold E. Palmer. 1968. London: Oxford University Press. v–vii. Print.
- Hitosugi, C. I., & Day, R. R. (2004). Extensive reading in Japanese. *Reading in a Foreign Language*, 16(1), 20–39. Retrieved from <http://nflrc.hawaii.edu/RFL/April2004/hitosugi/hitosugi.html>
- IMDb. [www.imdb.com](http://www.imdb.com)
- Krashen, S. (1982). *Principles and Practice in Second Language Acquisition*. New York: Pergamon.
- Lacayo, R. (2011). *George Washington: How the Great Uniter Helped Create the United States*. New York: Time Books.
- Lee, S. (2007). Revelations from three consecutive studies on extensive reading. *RELC Journal*, 38(2), 150–170. doi: 10.1177/0033688207079730
- Life. (2012). *Wonders of the Deep*. New York: Life Books.
- Mason, B., & Krashen, S. (1997). Extensive reading in English as a foreign language. *System*, 25, 91-102.
- Maxim, H. H. (2002). A study into the feasibility and effects of reading extended authentic discourse in the beginning German language classroom. *The Modern Language Journal*, 86(1), 20–35. doi: 10.1111/1540-4781.00134
- National Geographic Society. (2012). *National Geographic*, Vol. 222, No. 4.
- Palmer, H. (1968). *The Scientific Study and Teaching of Languages*. D. Harper (Ed.). London: Oxford University Press.

- Phillips, D. (2003). *Longman Preparation Course for the TOEFL Test: The Paper Test*. White Plains, N.Y.: Pearson Education, Inc.
- Pichette, F. (2005). Time Spent on Reading and Reading Comprehension in Second Language Learning. *The Canadian Modern Language Review*, 62(2). 243–262.
- Rose, R. (1997). *Twelve Angry Men*. London: Penguin Books, Ltd.
- Tanaka, H. & Stapleton, P. (2007). Increasing reading input in Japanese high school EFL classrooms: An empirical study exploring the efficacy of extensive reading. *The Reading Matrix* 7(1), 115–26.
- Tolkien, J.R.R., Wenzel, D., Dixon, C., & Deming, S. (2012). *The Hobbit* (illustrated ed.). New York: Del Rey.
- Vaughan, B. K., & Henrichon, N. (2006). *Pride of Baghdad*. New York: DC Comics.
- Verne, J., Rafter, D., & Ahluwalia, B. (2010). *20,000 Leagues Under the Sea* (illustrated ed.). New Delhi: Kalyani Navyug Media Pvt. Ltd.
- Yang, A. (2001). Reading and the Non-academic Learner: A Mystery Solved. *System*, 29. 451–466.

## Appendix A

### Interview Questions

1. The purpose of my research is to study the effects of an extensive reading program on students who are trying to learn to read in English as a second or foreign language. Extensive reading involves reading a single text about the same topic for an extended period of time, usually over the course of a few weeks. This is the teaching method that I intended to apply by having you choose from the books I made available to you. Did you read the books that you picked?
2. What did you think of those books?
3. Did you think you benefitted from extensive reading?
4. If you did benefit, how did you benefit?
5. Did you do extensive reading for any of your other classes?
6. How helpful would you say extensive reading was for preparing you to perform well on the TOEFL, compared to all other methods of English reading instruction taught to you at ALI (1=not at all helpful; 2=barely helpful; 3=somewhat helpful; 4=helpful; 5=very helpful)?
7. Did the piece of chocolate motivate you to keep reading your books?
8. You recently took the TOEFL exam through ALI. The last section of the TOEFL was section 3. This was the section where there were several passages of text, each about five paragraphs long, and you were supposed to read each passage and then answer a series of questions meant to check your comprehension. Some of these questions, for example, would ask you what the underlined word in a specific line of the text meant. Do you remember which part of the TOEFL I am referring to here?

9. Sometimes when people take tests like this, they might feel like, in order to finish the test on time, they should skip reading the passage altogether and go straight into the questions, and then, only looking through the reading passage to find the specific answer to whichever question they happen to be working on. Did you use this strategy of reading the questions and multiple-choice answers first, before actually reading the passage, in order to find the answer without actually reading the entire passage?
10. Indeed the TOEFL seems to openly encourage that specific sort of test-taking strategy. For example, a question may ask you “What does the word acquire in line 8 mean?” Questions like these explicitly reference a specific line in the reading passage with a number; it tells you exactly where to find the answer to the question you are looking for so you wouldn’t necessarily have to read the whole passage. Did you notice these questions? Did you try to answer those questions first before actually reading the passage?
11. Sometimes when students take these kinds of multiple-choice tests, they will engage in blind guessing, or filling in the bubbles quickly, without reading the question or its multiple-choice answers. Did you engage in blind guessing? Why?
12. Did you try to guess the meaning of unfamiliar words based on the context of the surrounding text, or did you simply skip over the words you did not recognize?
13. Another strategy people use when taking multiple-choice tests like the TOEFL is called “process of elimination.” Process of elimination is kind of like guessing in the sense that you do not know the right answer, however, you do know for sure that some of the possible answers provided are definitely wrong. If you can eliminate all of the answers you know are wrong, then the only answer remaining must be the right one. Hence, “process of elimination.” Did you use process of elimination to answer any of the questions in section 3 of the TOEFL exam?



14. Process of elimination often involves using prior knowledge in order to eliminate obviously wrong answers. For example, let us suppose that the reading passage was about some country – we'll pretend it was about France – and you just happened to be a French citizen yourself. Then one of the questions asks “According to the reading passage, what is the capital of France: A) Lyon B) Paris C) Orleans D) Bordeaux?” If you were a French citizen, you would likely be able to answer this question without even reading the passage. This is called using prior knowledge. Were you able to answer any of the questions in section 3 of the TOEFL like this?
15. When some people take tests like this, they might read a few words and then skip a few words, and then read a few more words and skip a few more words, as they choose. This is different from reading each and every word. This is a strategy called “scanning” or “skimming.” Did you use this strategy in order to get through the text quickly, or did you read every single word?
16. Besides choosing to skip certain chunks of the text at different intervals, one method of scanning, or skimming might be to try and visually locate what looks to be the most important information in the text. For example you might try to find basic syntactical phrases such as the [subject-verb-object] phrase as in “Jack drove his car,” or you would look for the proper nouns. Proper nouns are famous people, places or things that have names, such as Atlantic Ocean, President Obama, Pepsi, or Microsoft, and so on. If you used scanning or skimming, did you try and pick out the syntactic phrases or proper nouns as a strategy?
17. I imagine that if you saw a sentence like “President Obama crossed the Atlantic Ocean,” you might take that as the main idea of that paragraph and move on to the next paragraph in order to save some time. Did you skip ahead to the next paragraph once you felt like you understood the main idea of the paragraph you were reading?

18. Another strategy that some people use would be simply to read only the first and last sentence of each paragraph – and maybe one sentence in the middle – and hope that they got the main idea of the whole paragraph. Did you use that as a strategy?
19. One strategy people sometimes use in order to build overall comprehension of what they have read is to hold a summary, or the main idea, of each paragraph in their minds and then connect the main ideas of each paragraph in order to produce an overall understanding of the passage. Did you try to connect the information in each new paragraph to the paragraphs before it?
20. Do you think that section 3 of the TOEFL accurately reflects your ability to read and comprehend English text? Why or why not?
21. On a scale of 1 to 5 (1=very easy; 3=neither easy, nor difficult; 5=very difficult), how difficult was section 3 of the TOEFL exam, for you?
22. What was the most difficult aspect of section 3 of the TOEFL exam for you?
23. Would you prefer writing a response (short answer or essay) to a reading passage or article as a test of your English reading comprehension ability instead of the multiple-choice test?
24. About how many hours per week would you say you spend reading books or articles for your classes?
25. Do you remember the timed reading exercises that we did at the beginning of each class meeting? I am referring to the passage that I had you read through three times for a minute each time, and then you answered the ten questions on the back. Then I had you read a second passage, once, and I timed your reading to see how fast you could read. These exercises were supposed to prepare you for section 3 of the TOEFL. Do you remember this?

26. What are your thoughts and opinions about the timed reading exercises that we did at the beginning of each class?
27. How helpful would you say the timed reading exercises that we did at the beginning of each class were for preparing you to perform well on section 3 of the TOEFL (1=not at all helpful; 3=somewhat helpful; 5=very helpful)?