IMPLEMENTING EXTENSIVE READING IN UNIVERSITY EAP WRITING CLASSES

A DISSERTATION SUBMITTED TO THE GRADUATE DIVISION OF THE UNIVERSITY OF HAWAI‘I AT MĀNOA IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

IN

SECOND LANGUAGE STUDIES

AUGUST 2015

By

Jeongyeon Park

Dissertation Committee:

Richard Day, Chairperson
Thomas Hudson
Graham Crookes
Elizabeth Gilliland
Kamil Deen
for my family
ACKNOWLEDGEMENTS

First and foremost, I would like to express my sincere gratitude to my dissertation committee, Dr. Richard Day, Dr. Thom Hudson, Dr. Graham Crookes, Dr. Elizabeth Gilliland, and Dr. Kamil Deen for their guidance throughout the writing process. Completing my dissertation would not have been possible without their help and support.

I am deeply indebted to my academic advisor, Dr. Richard Day, for believing in me and for helping me complete this journey. I am truly thankful for his encouragement and guidance, which he continued to extend even during difficult times. My sincere appreciation goes to Dr. Thom Hudson and Dr. Elizabeth Gilliland for their thoughtful suggestions and constructive feedback on the details of my dissertation. Also, I am very grateful to Dr. Graham Crookes and Dr. Kamil Deen for their continuous support and encouragement.

My special thanks also go to ELI director Kenton Harsch and associate director Priscilla Faucette for their tremendous support in the data collection. The help I received from the other teachers and the students in the ELI have been hugely important to the completion of this study as well. I am also pleased to extend my thanks to Eunseok Ro, Gordon West, and Wei-Li Hsu for their help in analyzing my data and to Laurie Durand for proofreading my dissertation. I would like to acknowledge the generous funding I received from a Language Learning Dissertation Grant from *Language Learning* and the ‘Oihana Maikai‘i Fund from the Department of Second Language Studies at the University of Hawai‘i at Mānoa.

Last but not least, I am deeply grateful to my parents, the true mentors of my life, for standing by me with endless love and support and for teaching me to do my best throughout this long journey, and to my sister and best friend, for your love, laughter, and encouragement.
ABSTRACT

This study explores the effectiveness of incorporating extensive reading (ER) into an English for Academic Purposes (EAP) writing class at a university. Although previous research has shown positive effects of ER on diverse aspects of target language learning, little attention has been given to its usefulness for second language writing development. In addition, limitations such as long-term curricular commitments or time constraints have hampered the application of ER in higher education contexts. Therefore, the present study was designed in an attempt to investigate whether ER can be meaningfully implemented in an EAP writing class and whether it can bring about improvement in second language (L2) writing, with the hope of providing insight on this under-researched area.

Eighty-four students enrolled in intermediate EAP writing classes participated in the study. For the purpose of the study, six participating classes were divided into two groups: a control group and a treatment group. The treatment group’s (n = 44) classes were designed to help students regularly engage in ER in and outside the classroom across one academic semester, by substituting ER for a certain amount of in-class writing practice and homework assigned to the control group (n = 40). However, both groups otherwise followed the same curriculum, with similar lesson materials and instructional features, including teacher lectures, student activities, and major course assignments. Data were collected through pre- and post-tests administered at the beginning and the end of the semester in order to measure possible impacts of ER on the students’ writing development. An extensive reading survey and interviews were also conducted to examine in depth students’ reactions toward ER and their perceptions of the usefulness of ER.

The analysis of the student essays revealed that frequent exposure to easy reading materials in the form of pleasure reading brought about improvement in students’ writing
performance on the posttest. While both groups’ writing improved, a holistic rating indicated that the treatment group performed significantly better than the control group on the posttest. In support of this finding, an analytic rating of the essays showed specifically that the treatment group made greater gains in content, organization, vocabulary, language use, and mechanics. An objective rating, assessed with a variety of measures, revealed that the two groups showed different kinds of improvement; while the treatment group displayed increased accuracy and fluency, the control group exhibited increased fluency with decreased complexity across the two tests.

The extensive reading survey and the interviews found that the students had favorable reactions to engaging in ER, and that they believed it to be helpful in their English learning. In particular, the students recognized that ER integrated with writing activities played an important role in improving the quality of their target language learning experience as the semester progressed.

These findings offer several important implications. Adding empirical evidence to the current literature on ER, the study contributes to deepening our understanding of the effectiveness of ER in writing development. Moreover, the study points to a potentially important role for ER in an EAP context, thereby suggesting that ER’s applications in higher education could usefully be expanded. From a pedagogic perspective, the results of the study are hopeful in that they show that ER can be integrated into EAP contexts, and that it can help students increase their practice of reading and writing in the L2, which may in the long term play a critical role in building a solid foundation for academic literacies.
TABLE OF CONTENTS

Acknowledgments.......................................................................................................................... iii
Abstract ........................................................................................................................................... iv
List of Tables ................................................................................................................................... ix
List of Figures ................................................................................................................................. x

Chapter 1. Introduction .................................................................................................................. 1
  Overview of Previous Research on Extensive Reading................................................................. 1
  Gaps in the ER Literature ........................................................................................................... 2
    Gaps in ER and writing research. ............................................................................................... 2
    ER and its application to EAP contexts ...................................................................................... 4
  Overview of the Dissertation ....................................................................................................... 6

Chapter 2. Literature Review ......................................................................................................... 7
  Theoretical Perspectives on the Reading and Writing Relationship........................................... 7
  Perspectives on Linking Reading and Writing in L2 ................................................................. 11
  Extensive Reading Approach ..................................................................................................... 16
  The Effectiveness of ER on L2 Vocabulary ............................................................................... 18
  The Effectiveness of ER on L2 Reading .................................................................................... 24
  The Effectiveness of ER on Affective Domain ......................................................................... 28
  The Effectiveness of ER on L2 Writing .................................................................................... 31
  L2 Writing Assessment ............................................................................................................. 34
    Holistic assessment. ................................................................................................................ 34
    Analytic assessment ................................................................................................................ 36
    Portfolio assessment ............................................................................................................... 37
    CAF assessment ..................................................................................................................... 38
    Challenging issues of CAF. ..................................................................................................... 46
  Rationale for Applying Various Assessment Methods ............................................................. 47

Chapter 3. Method ......................................................................................................................... 49
  Research Purposes and Research Questions ............................................................................ 49
  Participants ................................................................................................................................. 50
  Instruments .................................................................................................................................. 53
    Background survey. ................................................................................................................ 53
    Writing test ............................................................................................................................. 54
    Extensive reading survey. ...................................................................................................... 56
    Post interviews ....................................................................................................................... 56
  Data Collection Procedures ...................................................................................................... 57
    ER workshop. .......................................................................................................................... 57
    ER library ............................................................................................................................... 59
    In-class ER. ............................................................................................................................ 60
    Out-of-class ER ...................................................................................................................... 62
    Writing activities .................................................................................................................... 62
    Book recording form .............................................................................................................. 64
    Teacher journal. ..................................................................................................................... 65
  Rating Procedures ..................................................................................................................... 66

vi
<table>
<thead>
<tr>
<th>Appendix</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Background Survey</td>
<td>159</td>
</tr>
<tr>
<td>B</td>
<td>Writing Test</td>
<td>161</td>
</tr>
<tr>
<td>C</td>
<td>Extensive Reading Survey</td>
<td>162</td>
</tr>
<tr>
<td>D</td>
<td>Interview Questions</td>
<td>165</td>
</tr>
<tr>
<td>E</td>
<td>Consent Forms</td>
<td>166</td>
</tr>
<tr>
<td>F</td>
<td>Suggested Topics for the 10-minute Writing</td>
<td>170</td>
</tr>
<tr>
<td>G</td>
<td>Book Recording Form</td>
<td>172</td>
</tr>
<tr>
<td>H</td>
<td>Assignment Log</td>
<td>173</td>
</tr>
<tr>
<td>I</td>
<td>Criteria for the Holistic Rating</td>
<td>174</td>
</tr>
<tr>
<td>J</td>
<td>Criteria for the Analytic Rating</td>
<td>176</td>
</tr>
<tr>
<td>K</td>
<td>Guidelines for Objective Measures</td>
<td>181</td>
</tr>
<tr>
<td>L</td>
<td>The 10-minute Writing Samples</td>
<td>184</td>
</tr>
<tr>
<td>References</td>
<td></td>
<td>186</td>
</tr>
</tbody>
</table>
LIST OF TABLES

1. Demographics of the Participants .............................................................................................................52
2. Design of the Study ........................................................................................................................................58
3. Intercorrelations among Raters ..................................................................................................................74
4. Descriptive Statistics for Holistic Rating ..................................................................................................75
5. Descriptive Statistics for Analytic Rating ................................................................................................79
6. Intercorrelations among the Five Analytic Categories ...........................................................................81
7. Rater Measurement Report for Four Raters ............................................................................................92
8. Difficulty Measurement Report for Five Categories .............................................................................94
9. Descriptive Statistics for Complexity Measures ....................................................................................95
10. Descriptive Statistics for Accuracy Measures ......................................................................................98
11. Descriptive Statistics for Fluency Measures ........................................................................................101
12. Amount of Time Spent and Number of Books Read for ER .................................................................106
14. Summary of Responses to Open-ended Questions on ER and the 10-minute Writing ........110
LIST OF FIGURES

1. Average Writing Scores based on Holistic Rating .................................................................76
2. Boxplots for Holistic Rating .................................................................................................77
3. Average Scores for Content ................................................................................................80
4. Boxplots for Content ...........................................................................................................82
5. Average Scores for Organization .........................................................................................83
6. Boxplots for Organization ...................................................................................................84
7. Average Scores for Vocabulary ..........................................................................................85
8. Boxplots for Vocabulary ......................................................................................................85
9. Average Scores for Language Use .......................................................................................86
10. Boxplots for Language Use ...............................................................................................87
11. Average Scores for Mechanics ........................................................................................87
12. Boxplots for Mechanics ......................................................................................................88
13. Variable Map from the FACETS Analysis .......................................................................91
14. Boxplots of C/T for Complexity .........................................................................................96
15. Boxplots of C/S for Complexity .........................................................................................97
16. Boxplots of EFT/T for Accuracy .........................................................................................99
17. Boxplots of EFC/T for Accuracy .........................................................................................100
18. Boxplots of EFC/S for Accuracy .......................................................................................100
20. Boxplots of W/C for Fluency .............................................................................................103
21. Boxplots of W/S for Fluency .............................................................................................104
22. Boxplots of WPM for Fluency

105
CHAPTER I

INTRODUCTION

Overview of Previous Research on Extensive Reading

Extensive reading (ER) aims to increase language learners’ exposure to a target language by leading them to read large quantities of easy, interesting materials. There is no agreement as to how extensive is extensive enough (Robb & Susser, 1989), but Day and Bamford (1998) suggested that one book per week may be sufficient to benefit language development and build good reading habits. They further proposed ten principles that can help ER programs to be successful by maximizing the benefits of this “student-centered” (Bamford & Day, 1998, p. 133) and pedagogically sound approach. The essence of the principles is that students should read as much as possible, should read for enjoyment, and should have freedom to choose what, when, and where to read (Day & Bamford, 2002). Given that this is the way reading for pleasure is done in the real world, ER seems to be an ecologically valid approach to learning to read.

Day and Bamford (1998) further recommended that materials at learners’ current linguistic level or a slightly easier level are ideal for developing fluency and confidence in second language (L2) reading. The extensive reading approach based on this theoretical background has drawn increasing attention and support from L2 and foreign language researchers and educators over the past two decades.

The potential role of ER in L2 learning has been much investigated. ER’s effectiveness has been documented in both small-scale case studies (e.g., Cho & Krashen, 1994; Judge, 2011; Leung, 2002; Nishino, 2007) and large-scale studies (e.g., Elley, 1991; Elley & Mangubhai, 1983; Robb & Susser, 1989). An early longitudinal study conducted in Fiji by Elley and Mangubhai (1983) demonstrated the diverse impacts of ER on general L2 proficiency particularly well.
Since then, a broad array of research has provided evidence that learners who engage in ER can improve in various aspects of L2 learning: reading comprehension (e.g., Al-Homoud & Schmitt, 2009; Bell, 2001), reading rate (e.g., Beglar & Hunt, 2014; Beglar, Hunt, & Kite, 2012; Iwahori, 2008), vocabulary (e.g., Pigada & Schmitt, 2006; Waring & Takaki, 2003; Webb & Macalister, 2013), spelling (e.g., Pigada & Schmitt, 2006; Polak & Krashen, 1988), and writing (e.g., Tsang, 1996, Hafiz & Tudor, 1990; Tudor & Hafiz, 1989). It is worth noting that one strand of this research has also attempted to show the effectiveness of ER by comparing it to more traditional approaches such as a skill-building instruction or intensive reading instruction (e.g., Al-Homoud & Schmitt, 2009; Taguchi, Takayasu-Maass, & Gorsuch, 2004).

In addition to its diverse impacts on language development, ER has been reported to have a positive influence on learner affect (e.g., Judge, 2011; Rodrigo, Greenberg, & Segal, 2014; Takase, 2007). Empirical evidence has shown that less motivated or reluctant readers became eager and enthusiastic readers after one semester, or even less, of exposure to ER (Cho & Krashen, 1994; Mason & Krashen, 1997). Although ER research has been largely limited to English for second language (ESL) or English for foreign language (EFL) contexts, a few noteworthy studies have attempted to apply ER to other languages and found parallel results, especially in regard to learners’ positive attitudes toward ER (e.g., Japanese: Hitosugi & Day, 2004; Tabata-Sandom, 2013; Spanish: McQuillan, 1996).

Gaps in the ER Literature

Gaps in ER and writing research. As seen from the overview of the literature, ER research has focused on examining its effectiveness for L2 learning, particularly with respect to reading (e.g., comprehension, rate) and vocabulary development. Only a few studies have looked into the relationship between pleasure reading and writing development in an L2. Although the
research is still scanty, and the existing studies have focused on different skills, the findings are illuminating. For example, Hafiz and Tudor (1990) and Tudor and Hafiz (1989) showed a positive influence of ER on writing readiness and accuracy. Tsang (1996) demonstrated that students who engaged in pleasure reading demonstrated increased descriptive writing skills, particularly those related to content (e.g., development of thesis, relevance to the topic) and language use (e.g., grammar). Mason and Krashen’s (1997) research corroborated Tsang’s (1996) findings: input-based instruction through pleasure reading was more effective than output-based writing practice for improving writing ability.

Krashen (2004) asserted that learners who read more read and write better, and that free voluntary reading is the most powerful tool to improve literacy levels. In particular, he suggested that writing competence is acquired not from simply “reading more” but from “self-motivated reading” (Krashen, 1984, p. 21). Furthermore, Krashen emphasized that good writing style comes from the combined endeavor of reading for meaning and writing to convey meaning (p. 37). Elley and Mangubhai’s (1983) so-called book flood study also provided empirical evidence that those who read more did show better writing ability. They argued that exposure to good models of L2 written language through reading is a natural way to learn how languages are used in context. In line with these arguments, Janopoulous (1986) found a strong correlation between pleasure reading and writing proficiency in L2.

Despite these early encouraging voices on the potential role of ER in L2 writing development, the research has apparently been insufficient to bring ER practitioners’ attention to the connection between reading and writing. This paucity of research partly motivated the present study.
ER and its application to EAP contexts. Day and Bamford (1998, pp. 41–42) suggested that ER can be incorporated (1) as a separate course, (2) as part of an existing reading course, (3) as a non-credit addition to an existing reading course, or (4) as an extracurricular activity, according to a program’s degree of commitment to the approach. ER programs to date have been mostly implemented by being integrated into an existing reading course or as an extracurricular activity. Moreover, while ER has been favored in primary and secondary schools, it has been less popular in higher education or in EAP (English for Academic Purposes) contexts, partly due to a practicality issue (Grabe, 2001; Macalister, 2008). Practicality seems to be a reasonable concern given the lower flexibility of predetermined curricula in colleges and universities (Ferris & Hedgcock, 2005). Even strong advocates of ER tend to be apprehensive of incorporating ER into the classroom when the aim of both teachers and students is rapid language development (Macalister, 2008). This time commitment required by ER is one of the main reasons that EAP contexts have little room for it (Grabe, 2001).

Some researchers, however, have called for expanding ER’s application to broader or higher educational settings. For example, Krashen (1995, 2003, 2004) has maintained his assertion that “light reading” (2003, p. 22) can be of great help for learners acquiring academic language competence by acting as a bridge to heavier reading. That is, reading easy materials a great deal can serve as a springboard for learners who need to prepare themselves for more complex and dense texts. Carrell and Carson (1997) further argued for the necessity of integrating both intensive and extensive reading instruction in an EAP reading curriculum. They contended that both play an important role in preparing students to handle the demands of academic reading, assisting their learning to read as well as their learning reading strategies. In support of these claims, Macalister (2008) attempted to show the possibility of providing a place
for ER in EAP settings. In his action research, ER was integrated into a 12-week university preparation EAP reading program in the form of 20 minutes of sustained silent reading a day. The participating students showed favorable reactions to ER, demonstrating, for example, their intention to increase the amount they read and more positive attitudes toward ER as the semester progressed (p. 253).

Building on the emerging consensus on the benefits of exposing students to input-rich environments to aid their transition to academic literacy, Grabe and Zhang (2013) argued that EAP students need to read extensively along with constant reading and writing practice to build confidence and fluency, which is fundamental to academic literacy (p. 20). They pointed out that one of the challenges EAP students face is a lack of experience in integrating reading and writing skills for academic purposes.

It is noteworthy that more and more researchers have recognized the possibility of introducing ER to EAP contexts; however, proposals to apply ER generally confine it to reading curricula. The intuitive logic of applying ER to L2 reading or vocabulary learning may be one reason for the limited scope of most research. Another reason may be that although the importance of reading and writing integration in academic contexts has been advocated by L2 researchers and practitioners (e.g., Belcher & Hirvela, 2001; Carson & Leki, 1993; Grabe, 2001), it is still common for reading and writing skills to be separately taught in EAP contexts. Furthermore, little empirical evidence of the connection between reading and writing exists; exactly how these two literacy skills interact is underresearched, and further investigation that can demystify the complex nature of the link between them is called for (Ferris & Hedgcock, 2005). The present study was therefore designed in an effort to expand the areas of ER’s
application as well as to promote reading and writing integration, with the hope of offering some insight into this less researched area.

Overview of the Dissertation

Chapter 2 reviews the literature relevant to this study and is divided into three sections. The first section discusses theoretical perspectives that suggest directions to explore for linking the literacy skills of reading and writing. The second section reviews studies on ER that illustrate the diverse impacts of ER on different aspects of language learning, providing an understanding of the current state of ER research. The third section describes how this study has arrived at certain writing assessment measures for data analysis by reviewing diverse measures available for evaluating second language writing. Chapter 3 presents the methodology of this study, first explaining the purposes of the research and the research questions, and then providing a detailed description of its participants, contexts, and procedures. In Chapter 4, the results of the quantitative and qualitative analyses of the data are reported. Chapter 5 discusses the findings of the study with regard to the research questions. Chapter 6 provides a brief summary of the major findings, explains the study’s limitations and pedagogic implications, and makes suggestions for future research.
CHAPTER II
LITERATURE REVIEW

This chapter consists of three major sections. In the first section, the reading and writing connection that provides the theoretical background of the present study is described. This section also discusses how the integration of the two literacies might enhance the process of L2 learning. The second section provides a detailed review of research that focuses on the effectiveness of ER in different aspects of L2 learning, including vocabulary development, reading, affect, and writing. The last section examines measures that have been used to analyze L2 writing performance; the research reviewed in this section has offered insights that affected the selection of the measures used for data analysis in the current study.

Theoretical Perspectives on the Reading and Writing Relationship

Traditionally, reading and writing have been considered distinct, and therefore they tended to be taught separately. Accordingly, research on reading and writing mostly progressed independently (Leki, 1993; Reid, 1993a). However, some scholars have claimed that the two skills are interrelated, and have attempted to examine their relationship (e.g., Belcher & Hirvela, 2001; Carrell & Conner, 1991; Carson & Leki, 1993; Esmaeili, 2002; Kroll; 1993; Leki, 1993). Hypotheses about the L1 reading-writing connection laid the foundations for understanding the two literacies in L2 (Hirvela, 2004), but explanations based on L1 learning were argued to be insufficient to explain how the two literacies interact in L2 (Eisterhold, 1990; Grabe, 2001). In addition, little empirical work is available to lend conclusive support to the relationship in L2.

One possible reason for the lack of research may be that the reading-writing relation in the L2 is more complicated and intertwined than in the first language (L1) (Hudson, 2007). Most
L2 learners have already acquired literacy in their native language, and this established L1 literacy ability plays a role by either facilitating or hindering their L2 development. Also, various L2 contexts, including ESL and EFL, can be another factor that makes it difficult to clearly establish how reading and writing integration enhances learning (Grabe, 2001).

One issue that has brought about much discussion in reading and writing research, yet has been more vigorously addressed in regard to L1, is directionality (Grabe, 2003). The issue is whether transfer of reading and writing abilities is (1) directional, meaning that ability acquired in one modality can be transferred to another one; (2) nondirectional, meaning that reading and writing are both driven by a single cognitive proficiency; or (3) bidirectional, meaning that reading and writing enhance each other.

The directional hypothesis posits that the reading and writing connection is directional, so that any transfer of skills is either from reading to writing or from writing to reading (Carson, 1990). Unlike the directional model, which claims that skills only transfer in one direction, the nondirectional model posits that transfer can occur in either direction because it assumes that reading and writing share “a single cognitive proficiency” (Eisterhold, 1990, p. 90). Therefore, improvement in one mode brings about improvement in the other. The basic notion of the bidirectional hypothesis is that reading and writing are interdependent but their relationship is “qualitatively different at different stages of development” (Eisterhold, 1990, p. 92).

Under the directional hypothesis, the read-to-write model emphasizes input in the process of learning to write, considering how learners can reproduce or reflect structural knowledge gained from reading, for example, comparison and contrast, in their writing (Eisterhold, 1990). How various texts can be used as a useful tool for learners to develop their writing skills is of key interest in this framework.
In contrast, the write-to-read model revolves around the idea that writing enables and empowers reading. Simply put, writing helps improve reading comprehension and retention of information (Stotsky, 1983). This model considers writing prior to reading as “an avenue to activating information that is already known to the potential reader and to prompting expectations about the possible content of the upcoming text” (Hudson, 2007, p. 274). Similarly, Hirvela (2004) noted that writing before, during, or after reading helps readers enhance their meaning-making process, which can consequently lead to the development of L2 reading skills. In addition, writing about one’s ideas, questions, or expectations not only encourages a deeper level of engagement but also provides room for interacting with the text while negotiating or interpreting the meaning embedded in the text (Zamel, 1992). Oded and Walters (2001) provided empirical evidence supporting these claims by showing that EFL students who summarized two given texts after reading scored higher on a subsequent reading comprehension test than students who simply listed examples used by the writer. That is, summarizing was found to be helpful in assisting students’ deeper comprehension of the texts, which in turn led to better performance on the test.

Drawing on the unidirectional hypothesis, Fitzgerald and Shanahan (2000) illustrated how reading and writing rely on analogous cognitive process and similar knowledge. They delineated four types of knowledge that reading and writing share (pp. 40–41): (1) metaknowledge (e.g., knowing that readers and writers interact, monitoring one’s meaning-making process); (2) domain knowledge about substance and content (e.g., vocabulary meaning, meaning created in context); (3) knowledge about universal text attributes (e.g., graphophonics, syntax, text format); and (4) procedure knowledge and skill to negotiate reading and writing (e.g., knowing how to access and use knowledge). The unidirectional model thus underlines that
“whatever processes a reader uses to make meaning of a text may also be the same processes that a writer uses to create meaning” (Carson, 1993, p. 87). It has been argued from this perspective that good readers are good writers in general (Grabe, 2003), and that “good writers are often good readers” (Reid, 1993a, p. 43).

Nevertheless, the relationship between reading and writing needs to be cautiously viewed. For example, Tierney and Shanahan (1991) found that reading and writing abilities are only moderately correlated in the range of .20 to .50. This demonstrates shared knowledge of reading and writing, but their correlation can explain no more than 25% of the variance. In other words, their findings imply that reading and writing are also independent processes, being “as separate as they are similar” (Fitzgerald & Shanahan, 2000, p. 43).

Similarly, one study investigated the relationships between reading and writing in both L1 and L2, and found that literacy skills can be transferred across modalities, but to different degrees (Carson, Carrell, Silberstein, Kroll, & Kuehn, 1990). Japanese and Chinese ESL students took an essay test and a cloze test in both their L1 and L2 (English), and their scores were compared using the Pearson correlation. Between reading and writing abilities in L1, a moderate correlation was found in Japanese ($r = .494$, $p < .01$), and a weaker correlation in Chinese ($r = .271$, $p < .05$). For the L1–L2 reading relationship, moderate correlations were found between Japanese and English ($r = .509$, $p < .01$) and between Chinese and English ($r = .366$, $p < .01$). However, for the L1–L2 writing relationship, weaker correlations were found between English and both languages (Japanese and English: $r = .230$, $p < .05$; Chinese and English: $r = -.019$, ns). Based on these results, the authors suggested that (a) the strength of the relationship between reading and writing may be different according to the L1 background, and (b) reading ability may transfer more easily than writing ability does. They also emphasized that other complex
Factors such as educational background and experience influence the reading and writing relationship, implying that the L2 literacy development of adult learners is complex.

**Perspectives on Linking Reading and Writing in L2**

Although several different hypotheses on the relations between reading and writing have been proposed, there seems to be a consensus among L2 researchers that reading and writing mutually support the development of literacy (Grabe, 2001; Hirvela, 2004). Particularly, Carson and Leki’s (1993) book *Reading in the Composition Classroom: Second Language Perspectives* sparked the interest for linking the two literacies in L2 (Grabe, 2001; Hirvela, 2004). As opposed to the more traditional view that reading is a decoding skill and writing is an encoding skill, the book discusses the reciprocal interaction of reading and writing. That is, reading is a good resource for building knowledge useful for writing, and writing reinforces that knowledge, as the skills share active meaning-making and interactive characteristics (Hirvela, 2004; Leki, 1993; White, 1994). In the same vein, Zamel (1992) underlined the complementary roles of reading and writing:

> [Reading] must allow for the ways in which readers contribute to and make connections with the text. Writing provides a unique opportunity for discovering and exploring these contributions and connections, for it allows the reader to dialogue with a text and find a particular way into it. (p. 468)

In order to promote the integration of reading and writing, learners must, Smith (1983) claimed, “read like a writer in order to learn to write like a writer” (p. 562). It is therefore important to help students read texts rhetorically by focusing on choices made by writers (Kroll, 1993). Likewise, students need to write like a reader based on the understanding that the texts they compose are “readings,” and thus further realizing the importance of meeting a reader’s
needs and expectations (Spack, 1985). Zamel (1992) argued that these approaches toward reading and writing can give students insights into how “reading lets us know writing, and writing lets us know reading” (p. 481) as they work in tandem to enhance each other.

Hirvela (2001) further pointed out that how to meaningfully connect the two literacies should be something that teachers explore in depth. The most important things for teachers to understand are the complexity of the relationship between reading and writing and how to apply their knowledge in their pedagogy in such a way that both literacies benefit (Carson, 1993; Eisterhold, 1990), as skewed literacy practices can produce imbalanced L2 improvement (Kim, 2012). For similar reasons, Ferris and Hedgcock (2005) asserted that an eclectic approach should be taken by educators in order to meet the diverse needs of learners and keep a pedagogic balance in the design of lessons and curricula. Despite the appeal of the idea of integrating reading and writing, there is a serious lack of empirical evidence showing whether it can in fact bring about greater improvement in L2 learning than teaching the two literacy skills separately (Carson & Leki, 1993; Grabe, 2001).

Both reading and writing are “equally important processes of a literacy event” (Carson, 1993, p. 85), but reading has received much more attention than writing (Ferris & Hedgcock, 2005). For example, Hirvela (2001, 2004) discussed how the reading-writing connection can be further enriched by applying reader response theory (Rosenblatt, 1978), “an approach to reading that primarily examines and values readers and how readers read texts, not how or why authors wrote them or how they are organized” (Hirvela, 2004, p. 46). In contrast to more traditional viewpoints centering on the author’s purposes or the text itself, this reader-based approach regards the experience of the readers—among the three elements of author, text, and reader that are involved in the reading process—to be key. In other words, from this perspective, it is critical
how readers make meaning while reading in association with their experience and interests by positioning themselves at the center of the reading activity.

A related issue centers on the use of literary texts for L2 learners; some researchers have argued that literary texts are beneficial in linking reading and writing, because these texts are likely to have a wide range of interpretations, which allows their readers to take an active role. Hirvela (2004) claimed that literary texts can be particularly well suited to writing instruction involving reading and writing in L2, as they can help students see the integration of the two literacy skills, which is important to further enrich their academic reading and writing skills. Hirvela went on to say that literature is usually effective for children because it facilitates their literacy development through pleasurable experiences, but it can also be advantageous for the education of students who are new to ESL environments, or more advanced-level students who are still in need of rich target input. Similarly, Hedgcock and Ferris (2009, p. 248) encapsulated the benefits of integrating literary texts into L2 instruction as follows: (1) cultivating cultural knowledge, (2) providing rich input for language acquisition, (3) supporting long-term motivation and building confidence, (4) offering topics to write about, and (5) promoting critical thinking skills.

Nevertheless, there has been a long-standing debate over the use of literature in composition classes (Belcher & Hirvela, 2000). Literary texts once played a significant role in the context of ESL or EFL, as they were considered to be the best examples of a language and therefore suitable for use in early foreign language education (Spack, 1985). However, since the focus of foreign language has moved to EAP, literature has begun to be considered “ill suited” for education when the primary purposes pertain to increasing linguistic proficiency (Hirvela, 2001, p. 111), because literature tends to use difficult language with high levels of structural and
lexical complexity. This trend has almost excluded any possibility of using literary texts in the ESL or EFL classroom, regardless of their potential value.

One of the core arguments against using literary texts lies in their limited application to the academic discourse that learners encounter in college. Furthermore, opponents strongly criticized the use of literature to make connections between reading and writing for college students. In other words, students are less likely to encounter literature than other genres in academia, and therefore it is not suitable for the purposes of academic writing classes. For instance, Horowitz (1990) cast doubt on the transferability of skills acquired through reading literature, claiming that writing about literature does not resemble the writing required in academic contexts.

In contrast, Spack (1985) stressed the value of literature in language learning; for example, she argued that short fiction can not only provide subject matter for writing but also offer resources for analysis such as theme, plot, character, setting, or point of view. Later, Spack (2004) conducted a case study of one Japanese student, Yoko, who attended a four-year university in the United States. This longitudinal study describes how Yoko struggled in her first year at the university due to lack of practice in academic reading and writing and how she developed academic literacies by engaging in reading and writing processes, constructing knowledge, and reflecting on her own learning process (p. 32). One of the interesting points that Spack reported is Yoko’s decision to read a few novels on her own in order to overcome her fear of taking a course that, like many university courses, had a heavy reading load. Yoko reported that she gained confidence by successfully finishing a few novels before she enrolled in the class (pp. 24–25). This intriguing description of Yoko’s case suggests the worth of further research on the potential role of literary texts in acquiring academic literacy. Related studies include one by
Costello (1990), which claimed that literary texts can be useful in literacy development as story telling (e.g., short stories and autobiographical essays) is a common genre in any culture.

In relation to Spack’s (2004) finding, pleasure reading has received some attention in the discussion of how to link reading and writing (Grabe, 2001; Hudson, 2007). Reading for pleasure is often linked to Krashen’s theory of second language acquisition (Carson, 1990; Carson et al., 1990). According to Krashen (1984), as speaking ability is acquired from comprehensible input with a low affective filter, writing ability is also developed via comprehensible input. In particular, voluntary pleasure reading that focuses on meaning is what develops the acquisition of writing. He asserted that writing practice is useful in aiding learners to familiarize themselves with writing processes and writing strategies, such as planning and revision. However, writing practice per se only helps writing performance, but does not help writing competence. Krashen (1984) stated:

Writing competence, it is hypothesized, comes only from large amounts of self-motivated reading for interest and/or pleasure. It is acquired subconsciously; readers are unaware they are acquiring writing competence while they are reading, and are unaware of this accomplishment after acquisition has taken place. It is reading that gives the writer the “feel” for the look and texture of prose. (p. 20)

Krashen admitted that reading alone is not enough for the acquisition of writing competence, although it is a necessary precedent; actual writing practice as well as writing instruction are needed to develop writing performance.

One study followed up by exploring Krashen’s (1984) hypothesis that pleasure reading leads to gains in writing proficiency (Flahive & Bailey, 1993). In this study, learners’ L1 and L2 pleasure reading time was compared to their reading comprehension test scores, argumentative essay scores, grammar test scores, and T-unit based analysis. The results showed that self-reported reading amounts were related to reading comprehension scores ($r = .49$, $p < .05$), and
that reading comprehension scores were related to holistic essay scores with a weak correlation
\((r = .35, p < .05)\). However, pleasure reading showed no relationship with grammatical accuracy
or complexity based on T-unit analysis. Krashen’s claim is appealing and has promoted the
notion of a reading and writing connection, but it still lacks an empirical foundation (Eisterhold,
1990; Grabe, 2001). The relationship between the two literacies will be discussed in detail in the
next section, along with the effectiveness of pleasure reading in other areas of language learning.

**Extensive Reading Approach**

In addition to the research on the role of reading in the development of writing, many
studies over the past two decades have researched the effectiveness of pleasure reading or ER in
other areas of language development. ER, which the present study draws on as a pedagogic
approach, is defined as “an approach to the teaching and learning of second language reading in
which learners read large quantities of books and other materials that are well within their
linguistic competence” (Day & Bamford, 1998, p. xiii). The formulation of ten principles of ER
by Day and Bamford (1998, 2002) further sparked L2 researchers’ and educators’ interest in its
impact on L2 learners’ language development. The principles describe characteristics of ER
along with practical guidelines that can help teachers understand what ER means and how to
implement ER more successfully in their specific setting. The ten principles are as follows (Day
& Bamford, 2002):

1. The reading material is easy.
2. A variety of material on a wide range of topics must be available.
3. Learners choose what they want to read.
4. Learners read as much as possible.
5. The purpose of reading is usually related to pleasure, information and general
   understanding.
6. Reading is its own reward.
7. Reading speed is usually faster rather than slower.
8. Reading is individual and silent.
9. Teachers orient and guide students.
10. The teacher is a role model of a reader.

The notion of reading easy materials a great deal for L2 development can be traced back to Elley and Mangubhai’s study in 1983. The study was conducted almost 30 years ago, yet it is worth looking at in detail here, as it is usually considered the first large-scale empirical L2 study on ER, and it provides a good description of the impact of reading. Elley and Mangubhai carried out this study in L2 primary schools in Fiji for about 20 months with the aim of making L2 acquisition similar to L1 acquisition (p. 55), hypothesizing that increased L2 exposure through interesting storybooks would be beneficial and lead to rapid development of L2. A total of 380 students in Classes 4 and 5 (9 – 11 years old) from eight rural schools were assigned to the following three groups: Shared Book group, Silent Reading group, and the control group. Approximately 250 books were provided to the first two groups. For the Shared Book group, the teacher chose a high-interest book and encouraged discussions among students by utilizing contents, illustrations, and new words of the book, followed by role-playing or writing activities, while the Silent Reading group students read books of their choice for 20 minutes a day without any follow-up activities. The control group conformed to the normal curriculum, focusing on an audio-lingual approach.

Pre and posttests eight months apart showed that students in the book flood groups (the Shared Reading and Silent Reading groups) demonstrated much greater gains than the control group students. Class 4 students did significantly better on their posttest in the reading comprehension test and the English structure test. Class 5 students showed significantly higher test scores in both the reading and listening comprehension tests. In their 12-month follow-up study, both receptive skills (i.e., reading and listening) and expressive skills (i.e., writing) had
improved. Elley and Mangubhai’s studies, with their findings of the diverse impacts of ER, was followed by numerous other empirical studies whose findings are addressed in the next section of this chapter in the order of vocabulary, reading, affect, and writing.

The Effectiveness of ER on L2 Vocabulary

Reading is an important means of expanding vocabulary size (Nation & Coady, 1988). Reciprocally, acquiring both breadth and depth of word knowledge is essential in order to be fluent in reading either in L1 or in L2 (Hudson, 2007). That is, a large vocabulary facilitates reading, and efficient word recognition further accelerates reading development (Koda, 2005; Laufer, 1992). In particular, reading a great deal allows multiple encounters with the same words and results in the expansion of sight vocabulary (Nation & Wang, 1999); “familiarity breeds automaticity” (Day & Bamford, 1998, p. 16). In addition, Huckin and Coady (1999, p. 182) argued that learners’ guessing of unknown words while reading can lead to vocabulary learning. They further enumerated the benefits of incidental vocabulary learning through ER as learning that is (a) contextualized; (b) pedagogically efficient, as it simultaneously assists both reading and vocabulary skills; and (c) individualized and learner-based. Nevertheless, while building a large recognition vocabulary is prerequisite for fluent and skillful reading, it is not easy for most L2 readers, who normally have limited exposure to L2 print through reading (Grabe & Stoller, 2002, p. 47).

Gains in vocabulary knowledge have been thus far one of the most studied areas with regard to the benefits of ER. Empirical evidence generally has supported ER as an aid to vocabulary learning as well as automaticity. In this respect, ER proponents stress the potential role of ER in expanding L2 vocabulary knowledge. They also contend that as engaging in ER
generally entails exposure to various genres of reading materials, general vocabulary knowledge can also increase along the way (Day & Bamford, 1998). Horst (2005) suggested that the true merit of an ER approach lies in offering ample opportunities through much repetition to increase the recognition of frequent vocabulary, not infrequent vocabulary. Therefore, the extent to which learners can increase their vocabulary growth through ER has been of interest to many researchers, as well as to practitioners concerned about the feasibility of integrating ER into their curriculum.

Previous studies (e.g., Laufer, 1992; Nation & Wang, 1999) have indicated that at least 95% of the running words in a text should be understood by learners in order for them to comprehend the texts. Later, Hu and Nation (2000) reported that learners should know 98% of the words in a text to comfortably read and comprehend the content with no help from dictionaries. Several studies have tapped into the importance of choosing appropriate ER materials with regard to learner comprehension and further supported the use of graded readers as a resource. For example, a corpus-based study by Nation and Wang (1999) specifically discussed the role and potential usefulness of graded readers in vocabulary development by examining coverage, density, and repetitions of vocabulary in one graded reader series, Oxford Bookworms. In the series’ 42 books, they found that, across all levels, almost 40% of the vocabulary occurred ten times or more, thereby providing a good chance of cumulative learning. The authors also suggested that learners read at least one graded reader per week to ensure they meet the same vocabulary again soon, building a threshold for vocabulary acquisition. This empirical finding seems to uphold Day and Bamford’s (1998) one-book-per-week suggestion. Furthermore, Nation and Wang (1999) found that words from the lower levels are repeated more at the higher levels, and thus suggested the benefit of reading more at the higher levels. As the
authors admitted, they cannot generalize this finding to all existing graded readers since the study investigated only one series of graded readers. However, the findings seem meaningful in that they show that well designed graded readers can indeed provide learners a good opportunity for vocabulary learning.

Day and Bamford (1998) suggested that children’s books can be good material for ER, but there has been little research on the effectiveness of using them for L2 adult learners’ vocabulary learning. Recently, another corpus-driven study (Webb & Macalister, 2013) compared lexical loads and suitability of texts written for L1 children and adults to graded readers. In this study, the School Journal published by the New Zealand Ministry of Education, which contains a collection of different genres such as poems, plays, and articles for 7 to 13 year olds, was chosen for children’s texts, and the Wellington Written Corpus (WWC) was selected for texts written for older readers. Thirty-three graded readers from the Oxford Bookworm series (levels 1 to 4) were chosen for the comparison. The authors found that 10,000 word families were needed to reach 98% coverage of the texts written for children and older readers; in other words, readers should know the most frequent 10,000 word families to reach 98% coverage as a lexical threshold to comprehend the texts easily. In contrast, the graded readers reached the same amount of coverage with only 3,000 word families. In addition, among the three, graded readers showed the highest repetition rate for word families outside of the most frequent 2000 word families, with a 24% repetition rate (i.e., 10 times or more), compared to 15% and 8% in the School Journal and the WWC, respectively (p. 313). These findings indicate that graded readers are also effective for learning lower-frequency vocabulary, and that texts for children provide a better chance of vocabulary learning than those written for adults. Webb and Macalister concluded that texts or books written for children might not be as suitable as graded readers for
L2 learners, who usually have a small range of vocabulary.

Two earlier studies further supported the importance of choosing the appropriate materials for the development of vocabulary. In Pitts, White, and Krashen’s (1989) study, where non-graded readers were used, the reading text was above the participants’ reading level and half of the students could not even finish a book, which deprived them of opportunities to learn words incidentally. In contrast, Day, Omura, and Hiramatsu (1991) found significant incidental vocabulary learning when students were allowed to read materials at their level.

Several case studies also have shown positive evidence that supports the benefits of ER for vocabulary learning. Pigada and Schmitt (2006) explored whether ER could lead to increased word knowledge in aspects of spelling, meaning, and grammar. A 27-year old Greek learner of French read a total of four books, spending approximately 60–90 minutes per book. The researchers interviewed the participant before and after the ER treatment to examine the acquisition of the target words in terms of the aforementioned three aspects. The target words included 70 nouns and 63 verbs (133 words in total), which were divided into six subgroups according to the number of encounters (i.e., 1, 2–3, 4–5, 6–10, 10+, and 20+) to explore the effects of frequency on vocabulary learning. Overall, a considerable amount of gain was found; some degree of learning was observed in 87 out of the 133 target words (65.4%). However, only spelling showed relatively strong enhancement, with a lesser extent of learning for meaning and grammar. Moreover, in terms of the effects of word frequency, echoing Nation and Wang’s results (1999), about 10+ exposures turned out to be sufficient to enhance vocabulary learning. Taking these studies together, it seems clear that the more encounters a learner has with words in texts, the more learning can occur.
Similarly, Cho and Krashen’s (1994) case study with four ESL learners examined whether vocabulary could be acquired from pleasure reading. A series written at the 2nd grade level, “Sweet Valley Kids,” was chosen for the ER material. Cho and Krashen noted that no specific amount of reading was required each day, as it was meant to be free voluntary reading. Therefore, the reading amount (ranging from 8 to 23 volumes) and the reading period (ranging from 2 weeks to 2 months) varied across the four participants. Nevertheless, all the learners acquired vocabulary at a rate more or less comparable to native speakers’ general rate per year of vocabulary acquisition through reading (i.e., 3000 words per year from reading one million words, according to Nagy, Herman, & Anderson, 1985, cited in Cho & Krashen, 1994).

The literature discussed so far endorses the view that learners can increase their vocabulary knowledge while engaging in ER. However, whether such gains in vocabulary lead to long-term retention was another concern for ER practitioners (Huckin & Coady, 1999). Several studies have thus attempted to answer this question. Waring and Takaki (2003) examined incidental vocabulary learning from reading graded readers through immediate, one-week, and three-month delayed posttests. Overall, students were able to learn new words from context after one hour of reading, but about half of the words were forgotten three months later, when measured through a word-form recognition test, a meaning test, and a multiple-choice test. On the immediate word-form recognition test, the students were able to recognize 15.3 words out of 25 in the reading; the more frequent the word, the higher the rate of recall. However, this recognition gradually faded as time passed. In terms of meaning, the students correctly identified 4.6 words of 25 in the immediate posttest, but this dropped to 0.9 words after three months. Words that appeared less than 15 times had almost no chance of being learned.
In a similar vein, Kweon and Kim (2008) investigated whether vocabulary could be incidentally learned and further retained through ER. Twelve Korean learners of English were assigned to read three young adult books over five weeks, reading three or four chapters per day at home. The most frequent 367 content words were divided into three word classes (i.e., noun, verb, and adjective) within three bands of frequency: 20 or more, 7–19, and 1–6 occurrences. The self-reported test results showed that the vocabulary knowledge of the participants increased significantly between the pretest and the immediate posttest in all three word classes, and the delayed posttest showed that this increased knowledge was largely retained one month later. As expected, the retention rate was higher for more frequent words than for less frequent words across all three word classes, stressing the important role of frequency in incidental learning. Incidental word learning through reading, with moderate retention rates, appears to be supported by this study. However, requiring students to read four to six hours per day and giving them a detailed comprehension quiz in each class does deviate somewhat from the core tenets of ER. Whether the vocabulary gain can solely be attributed to ER seems questionable in this study.

The general picture that emerges from these findings is that learners do acquire vocabulary, albeit moderately, through ER. However, as Pigada and Schmitt (2006) noted, whether reading alone can lead to successful vocabulary acquisition is still somewhat uncertain. More empirical studies are called for to address this issue, especially taking into account the findings that participants who consulted with dictionaries while reading improved their word knowledge more (e.g., Cho & Krashen, 1994) and that reading accompanied by vocabulary instruction led to better gains compared to reading-only instruction (e.g., Zimmerman, 1994).
The Effectiveness of ER on L2 Reading

In addition to the benefits of ER on vocabulary learning, Grabe and Stoller (2002) claimed that ER is one of the most effective ways to improve reading fluency. A strand of research has attempted to demonstrate the positive impact of ER on ESL or EFL learners’ reading rate and reading comprehension, which are critical components of fluent reading (Day, 2011; Grabe & Stoller, 2002). Researchers have argued that reading easy and interesting materials assists effortless word recognition (Grabe, 1991; Taguchi et al., 2004) as it can enable L2 readers to pay more attention to the meaning of the texts, giving them a better chance of learning from and comprehending the texts (Day & Bamford, pp. 18–19). However, findings have not been conclusive, not only because of a lack of empirical evidence but also because of problematic research designs (Beglar, Hunt, & Kite, 2012).

For example, Iwahori (2008) investigated the effectiveness of ER on the reading rate of high school students in Japan. Individual students increased their word-per-minute (wpm) by approximately 30% after the ER treatment. The fact that students made a significant improvement over a relatively short period of time (seven weeks) seems remarkable. However, as Iwahori admitted, whether the observed gain could be solely ascribed to ER is unclear, due to the absence of a control group. Moreover, the students’ reading comprehension was not tested, so the study tells us nothing about whether the improved reading rate was achieved by sacrificing comprehension.

Likewise, a series of research studies has attempted to show that ER is a more beneficial approach to developing basic reading rate and reading comprehension by comparing it with other reading approaches, such as skill-building instruction or intensive reading instruction. In Bell’s (2001) study, while the experimental group engaged in ER, the control group received intensive
reading instruction heavily based on reading short texts, answering comprehension questions, grammar, and vocabulary lessons. Both groups also did a great deal of reading in class and at home across two semesters. The results were surprising. The ER group was four times faster on the reading speed test (wpm) and had three times higher scores on the comprehension test than the traditionally taught control group. However, the results should be cautiously interpreted, as reading rate and comprehension were separately tested utilizing different texts. Again, it could be the case that the substantial gain in reading rate was achieved at the expense of comprehension.

Comparable results have been obtained in several other studies. Robb and Susser (1989) randomly assigned a fairly large number of students ($N = 125$) to two reading classes: an ER class and a skills-building class. In addition to in-class reading, the ER class students were required to read a minimum of 500 pages of unsimplified books at home along with short summaries. In contrast, the skills-building class spent most of their class time reading textbooks to learn reading skills, and this was done as homework as well. The post comprehension reading test scores were similar between the two groups, indicating that both types of reading instruction were effective. However, it is intriguing to note that the ER class scored higher on understanding the important facts and guessing vocabulary from context (p. 244). Furthermore, in line with Iwahori (2008) and Lai (1993), the reading rate of the ER class was significantly faster than that of their counterparts. In spite of these positive gains, however, it should be kept in mind that the ER class spent twice as much time reading at home than their counterparts. The ER group’s increased reading rate and comprehension ability could simply be the result of more exposure to texts.

In a similar manner, Al-Homoud and Schmitt’s (2009) study examined the effectiveness of an ER class by comparing it with a traditional class over a 10-week period. The traditional
class focused on intensive reading and vocabulary exercises, requiring students to read about 100 pages of a textbook to learn 800–1200 vocabulary items along with various reading skills. In contrast, the ER group simply provided students with 150 graded readers of various genres and required them to read at least one graded reader per week. Only the ER group made a significant, yet very small, gain on their comprehension posttest after 10 weeks. For reading rate, both groups improved significantly on their posttest, and the ER group students made significantly larger improvements than their counterparts. However, as in the studies by Iwahori (2008) and Bell (2001), the researchers did not test students’ comprehension of the texts used for measuring reading rate.

Beglar et al. (2012) pointed out the problematic research design of some of the previous research (e.g., no control group or absence of comprehension measures) and attempted to address these issues. Their study delved into the effectiveness of pleasure reading on the development of reading rate. While their intensive reading group showed an almost negligible reading rate gain, three pleasure reading groups achieved statistically significant gains over two academic semesters. To ensure that their reading rate gain was not achieved at the expense of comprehension, the reading rate tests used four narrative reading passages and were followed by multiple-choice comprehension questions. All groups maintained their high reading comprehension scores, which were above 80%, on the posttest. The study also found that whereas the amount of simplified text students read had a moderate, yet positive, correlation with their reading rate gain (Pearson correlations, $r = .39, p < .01$), the amount of unsimplified reading was negatively correlated with reading rate ($r = -.15, p = ns$).

These findings were upheld in Beglar and Hunt’s (2014) recent study. The amount of reading was generally found to be associated with reading rate gains. In particular, ER was more
effective for students at a low reading proficiency level whose reading rate was below 100 wpm. Also, it was confirmed that greater reading rate gains were associated with more simplified texts and fewer unsimplified texts. The authors further analyzed 14 pairs of participants who read the same number of standard words but two different text types: only simplified texts or both simplified and unsimplified texts. Interestingly enough, those who only read simplified texts showed significantly higher reading rates. Another noteworthy finding is that lower-level simplified texts were more advantageous than higher-level texts for the improvement of reading rates.

Taguchi et al. (2004) compared two approaches to facilitating reading fluency and comprehension: ER and repeated reading. While the ER group was encouraged to read graded readers they selected themselves, the repeated reading group read the same passage five times: a timed first reading, two readings while listening to an audiotape, and two silent readings. Both groups performed similarly on the post comprehension test. However, the repeated reading group showed a slightly higher reading rate (wpm) than the ER group in their post reading fluency test.

Taken together, accumulated empirical findings seem to indicate that an ER approach is either superior to or just as effective as an intensive approach for the improvement of L2 reading rates and comprehension. Also, the longer the experimental period, the greater the gains observed (Beglar & Hunt, 2014; also see Nakanishi, 2015 for meta analysis). As Beglar et al. (2012, p. 696) suggested, comparing various approaches in the same study (e.g., ER, timed reading, repeated reading, intensive reading) might shed more light on the pedagogic implications of methods of teaching L2 reading fluency and provide a more comprehensive understanding. In addition, as Yamashita’s (2008) study indicated, certain improvement can probably be manifested more quickly in reading than in other skills. While Japanese ESL
students’ reading comprehension ability significantly improved in this study, their L2 linguistic ability at the micro level such as vocabulary, spelling, and morphosyntax measured through cloze tests did not show significant gains after 15 weeks of ER engagement. Therefore, further research that can provide a more holistic view still seems to be necessary.

**The Effectiveness of ER on Affective Domain**

A good number of studies have also investigated whether the impact of ER could be extended to other domains such as affect. Numerous studies have attempted to explain how affective factors like motivation and attitude might interact with ER, which has its central tenet as the importance of reading for pleasure.

Day and Bamford (1998, p. 28) accounted for learners’ decision to read in the L2 through the expectancy value model, which consists of four main variables: materials, reading ability, attitudes, and sociocultural environment. They explained that the first two are related to the *expectation* of successful L2 reading, and the latter two concern the *value* attached to L2 reading. L2 reading is more strongly influenced by reading materials and attitudes than ability and environment. Nishino’s (2007) case study supported this claim. Nishino closely examined two beginning EFL learners’ motivational changes over two and a half years of ER experience, and found a strong relationship between the learners’ motivation and interesting reading materials. Reading materials were particularly related to the students’ “flow experience” (Csikszentmihalyi, 1990) and confidence. It is also intriguing to find out that changes in the students’ motivation over time were associated with reading materials, external factors (e.g., university entrance exams), and successful experiences of ER. Similar findings emerged from another longitudinal case study (Judge, 2011), which found that student motivation to read is influenced by the
interplay of intrinsic and extrinsic motivation without any discernable patterns, changing dynamically over time.

Likewise, drawing on Dörnyei and Ottó’s (1998) process model of L2 motivation, de Burgh-Hirabe and Feryok (2013) qualitatively analyzed learners’ motivational change over time, before, during, and after the ER treatment. Nine Japanese learners of English voluntarily participated in the out-of-class ER study. Based on participant interviews and journal entries, the researchers concluded that the participants’ motivation to read was dynamic, and that it changed due to various factors, including individual participants’ goals, self-regulation, and the availability of ER materials. Like Judge (2011), de Burgh-Hirabe and Feryok found that not only intrinsic values but also external demands, such as exam preparation and time-related concerns, were influential in the participants’ decisions to keep reading books. Another interesting finding of this study is that although the participants began with positive perceptions of ER that initially led them to engage in it actively, this was not sufficient to overcome other external demands and sustain their engagement in ER, regardless of their enjoyment or positive attitudes. This finding echoes Yamashita’s (2004) claim that merely having positive feelings about ER is insufficient for learners to keep reading.

In line with Mori’s (2002) finding that a belief in the *intrinsic value of reading and learning English* was one of the most significant predictors of the amount that L2 readers read, Takase (2007) found *intrinsic motivation* for L1 and L2 reading to be the most influential factor in reading. No positive relationship between L1 and L2 reading motivation was established, however. Takase’s follow-up interviews revealed that students’ L1 reading habits were not associated with their L2 reading habits, due in part to their different reading ability in the L1 and L2. Moreover, students responded that they tend to have flow-like experiences more when
reading in the L1 than in the L2. Recently, Kirchhoff (2014) specifically investigated the flow experience of learners while reading graded readers. Although students responded that they experienced flow during reading, their experiences seemed to be insufficient to stimulate them to keep reading. No correlation was found between frequency of flow experience and amount of time spent on reading. Nonetheless, it is worth noting that students’ flow experience was most influenced by the content of the book (e.g., fun and curious stories), which lends support to Day and Bamford’s (1998) claim that reading materials are one of the most important factors that affect L2 reading.

Rodrigo et al. (2014) investigated how different reading interventions (ER vs. no-ER) affected low-literate adults’ reading patterns. Over a period of three months, both groups’ responses indicated that they had very positive attitudes toward reading and enjoyed reading. However, only the ER group showed positive and statistically significant change in their reading habits, and more surprisingly, this was sustained even six months after the intervention. The authors contended that having a positive attitude toward reading may not be enough to develop reading habits, but that “a healthy library” (p. 85), easy access to books, and freedom to choose books together can contribute to the development of reading habits.

All these empirical findings seem to support Day and Bamford’s (1998) “extensive reading bootstrap hypothesis” (p. 30), which assumes that students’ successful reading experiences stimulate positive attitudes toward L2 reading, which in turn promotes motivation to read in the L2, entailing subsequent ER experiences. This cycle seems to be key in fostering a habit of reading, which is one of the primary goals of ER. In the same vein, Krashen’s (1995, 2004) pleasure hypothesis proposes that enjoyable activities facilitate language acquisition and that simply providing students with a variety of interesting reading materials and time to read can
result in increased motivation and positive attitudes toward reading. Cho and Krashen (2002) demonstrated that even a single positive experience with reading can be influential. In addition, empirical evidence from other studies has highlighted how ER helps reluctant readers become more eager readers (Cho & Krashen, 1994; Mason & Krashen, 1997). Day and Bamford (1998) also noted that L1 reading attitude is one of the factors in L2 reading attitude. This was exemplified in Yamashita’s (2004) study, which claimed that four L1 and L2 reading attitude variables—anxiety, comfort, value, and self-perception—are related, but have different degrees of transferability (p. 10). To be specific, the value that students ascribe to reading turned out to be more transferrable from L1 to L2 than anxiety and comfort. Moreover, L2 proficiency did not play an important role in the transfer of reading attitudes.

The Effectiveness of ER on L2 Writing

Although ER has gained a fair amount of attention and been researched actively for the past several decades, surprisingly few studies have inquired into the effectiveness of ER on writing. Nakanishi’s (2015) recent meta-analysis shows this paucity of research; among 34 studies included in the analysis, only one study, by Tsang (1996), explored the impact of ER on the writing domain of L2. Supporting earlier claims that students who read more read better and even write better (Krashen, 1984, 2004) and that reading books helps students “learn naturally, from context, and provide[s] excellent models of written English” (Elley & Mangubhai, 1983, p. 56), Janopoulos (1986) found that the amount of pleasure reading in L2 was correlated with L2 writing proficiency; a strong (.76) positive correlation was observed (Yule’s Q statistics). The amount of reading in L1 had no significant correlation with L2 writing proficiency. A few more empirical studies have followed up on this finding.
For example, Tudor and Hafiz (1989) implemented a three-month ER study in the UK as an extracurricular program with a small group of ESL learners of Pakistani origin. Sixteen students voluntarily met with the researcher after school for one hour, five days per week. Among the seven tests including reading and writing tests, interestingly, the students who attended the ER program improved markedly on the writing tests. In order to further examine the effectiveness of the treatment, the pre and post written production of the experimental group was analyzed in detail in terms of writing readiness (i.e., writing fluency), vocabulary range, syntactic maturity, and accuracy of expression. The students significantly improved in writing readiness and accuracy of expression. In contrast, T-unit length, sentence length, and the ratio of complex sentences pertaining to syntactic maturity decreased. In other words, the students produced more correct expressions but in simpler sentences. Tudor and Hafiz (1989) explained that this intriguing result might have been derived from reading the simplified language or short sentences in the graded readers, which may imply an association between output and input. Gains in syntactic, semantic, and spelling accuracy clearly indicate, however, that the ER program played an important role in consolidating the students’ command of English. Nevertheless, the results require careful interpretation. As the two ESL classes that served as a control group did not have any extracurricular classes, as the experimental group did, the improvement cannot be clearly explained as a result of the ER. On top of that, the writing of the experimental group was not compared to that of the control group.

Tsang (1996) compared the effectiveness of different writing programs (i.e., regular plus ER, regular plus frequent writing practice, and regular plus mathematics) on L2 writing with students attending an English medium secondary school in Hong Kong. Over 24 weeks, the ER program was required to read eight books and write reviews, while the writing program was
given additional writing tasks involving various contents and genres. The students’ pre- and post-tests were analyzed using five analytic scales. The ER program students did better, particularly in content and language use, compared to the other two groups. No statistically significant gains were reported in organization, vocabulary, and mechanics. It is noteworthy that the output-based class was less effective than the input-based class.

Similarly, Mason and Krashen (1997) explored the effectiveness of writing by dividing students into three groups: an English response group, a Japanese response group, and a comparison group. Whereas the comparison group practiced intensive reading with cloze exercises in class and at home, the two experimental groups devoted the same amount of time to ER and writing activities in L1 and L2. All students were asked to write summaries in English of the books that they read at the beginning and end of the semester, and two raters holistically scored the summaries to see if there were any improvements after one academic year. Interestingly, the Japanese response group made significantly better gains than both the English response group and the comparison group. However, unlike in Tsang’s (1996) study, the summaries were holistically evaluated on a scale of 1–6, from worst to best, and specific components (e.g., content, language use) were not separately considered. Mason’s (2004) more recent study with three groups—a Japanese summary group, an English summary group with corrective feedback, and an English summary group—corroborates his earlier findings with similar results. No significant differences were found among the groups in cloze tests, TOEIC, and error-free clause tests. However, the Japanese summary group made the greatest gains in the three measures. It is intriguing that writing summaries in their L1 was more helpful to the students than writing summaries in their L2.

As discussed, there have been a few attempts to show the association between ER and
writing improvement in L2, but findings have been less than conclusive. Therefore, the present study hopes to provide a more holistic view of the relationship between ER and L2 writing by analyzing the participants’ written production in depth. The third section of this chapter will discuss the methods employed in the present study to analyze the L2 writers’ texts.

L2 Writing Assessment

In general, analyses conducted in L2 writing research had three purposes: (1) locating areas of improvement related to instruction, (2) comparing different groups of writers (e.g., different proficiency groups), or (3) finding effects of specific types of interventions or tasks (Polio, 2012, p. 144). This section discusses writing assessment methods along with specific measures that have been commonly used in analyzing L2 written products for these purposes.

Holistic Assessment. Holistic scoring, which concerns the overall quality of an essay, is a measure in which raters assign a single grade based on their total impression of a whole text. The assumption that underlies this scoring approach is that “writing as a whole is more than the sum of its parts” (White, 1994, p. 231). When scoring holistically, raters are generally instructed to read an essay quickly, using a rubric, and not to focus on details too much, as it is thought that “with increased speed comes increased validity and reliability” (McColly, 1970, as cited in Charney, 1984, p. 74). Holistic scoring has made direct assessment (i.e., essay testing) more feasible because of its time efficiency and consequent cost effectiveness. It still remains a common way to assess a piece of writing. In particular, holistic scoring has been argued to be an appropriate and viable tool, not only for large-scale assessments such as placement or proficiency tests (e.g., the Test of Written English administered by the Educational Testing
Service) where a large number of students must be scored in a relatively short time (White, 1984), but also for classroom assessment and exit testing (Hayes, Hatch, & Silk, 2000).

However, one critical pitfall of holistic scoring lies in its high subjectivity, which threatens its reliability. Lack of consistency can derive from “different standards of severity” (p. 24) held by different graders or from one grader being inconsistent in applying the standards (Jacobs, Zinkgraf, Wormuth, Hartfiel, & Hughey, 1981). In addition, individual raters may have different values when it comes to good writing. Therefore, it is hard to avoid subjectivity and low inter- and intra-rater reliability in holistic rating. In order to achieve a higher degree of rating agreement by reducing subjectivity in the scoring of compositions, Jacobs et al. (1981, cited in Perkins, 1983, p. 654) suggest seven steps:

1. Adopt a holistic evaluation approach;
2. establish criteria to focus readers’ attention on significant aspects of the compositions;
3. set a common standard for judging the quality of the writing;
4. select readers from the same backgrounds;
5. train readers until they can achieve close agreement in their assessments of the same papers;
6. obtain at least two independent readings of each composition; and
7. monitor the readers periodically during the evaluation to check their consistency in applying the standards and criteria of evaluation.

In the same vein, White (1984, p. 404) suggested six similar procedures useful for achieving high reliability of scoring. He specifically asserted that more reliable results could be achieved by helping raters internalize the explicit scoring criteria. Supporting this idea of implementing rater training, several researchers (Casanave, 2004; Hamp-Lyons, 2001; Homburg, 1984; Reid, 1993b, Weigle, 2002) have argued for the importance of articulating clear criteria that raters can rely on consistently. Weigle (2002) emphasized that rater training including rater norming sessions with “a set of anchor or benchmark scripts” (p. 112) is of critical importance in holistic scoring to help raters conform to a rubric, which eventually leads to increased inter- and
intra-rater reliabilities. Without these anchor papers, the descriptors in the rubric may not be consistently interpreted (Weigle, 2007). However, Huot (1990) expressed concern that using rubrics along with speed reading may limit the raters’ natural process of reading, and thus “a personal stake in reading might be reduced to a set of negotiated principles” (p. 211). While acknowledging the importance of employing a reliable scoring guide, Huot argued that individual raters’ personal reactions emerging from a fluent reading process should not be overly restricted.

Another issue on which argument persists is that we do not really know how individual raters arrive at their decisions even when they are using the same criteria or rubric (Casanave, 2004). Cumming, Kantor, and Powers (2002, pp. 67–68) claimed that rater training along with monitoring of the scoring process can result in reliable holistic rating, but they pointed out, echoing Huot’s (1990) argument, that collapsing complex writing traits into single score points does not seem to reflect the richness of student writing. Moreover, such scoring obscures the specific judgments, or “decision-making behavior” (p. 68), used by each rater. Think-aloud protocols revealed as many as 35 decision-making behaviors (e.g., read or reread composition; assess reasoning, logic, or topic development; classify errors into types; see Cumming et al. 2002 for more information) of 17 raters on TOEFL essays. Handwriting was also found to influence raters’ judgment.

Analytic assessment. Although it is generally argued that holistic scoring is efficient, less time consuming, and therefore economical (Huot, 1990; Perkins, 1983; White, 1984), some scholars assert that analytic scoring is more useful in that it enables teachers to provide meaningful, organized feedback to students on their compositions by means of various criteria (Paltridge, Woodrow, Harbon, Phakiti, Shen, Stevenson, & Hirsh, 2009; Weigle, 2002, 2007).
With these advantages, analytic scoring is claimed to fit well into classroom writing assessments, especially for growth measurement (Wolcott & Legg, 1998). While, under holistic scoring, longer essays tend to receive higher grades with various subskills of writing overlooked, analytic scoring allows teachers to provide students with more customized diagnostic feedback, specifically identifying where their strengths and weaknesses lie. For example, Jacobs et al.’s (1981) ESL Composition Profile, one of the most commonly adopted analytic scales (Hamp-Lyons, 1990), guides teachers to assess student compositions by breaking them down into five subskills: content, organization, vocabulary, language use, and mechanics. This type of subdivided criteria can make rater training easier, and it is particularly useful for evaluating ESL learners’ performance, which tends to demonstrate uneven development of writing (Weigle, 2007). However, teachers should keep in mind that if categories are too narrow or complicated, students can be turned off and overwhelmed by the complex results (Wolcott & Legg, 1998). In this regard, researchers have advised teachers to be cautious not to overly criticize students’ writing (Paltridge et al., 2009) and not to allow a backwash effect to influence their instruction (Bacha, 2001).

**Portfolio assessment.** Portfolio assessment evaluates a student’s collection of written work to measure student growth in writing over an extended period of time. This method of assessment allows students to demonstrate their writing ability in various genres and registers under natural circumstances with no time constraints. Some scholars (e.g., Hamp-Lyons, 2011; White, 1994; Wolcott & Legg, 1998) therefore consider this to be a more valid approach that “relate[s] more directly to pedagogy” (Kroll, 1998, p. 231; see also Weigle, 2007) than a “snapshot approach” (Hamp-Lyons & Kroll, 1996, p. 53), which generalizes students’ writing ability based on a single essay written under time constraints. Because portfolio assessment
allows students to demonstrate their ability across different genres of writing through constant revision and editing, it allows teachers to assess their students’ “growth and achievement” in writing (Weigle, 2007, p. 204). With the movement from product-focused writing to process-oriented writing, portfolios have drawn attention as a teaching, learning, and assessment tool (Romova & Andrew, 2011).

Although some of the validity problems faced by writing assessment of one-time impromptu writing can be avoided by portfolio assessment (White, 2005), it still has some reliability issues (Weigle, 2007, p. 205). For example, a complex scoring process that involves the assessment of a mixture of essays containing different degrees of strengths and weaknesses makes it hard to assign a single score to a portfolio (Kroll, 1998; White 2005). Selected writing drafts from an individual student’s portfolio may vary a great deal in quality, adding to the difficulty of giving it a single score. In addition, portfolio assessment requires much effort and time from both teachers and students, which in turn leads to practicality issues.

**CAF Assessment.** With general agreement among scholars that second language performance is multifaceted in nature (Housen & Kuiken, 2009), the triad of complexity, accuracy, and fluency (CAF) has also been widely analyzed in L2 writing assessment. Most of these investigations utilize CAF as dependent variables in evaluating the effect of instruction on L2 performance (Pallotti, 2009), but more recently, since Skehan (1996) put forward a proficiency model, the triad itself has been a center of attention (Housen, Kuiken, & Vedder, 2012). Housen, Kuiken, and Vedder (2012, pp. 299–300) noted that complexity, accuracy and fluency each represent heuristic dimensions for guiding systematic inquiry and observation of L2 performance, proficiency and development and they are most frequently used as dependent variables to assess variation in these areas with respect to independent variables such as level of...
acquisition, stage of development, type of instruction, learning context or task features.

The analysis of CAF has been mostly carried out with oral production (Gunnarsson, 2012), and although it is argued that the processes of L2 speech and writing production have much in common (Ellis & Yuan, 2004), the following section focuses specifically on reviewing how the triad has been defined and what kinds of objective measures have been employed in assessing L2 written production. However, it has to be noted that unified or explicit definitions are unavailable despite the fact that CAF has become a prominent indicator of L2 performance or L2 knowledge, due in part to inconsistent findings and interpretations (Housen & Kuiken, 2009).

**Complexity.** In second language acquisition, complexity has been investigated in two broad strands of research. One strand explores the impact of different instructional methods on the acquisition of more or less complex structures of a target language. The other utilizes complexity “as a basic descriptor of L2 performance and as an indicator of L2 proficiency” along with accuracy and fluency (Bulté & Housen, 2012, p. 21). The current study belongs to the second strand, considering complexity as one of the characteristics of students’ writing performance. Ortega (2003) defined linguistic complexity, or, more specifically, syntactic complexity, as “the range of forms that surface in language production and the degree of sophistication of such forms” (p. 492). She claimed that syntactic maturity is an important index in that learners’ language development is displayed through their ability to appropriately use syntactic repertoires. That is, syntactically more advanced learners are more able to demonstrate varied and sophisticated structures in a piece of writing.

Traditionally, complexity measures have been based on the T-unit, which refers to “one main clause plus any subordinate clause or nonclausal structure that is attached to or embedded in it” (Hunt, 1970, p. 4). Since Hunt introduced this “minimal terminable unit” (p. 4) as an index
of growth of syntactic maturity, it has been one of the measures most widely and frequently used in assessing syntactic complexity of written and speech production in L1 and L2 development (Bardovi-Harlig, 1992). According to Wolfe-Quintero, Inagaki, and Kim’s (1998) substantive review of CAF measures, the T-unit complexity ratio (i.e., clauses per T-unit) was the measure most applied in the literature (e.g., Casanave, 1994; Hirano, 1991).

Scholars generally do acknowledge the usefulness of T-units, but some researchers have proposed other measures that are helpful in examining the syntactic complexity of written production. According to Hunt (1965, cited in Wolfe-Quintero et al., 1998, p. 70), written language develops from coordination to subordination and then to reduced phrases. That is, L1 children’s use of run-on sentences at early ages diminishes as they begin to use more sophisticated structures, such as subordinate or embedded sentences. Taking this developmental trend into account, Bardovi-Harlig (1992) argued that a sentence-based analysis is superior to a T-unit analysis as it reflects students’ syntactic knowledge more accurately; coordination in writing better captures L2 competence. In his study, the analysis of 86 advanced ESL students’ compositions at seven levels of proficiency showed that as the students’ proficiency increased, the number of coordinated sentences decreased. In keeping with this, Ishikawa (1995) reported that she found the ratio of clauses to sentences (C/S) to be the best index for measuring the growth of beginning level students’ syntactic maturity over a short period of time. Considering that learners at this level tend to produce fragments rather than full T-units, she asserted that the clause is a better index than the T-unit. Casanave (1994) applied this coordination index to create individual student profiles of writing improvement across three semesters; however, Casanave failed to observe a trend of decreased coordination with increased proficiency, instead finding an increase in coordination or little change. In Homburg’s (1984) study, the number of dependent
clauses per composition and the ratio of the number of dependent clauses per independent clause (DC/C) increased with proficiency, which supports Hunt’s (1965) earlier claim that subordination and relativization tend to increase as student proficiency increases.

Ortega (2003) examined 25 studies to explore syntactic complexity measures that have been used in college level ESL or EFL research. Overall, among the six most frequently used measures, ESL groups produced statistically significantly higher means than EFL groups in mean length of sentence and mean length of T-unit. Ortega also found that while almost negligible effect sizes were observed in studies of three months or less in EFL settings, relatively moderate changes were found in a group of ESL learners (e.g., Larsen-Freeman, 1983). Among the measures observed, the mean length of T-unit was the only measure employed by all six longitudinal studies; notably, the longer the study, the larger the effect sizes that were found (e.g., Casanave, 1994). Ortega (2003) pointed out, however, that syntactic complexity shows learners’ active knowledge of L2 structure, but these measures are neither “absolute developmental indices” nor “direct indices of language ability” (p. 494). That is, greater complexity is not necessarily a sign of more developed or better written performance. Polio (2001) also noted that complexity may indicate development in L2 but “not necessarily quality” (p. 97), because sentence variety rather than complexity contributes more to quality of writing. Norris and Ortega (2009, pp. 563–564) recently suggested that while T-unit analysis is suitable for intermediate or advanced level students who are capable of producing full sentences, clause-based analysis (e.g., coordinate index) is more suitable for low proficiency students.

**Accuracy.** Accuracy refers to learners’ ability to produce language free from errors, and as “the simplest and most internally coherent construct” (Pallotti, 2009, p. 4), it usually has been

---

1 Mean length of sentence, mean length of T-unit, mean length of clause, mean number of T-units, mean number of clauses per T-unit, and mean number of dependent clauses per clause.
measured in comparison with target-like production, that is, native speakers’ production (Wolfe-Quintero et al., 1998). From this perspective, deviations from native speakers’ production are strictly considered errors. Housen, Kuiken, and Vedder (2012), however, argued that such deviation as well as the decision of what constitutes norms can be relative in nature, and therefore, “A” in CAF should rather be interpreted as appropriateness and acceptability (p. 4). Likewise, Pallotti (2009) noted that it should be kept in mind that a grammatically accurate production does not always necessarily mean “communicatively effective” production. This argument is in accord with Casanave’s (2004, p. 68) view that accuracy per se does not embrace the accuracy of content.

The purpose of measuring accuracy has generally been to evaluate the effects of a certain pedagogic intervention, program, or task (Polio, 2001). The most commonly used measures for accuracy are the total number of error-free T-units (EFTs) (e.g., Casanave, 1994), the total number of error-free clauses (EFCs) (e.g., Ishikawa, 1995), or the total number of errors with classification (e.g., Kroll, 1990) or without classification (e.g., Zhang, 1987). Wolfe-Quintero et al.’s (1998) review indicates that relatively stable results have been reported with EFTs, which show a consistent tendency to increase with proficiency (e.g., Hirano, 1991; Homburg, 1984; Perkins, 1983). The EFT has been demonstrated to be a useful unit of measurement, particularly in distinguishing learners at different proficiency levels; however, careful interpretation is needed in part due to the fact that the scope of the term “error-free” has either not been explicitly defined or has varied from study to study. The type or magnitude of errors appears to have been decided depending mostly on researchers’ preferences. For example, in some studies (e.g., Henry, 1996), an error-free T-unit is accurate in morphosyntax, lexicon, spelling, and punctuation, whereas in others, an error-free T-unit needs to be free of only morphosyntactic and lexical
errors (e.g., Casanave, 1994) or only morphosyntactic errors (e.g., Arnaud, 1992). Critics have attributed inconsistent research findings on accuracy to this inconsistent application of definitions (Wolfe-Quintero et al., 1998).

Clause-related measures have been found useful as another type of accuracy index. Ishikawa (1995) measured low-proficiency EFL learners’ writing changes over eight weeks. Two groups were given the same picture story but different tasks: a guided writing task and a free writing task. Twenty-four objective measures were employed to analyze their pre and post writings. Both groups changed significantly on one measure, error-free clauses per sentence. The guided writing task group alone significantly changed on seven more measures, most of them clause-related. This seems to support Norris and Ortega’s (2009) suggestion that T-unit analysis may be unsuitable for low proficiency students because a certain level of development is necessary for T-units to be explanatory.

However, merely counting EFTs or EFCs can be criticized for not clearly showing how errors are distributed or what types of errors are involved within one T-unit (Bardovi-Harlig & Bofman, 1989; Kuiken & Vedder, 2007; Polio, 2001). In other words, the quantity of errors, but not their quality, or their nature, can be described by this kind of assessment (Polio, 2001). For example, one T-unit containing a single error is treated as the same as one with multiple errors. As such, “clusters of errors are lost in the quantification of the data” (Bardovi-Harlig & Bofman, 1989, p. 22). A few studies have analyzed types or degrees of seriousness of errors to address this issue. For example, Bardovi-Harlig and Bofman’s (1989) study counted the total number of errors per clause (E/C) along three parameters: syntactic (e.g., errors of word order),

\footnote{T-units and error-free clauses per composition, error-free clauses per T-unit, total words in T-unit, error-free clauses/total clauses, total words in error-free clauses, and words in error-free clauses/total words in clauses.}
morphological (e.g., errors in inflectional, nominal, or verbal morphology), and lexical (e.g., idiomatic) errors. Similarly, Homburg (1986) classified errors into three levels from minor to serious based on their effect on comprehensibility or readability.

Another issue is that counting EFTs does not distinguish students who produce accurate yet short T-units from those who take risks and produce more complex T-units but more errors. This distinction seem crucial, as there could be possible trade-off effects between accuracy and complexity in relation to learners’ risk-taking behavior (Wolfe-Quintero et al., 1998). Casanave (2004) also noted the possibility of an inverse relationship between accuracy and fluency: as “attention to one goes up, attention to the other goes down” (p. 68). However, there is a paucity of research on how the elements of the CAF triad are interconnected in L2 written development.

Taking these studies together, it seems to still be open to question whether there is a linear relationship between EFTs or EFCs and proficiency. Considering Polio’s (2001) argument that accuracy reflects writers’ attention as well as representing the explicit knowledge of L2 learners, whether accuracy per se can be interpreted with regard to L2 development remains unclear.

**Fluency.** Much research has been conducted on L2 oral fluency, but little on L2 written fluency (Polio, 2001). Traditionally, fluency has broadly referred to learners’ overall L2 proficiency (Lennon, 1990). Schmidt (1992) defined oral fluency as “the processing of language in real time” (p. 358), involving “speed and ease of processing” (p. 358) as common components. Schmidt therefore argued that fluency is based on learners’ “automatic procedural skill” that requires little attention or effort, tying it to a performance aspect. He noted that these features also characterize other modalities, such as writing.
Casanave (2004) specifically defined fluency as “writers’ ability to produce a lot of language (or to read) without excessive hesitation, blocks, and interruptions” (p. 67). In a narrower sense, Wolfe-Quintero et al. (1998) viewed written fluency as “a measure of the sheer number of words or structural units a writer is able to include in their writing within a particular period of time” (p. 14). However, Polio (2001) criticized this definition of fluency because including structural units (e.g., T-units or clauses) could make the interpretations of research findings arbitrary. For instance, students who wrote longer structures or compositions but fewer numbers of T-units or clauses would be considered less fluent.

The available definitions of fluency at least agree that speed and ease of writing characterize fluency in L2 writing. Moreover, reflecting the multidimensionality of fluency, various measures have been applied to it. Counting the number, length, or rate of production units (e.g., words, clauses, or sentences) has been a common way to measure fluency in writing (Wolfe-Quintero et al., 1998). Traditionally, the total number of words in a written production has been one of the widely used measures in assessing overall fluency of writing, and it has been shown that more proficient students tend to write longer compositions than less proficient students (Homburg, 1984). However, studies by Henry (1996) and Ishikawa (1995) that used total number of words as a measure found no consistent results. T-unit length (e.g., Homburg, 1984; Larsen-Freeman, 1983) and error-free T-unit length (e.g., Tedick, 1990) have rather consistently shown linear relationships with proficiency levels, irrespective of target languages or tasks (Wolfe-Quintero et al., 1998).

Chenoweth and Hayes (2001) examined burst length, “the production of language in parts” (p. 83), to measure fluency, assuming that writers with more linguistic experience would be able to retrieve words more easily. On the other hand, beginning level, or less fluent, L2
writers who are more occupied with spelling or word level concerns may have little space available in working memory to attend to content and organization, which characterize higher level skills in writing. Gunnarsson (2012) utilized words per burst (p. 251) as well, counting the number of words produced consecutively without pausing, in analyzing written production in real time in a computer-mediated test.

Whereas many studies have been carried out to probe L2 oral fluency (e.g., Lennon, 1990; Schmidt, 1992), comparatively few have addressed L2 written fluency. This may be in part due to the fact that correctness (i.e., accuracy) or content has usually been prioritized over fluency in written production (Lennon, 1990).

**Challenging issues of CAF.** Accumulated empirical evidence in general seems to have led to a relatively satisfactory agreement among researchers on the usefulness of the CAF triad in understanding the development of L2 writing, and as shown, diverse measures have been employed to gauge CAF. Nevertheless, some scholars have questioned the validity of these CAF objective measures, asserting that they do not assess writing ability per se, but simply measure mastery of writing conventions (Charney, 1984). Norris and Ortega (2009) noted that as complexity, accuracy, and fluency are dynamic, being “developmental in nature, growing, and changing all the time” (p. 556), it is challenging to provide a complete picture of how they develop in response to interlanguage system change. In addition, existing definitions of CAF are not uniform enough to capture the triad’s multicomponentiality, and, as a result, its operationalization as an indicator of L2 performance or proficiency raises another issue: How adequately, reliably, and validly can it be measured?

For example, Perkins (1983) argued that objective measures put little value on writing’s communicative function, which is manifested by a combination of numerous factors of writing,
such as coherence, organization, tone, and style. Also, the T-unit’s use as one of the most commonly applied measures has been criticized by some researchers because a T-unit analysis does not go beyond syntactic maturity, neglecting communicative competence, or the appropriateness of writing (Gaies, 1980), which is unquestionably an important dimension in target language development. T-unit analysis has also been criticized because it merely breaks up all sentences, “treat[ing] all conjunctions as semantically null” (Bardovi-Harlig, 1992, p. 302).

Rationale for Applying Various Assessment Methods

Various measures for assessing L2 learners’ written performance have been discussed in this section, largely divided into subjective (i.e., holistic and analytic ratings) and objective (i.e., CAF) measures. To sum up, the use of holistic or analytic measures based on impressionistic judgments of raters has been praised for high validity, yet there is a major pitfall in its subjectiveness. On the other hand, the measures of the CAF triad, which mainly rely on frequency counts, have been claimed to be highly objective and thus reliable, but these no-reading-for-meaning measures have been criticized for not providing an intuitive overview of written production. Polio (2012) argued that in order to obtain a holistic picture of interventional effects, utilizing one measure can be misleading.

Recently, while Pallotti (2009) agreed that CAF measures are useful for describing linguistic performance, he firmly asserted that adequacy, “the appropriateness to communicative goals and situation,” (p. 16) should also be considered as an independent construct in order for learners’ performance to be more meaningfully interpreted. From this perspective, evaluating written production using only the CAF triad is considered to be overly narrow. This seems in line with Paltridge et al.’s (2009) assertion. They argued that writing assessment measures should
consider not only linguistic features (e.g., accuracy, complexity) and quality (e.g., organization, coherence) but also practicality in relation to the need and purpose of the assessment (e.g., low stakes vs. high stakes assessment), as none of the single measures alone is perfect (Wolcott & Legg, 1998). There can be no doubt that employing only one measure cannot account for the multidimensionality of L2 writing performance or development. However, not many studies have employed both objective and subjective measures together to analyze L2 written work. Therefore, the present study employs both subjective and objective measures to analyze diverse aspects of features of written production.
CHAPTER III

METHOD

This chapter discusses the research methodology. First, the purposes of the study and the research questions are explained. The context of the study is then described in detail, and the data collection procedures are thoroughly described as well. The last section explains the statistical methods employed for the data analyses.

Research Purposes and Research Questions

One of the primary aims of the present study is to investigate whether exposure to a print-rich environment through ER has positive influences on the participants’ writing. Furthermore, if ER does affect writing, the study considers how the effects appear in student essays is a matter of interest; more specifically, the study observes to what extent and in what areas any improvement results from the ER treatment. As only a few studies have investigated the effectiveness of ER on writing to date, looking into the relationship between the two constructs is the first motivation of the present study.

Another important goal of the study is to explore student perceptions of and reactions to ER both in general and specifically in an EAP context. Because this study is a first attempt to investigate the possibility of incorporating ER into an academic writing class, whether the participants find ER to be useful in improving their writing skills as well as their overall English is of key interest. Furthermore, the study seeks to provide pedagogic suggestions for how ER can be meaningfully implemented to create more reading and writing practice in an EAP context, thereby enhancing students’ literacy development. With these aims in mind, the study investigates the following research questions:
1. How does extensive reading affect ESL students’ overall quality of writing?

2. How does extensive reading impact ESL students’ writing development in terms of complexity, accuracy, and fluency?

3. What are students’ general reactions to and perceptions about the usefulness of ER?

4. What are the benefits of incorporating extensive reading into an EAP writing class?

Due to the exploratory nature of this study, it does not test particular hypotheses. Nevertheless, each research question was formulated with some expectations about what might be found:

1. ER is expected to result in some quality differences between the treatment group and the control group, with certain aspects of writing more influenced than others.

2. ER is expected to have an impact on students’ complexity, accuracy, and fluency in L2 writing; however, different degrees of change or improvement are anticipated.

3. Students are expected to react favorably to ER in general, but some degree of reluctance or negative attitude is also anticipated depending on individual student perceptions and experience.

4. The integration of ER and writing activities is expected to create diverse opportunities to aid students’ development of reading and writing in English.

**Participants**

The participants were 84 ESL students at the University of Hawai‘i at Mānoa in the United States. According to the University, international and immigrant students who do not
meet the University’s exemption criteria\(^3\) must take the placement test. The test is composed of writing, listening, and reading sections, and administered by the university’s English Language Institute (ELI) whose primary purpose is “to provide English instruction for students who have been admitted to the university and who do not speak English as a native language…in order to facilitate their academic studies” (http://www.hawaii.edu/eli). Based on the test results, some students are required to enroll in specific ESL courses offered by the ELI; they may be required to take from one to six courses in the areas of listening and speaking, reading, and writing in English. If students demonstrate adequate English proficiency on the test, they are exempted from taking ESL courses. For the writing portion of the placement test, three raters read individual students’ essays to decide on their placement in either the intermediate level (ELI 73) or the advanced level (ELI 83) class.

The participants for the present study were students enrolled in ELI 73, *Intermediate Academic Writing*. This course was selected for several reasons. First, while ER research has mostly been conducted with and found benefits of ER for students at a low level (e.g., Al-Homoud & Schmitt, 2009; Nishino, 2007), there is a paucity of research dealing with intermediate learners. Another reason was to meet one of the primary goals of the study, which is to explore the possibility of integrating ER into an EAP context. Because students at this level were assumed to need to develop both their general and academic writing abilities as they transition to an academic community, they were thought more suitable for the purposes of the study than students who had reached an advanced level of academic writing. The goals of ELI 73 described in the course syllabus also align with the purposes of the study:

\(^3\) A score of 100 or above on the internet-based TOEFL, a score of 250 or above on the computer-based TOEFL, or a score of 600 or above on the paper-based TOEFL exempts students from taking the placement test.
ELI 73 is designed for undergraduate and graduate students...who need to develop their general and academic writing abilities in English as an additional language. Students in this course will explore differences between personal and academic writing, develop written fluency, and improve command over textual, rhetorical, and discursive conventions common in academic writing.

Due to the limited number of students enrolled in ELI 73 per semester, the data collection took place over three consecutive semesters: Fall 2013 (2 control classes and 1 treatment class), Spring 2014 (1 treatment class), and Fall 2014 (1 control class and 1 treatment class). Six ELI 73 classes participated in total; three classes were assigned to be control groups and three to be treatment groups. Students attended this 16-week writing class for one semester, amounting to 32 face-to-face class meetings. Each class lasted for 75 minutes.

Table 1

Demographics of the Participants

<table>
<thead>
<tr>
<th>Nationality</th>
<th>N</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Japanese</td>
<td>32</td>
<td>38%</td>
</tr>
<tr>
<td>Korean</td>
<td>22</td>
<td>26%</td>
</tr>
<tr>
<td>Chinese</td>
<td>14</td>
<td>17%</td>
</tr>
<tr>
<td>Taiwanese</td>
<td>4</td>
<td>5%</td>
</tr>
<tr>
<td>Spanish</td>
<td>2</td>
<td>2%</td>
</tr>
<tr>
<td>Thai</td>
<td>2</td>
<td>2%</td>
</tr>
<tr>
<td>Finnish</td>
<td>1</td>
<td>1%</td>
</tr>
<tr>
<td>German</td>
<td>1</td>
<td>1%</td>
</tr>
<tr>
<td>Icelandic</td>
<td>1</td>
<td>1%</td>
</tr>
<tr>
<td>Indonesian</td>
<td>1</td>
<td>1%</td>
</tr>
<tr>
<td>Russian</td>
<td>1</td>
<td>1%</td>
</tr>
<tr>
<td>Slovak</td>
<td>1</td>
<td>1%</td>
</tr>
<tr>
<td>Swiss</td>
<td>1</td>
<td>1%</td>
</tr>
<tr>
<td>Vietnamese</td>
<td>1</td>
<td>1%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Gender</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Female</td>
<td>55</td>
<td>65%</td>
</tr>
<tr>
<td>Male</td>
<td>29</td>
<td>35%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Academic Status</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Undergraduate</td>
<td>62</td>
<td>74%</td>
</tr>
<tr>
<td>Graduate</td>
<td>22</td>
<td>26%</td>
</tr>
</tbody>
</table>
Of the 95 students enrolled in ELI 73 across the three semesters, 84 students’ data (44 in the treatment group and 40 in the control group) were included in the study, because some students missed either the pretest or the posttest, which were essential to the data analysis. Participants included undergraduate and graduate students, and they had diverse cultural and academic backgrounds. The demographic information of the 84 participants is summarized in Table 1. The majority of the students came from Asia, with the greatest numbers from Japan (38%), Korea (26%), and China (17%). Sixty-five percent were female and 35% were male, while 74% were undergraduates and 26% were graduate students. Among the 62 undergraduate students, 44 (71%) were exchange students who had come to the U.S. for a study abroad program of one or two semesters.

The length of time the students had studied English varied, ranging from a minimum of 6 years to a maximum of 20 years with a mean of 10 years and a median of 10 years. The length of their residence in English-speaking countries also varied: less than a month \((n = 68, 80.95\%)\), between one month to six months \((n = 9, 10.71\%)\), and more than a year \((n = 7, 8.33\%)\). Most of them were new to the university as either freshmen or first year graduate students.

**Instruments**

For data triangulation, three different types of data were collected: surveys (i.e., a background survey and an extensive reading survey), writing tests (i.e., a pretest and a posttest), and interviews (i.e., post interviews). This section provides a detailed description of each instrument along with the rationale for using it.

**Background survey.** A background information form was used to obtain participants’ demographic data (see Appendix A). The study used the ELI in-house background information
form, which also asks for personal information, such as names and email addresses. However, this information was neither used for research purposes nor included in any of the dissertation.

The survey also asked students about their years spent learning English and length of residence in English-speaking countries, as it was assumed that their English learning experiences would be as diverse as their backgrounds. Moreover, as a brief needs analysis, the survey asked students to identify their personal goals for this course (e.g., improving organization, grammar, etc.) with regard to their perceived strengths and weaknesses in writing in English. These questions were included in the survey in an attempt to learn what students wanted to get out of the course and to what extent the course fulfilled their expectations.

An additional set of questions was related to students’ reading and writing in L1 and L2 to ensure that students in the control and treatment groups were not very different from each other in terms of their reading and writing habits at the onset of the data collection. For instance, students were asked about their attitude toward English reading and writing in general, as well as to give a rough estimate of the average amount of time they spent reading and writing in English per week for both personal and academic purposes.

**Writing test.** The ELI 73 diagnostic test is normally administered during the first week of the semester to familiarize course instructors with their students’ strengths and weaknesses and to ensure that students are in the appropriate level of class. Thus, this in-house test was used as a test instrument for the study; test time and instructions were slightly modified for the data collection (see Appendix B). Students were instructed to write an essay for 60 minutes on the writing prompt, “What is your attitude toward writing?” Along with this question, the test materials included five different authors’ perspectives on writing, which the students could use in
presenting their points of view. The evaluation criteria were briefly introduced on the test instruction handout and also verbally explained to students before they began to write.

This particular choice of topic was made for several reasons. First, the topic is general and familiar to everyone. In a timed test, a familiar topic can allow students to demonstrate the writing skills that are truly representative of their overall writing ability (Kroll & Reid, 1994). Second, the topic is culturally accessible and does not require any content knowledge; selecting a prompt that all students have equal access to is crucial (Horowitz, 1991). Third, this topic allows the researcher to learn about individual students’ writing attitudes at the beginning and the end of the semester, offering additional valuable information for the study. Finally, this topic allows students not only to reflect on their perceptions about writing but also to argue for or against the provided authors’ points. In contrast to the narrative essay topics used in previous ER studies, for example, “My Family” (Lai, 1993) or “My Favourite Person” (Tsang, 1996), the present study selected a more formal and level-appropriate topic that could allow students to demonstrate their skills in both general and academic writing, which also aligned with one of the aims of this writing course.

Both the control group and the treatment group completed this essay test at the beginning (pretest) and the end of the semester (posttest) in a timed and supervised setting. The pretest was intended to ensure that there was no significant difference in terms of English writing ability between the two groups at the onset of the study, although the students had already been evaluated as intermediate level based on their placement test results. The same essay test with the same prompt was administered as a posttest after 15 weeks. The same prompt was used to allow the two tests to be readily compared to assess writing improvement. That is, the present study aimed to (1) measure how much students in both groups had improved after one semester of
writing course and (2) compare improvement between the treatment group and the control group on different aspects of writing.

**Extensive reading survey.** An extensive reading survey was developed to inquire into the treatment group students’ perceptions about ER (see Appendix C). This retrospective survey focused mainly on two constructs: (1) students’ attitudes toward ER: enjoyment (items 1, 2, 7, 8, and 12), comfortableness (items 5 and 6), confidence (item 10 and 19); and (2) students’ evaluation of the usefulness of ER in light of their English learning across different skills: reading (items 3 and 13), writing (items 4, 9, 14, 15, and 16), vocabulary (item 11), speaking (item 17), and listening (item 18). Nineteen items to be rated on a 5-point Likert scale were included in the survey. Relatively high reliability was achieved (Cronbach’s alpha = .898). The survey ended with five open-ended questions to investigate students’ more in-depth thoughts and ideas. These questions asked students to freely write about their one semester of ER experience in terms of their attitudes, likes, dislikes, and perceptions of ER’s usefulness, along with any suggestions they might have.

**Post interviews.** As Brown (2001) noted, “the flexibility of interviews allows the interviewer to explore new avenues of opinion in ways that a questionnaire does not” (p. 78). This study included interviews with the treatment group students in addition to the survey with the aim of exploring their reactions to ER in detail. Because this study integrated ER into an EAP writing class, which had not been attempted or reported in mainstream literature before, the extent to which the students perceived ER as useful is of great interest. Whether students felt that they had benefited from the integration of reading and writing through ER was also explored. Control group students were also interviewed in order to examine their thoughts about ELI 73 in general as well as their attitudes toward and experiences with reading and writing in English.
The interview led by asking general questions first, with follow-up questions asked when necessary (see Appendix D for a full list of interview questions). However, rather than asking the questions one-by-one, I aimed to have a casual conversation, bringing up the questions naturally during the talk. The hope was that the students would talk more freely about their feelings and thoughts if the atmosphere were comfortable and conducive to storytelling.

All interviews were carried out when the semester was over to avoid any influence of the teacher-student relationship on the interviewees’ responses to the questions. Students voluntarily participated in the interviews; they were informed of their rights as research participants and asked to read and sign a consent form (see Appendix E). The interviews were conducted in a library on campus, and each lasted approximately 30 minutes to an hour. The interviewees were compensated for their time with a $5 or $10 gift card depending on how long the interview lasted. All the interviews were audio-recorded and later transcribed.

**Data Collection Procedures**

The study is a semester-long, longitudinal study; data collection took place at the beginning and at the end of each semester to compare the effects of the main treatment. In week 1, before any type of writing instruction had taken place, the background survey and the pretest were administered during the regular class period in both control and treatment groups. In week 16, when the treatment was over, students in both groups took the posttest. Only the treatment group students were asked to answer the extensive reading survey. Once the semester was over, volunteers from both groups were interviewed to further investigate students’ thoughts and ideas about ER as well as about their reading and writing experiences. Table 2 below summarizes the design of the present study and the instruments employed.
Table 2

*Design of the Study*

<table>
<thead>
<tr>
<th></th>
<th>Control Group</th>
<th>Treatment Group</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><em>(n = 40)</em></td>
<td><em>(n = 44)</em></td>
</tr>
</tbody>
</table>

*Week 1*
- Background Survey ✓ ✓
- Essay Test (Pre) ✓ ✓

*Week 2-15*
- Writing Instruction ✓ ✓
- Extensive Reading ✓

*Week 16*
- Essay Test (Post) ✓ ✓
- Extensive Reading Survey ✓

*After semester*
- Post Interview ✓ ✓

I taught the treatment groups as the researcher and instructor, and other ELI writing instructors taught the control groups. Taking into account the nature of this classroom-centered research where instructor and student interaction plays an important role, it would have been better if the same teacher had taught all the classes; however, this was not possible. Therefore, in efforts to ensure that student experiences in both groups were as similar as possible, the classes shared the same features, including lectures, major writing assignments, class activities, and pair or small group discussions. Moreover, while no two groups can be exactly the same, the instructors worked to ensure that the two groups received the same kind writing instruction as nearly as possible, with all instructors using the same lesson materials along with the same course syllabus and weekly schedule. Furthermore, the ELI 73 instructors observed each other’s classes. Such observation is in fact required in the ELI as a professional development activity to provide a chance for the instructors to learn from their colleagues and reflect on their own teaching skills. All the participating ELI 73 instructors used the observations as another opportunity to ensure that the design of their lesson plans and activities was as similar as possible.
The course included common features of EAP writing instruction, such as freewriting, summarizing, and paraphrasing, as well as practice involving various genres such as personal writing in addition to academic writing genres. Also, both groups used the process approach to writing, and therefore class activities included planning, drafting, and revising recursively, with less focus on grammar and syntax.

**ER workshop.** On the first day of class, students were informed that the control group was designed as a traditional writing class whereas the treatment group would spend a certain amount of time in and outside class for ER and associated short writing activities. Students were given the freedom to choose a class section that they preferred. Almost all students preferred to stay in their original class; some students may not have had a choice but to remain due to inflexible course schedules. However, none of the students expressed dissatisfaction about their class selection.

A brief workshop was provided to the treatment group on the first day of class to introduce the concept of ER and explain how the course would integrate ER into class activities. To avoid a possible Hawthorne effect, however, students were not told that ER might be useful for improving their writing. Day and Bamford’s (2002) ten ER principles, on which the present study draws, were explained to the students using handout material. Students were also periodically reminded about these principles throughout the semester.

In order to make sure students read books within their reading ability, which is one of the central tenets of ER, they were introduced to how to choose appropriate books for their level during the workshop. First, students were allowed to choose books they wanted to read by, for example, looking at the covers, reading through the blurbs on the backs, or reading the first couple of pages. Then, they were instructed to read the first page or a random page of the book.
Following Nation’s (1990) suggestion, students were advised to go one level down if they found more than three to five unknown words on one page. They were also advised to check out higher-level books if they found their first choice too easy.

In sum, following the ten principles, students were allowed to freely choose any book that suited their interest, and more importantly, they were also instructed to read books within their reading ability at their own pace. It is worth mentioning that, following principles 9 and 10, the instructor of the treatment groups played multiple roles, not only as a facilitator who guided and encouraged students to engage in reading throughout the semester, but also as a reader, or ER participant, who read books and shared reading experiences with the whole class. Teacher modeling has been advocated by ER practitioners (e.g., Day & Bamford, 1998; Stoller, 2015) as a crucial component of successful ER that helps students see the value of reading. In addition, it was hoped that such participation would help to create a sense of a whole-class reader community. Every effort was made to provide a pleasant atmosphere in which students could comfortably engage in reading.

**ER library.** Because of the limited linguistic ability of L2 learners, the most popular reading material for ER programs has been language learner literature such as graded readers. While Day and Bamford (1998) noted that carefully selected children’s books, newspapers, magazines, and even comics can also be good sources for ER materials, a recent corpus-driven study by Webb and Macalister (2013) showed that graded readers are more suitable for L2 readers than children’s books (i.e., books written for learning L1) because graded readers are written around specific sets of word families to enhance efficient vocabulary learning. Webb and Macalister argued that texts written for children may not be as efficient as graded readers for improving vocabulary learning and reading comprehension due to heavier lexical demands and
more “children’s genre-specific vocabulary” (p. 314) which, according to the authors, are less common in texts or books written for older readers (e.g., fairy, dragon, p. 314).

The creation of the classroom library for this study took both arguments into consideration. The library contained mainly graded readers from levels two to six (300 to 1800 headwords), published by Oxford (i.e., The Oxford Bookworms), Cambridge (i.e., Cambridge English Readers), and Penguin (i.e., Penguin Readers). Books written for 3rd to 6th grade L1 English children (i.e., young readers) were also included, based on the assumption that those who could relatively easily finish short and easy texts such as the graded readers might want to challenge themselves with these slightly longer and more difficult books. Moreover, the inclusion of such books allowed the library to meet the various topical interests of the students and to provide space for further learning, thereby helping L2 learners gradually expand their “reading comfort zone” (Day, 2011, p. 10). The importance of providing students with a variety of reading materials was also supported by Pilgreen (2000), who emphasized the necessity of offering texts that students “not only want to read—but can read” (p. 9, emphasis in the original) and making the classroom library “genuinely inviting” (p. 9) to students.

The classroom library contained 258 books altogether (graded readers: 219 books; young readers: 39 books), with a wide range of topics and genres including both fiction (e.g., classic, romance, suspense, science fiction, mystery) and nonfiction (e.g., biography, history). Moreover, students were free to bring their own reading materials (e.g., books, magazines) if they did not find any book that interested them in the classroom library. To ensure that the outside materials complied with the core ideas of ER (i.e., reading easy materials within students’ linguistic levels), the students were advised to discuss their reading materials with the instructor. As there was no lockable cabinet in the classroom, the classroom library had to be based on a delivery system...
due to security concerns. Ten minutes before the class, around 100 books were attractively arranged at the front of the classroom, divided into the two categories of fiction and nonfiction, for browsing and checking out.

**In-class ER.** The first 20 minutes of each class was assigned to ER in the treatment group; 15 minutes of individual and silent in-class ER was followed by five minutes of pair or group discussion in the treatment group. This replaced a 20-minute freewriting activity in the control group, which was usually composed of 10 minutes of freewriting and 5 to 10 minutes of group discussion on the freewriting topic.

**Out-of-class ER.** Both groups were required to spend an average of one and a half to two hours on homework per week; however, the control and treatment groups received different homework assignments. While the control group was given textbook-based homework each week (e.g., textbook reading, writing brief responses or answers, grammar practice, etc.), the treatment group was required to continue to read their chosen books and complete a short writing activity at home as an extension of their in-class ER. This type of out-of-class reading, according to Day and Bamford (1998), is good for building a reading habit in the long term. Textbook-related homework was occasionally given to the treatment group, but this was very limited.

**Writing activities.** Fifteen minutes of ER in class was followed by five minutes of pair or group discussions utilizing the students’ freewriting homework. This freewriting activity, termed “the 10-minute writing” in the study, was directly related to the books that the students were reading each week. The activity was designed based on the assumption that if they were asked to write about what they were reading, they would have more to say and therefore more to write. In addition, by providing them the chance to clarify what they learned from the reading, this writing activity aimed to help students better understand the texts they read, which might further
stimulate their interest in continuing to read. Moreover, this activity was expected to encourage students to voice their opinions in writing more freely than they might when asked to produce a correct analysis, which is usually the case for most academic reading and writing activities. Furthermore, given that many ESL students tend to be very concerned about making grammatical mistakes, doing this worry-free type of writing was assumed to contribute to the development of writing fluency, which is one of the main goals of ELI 73.

Writing book reports or summaries has been used in ER studies as a post activity that allows teachers to monitor their students’ reading and students to express their feelings about the books (e.g., Mason, 2004; Robb & Susser, 1989; Takase, 2007; Yamashita, 2008). However, in order to motivate engagement with the books and create diverse writing opportunities, the present study provided a variety of writing topics each week. This list of examples illustrates the types of topics and questions that students were given:

- To help students practice summarizing: “Summarize the story so far, including the main themes, events, and characters”
- To help students practice describing: “Choose characters in the book that you either liked or disliked, and explain why. You can describe their personality or behavior in relation to some events in the story”; “Describe one of the most important events in the story and explain how the main character dealt with or reacted to that specific incident”
- To evaluate students’ overall understanding of the text while stimulating their creative writing: “Write new or different endings for the story. If you were the author, how would you end the story?”; “If you could create a character or a thing to make the story better or more interesting, who or what would that be?”
- To help students engage more closely with a book: “Choose one character in the story.
Then, think of one gift that you want to give that character, and explain briefly why you have chosen that particular gift

• To promote students’ critical thinking by analyzing similarities and differences: “Write about cultural aspects that you found different from yours”; “Can you relate the story to your life? If you were in the same situation, would you act differently from the character(s) in the story?”

• For students who read nonfiction: “Did the book provide the information that you expected? To what extent is the provided information similar or different from your knowledge about the topic?”

This series of questions designed for the 10-minute writing seems simple but was devised to create ample opportunities to interact with the texts while responding, summarizing, describing, and analyzing the content. In addition, these types of writing are commonly encountered in the academic setting, and it was assumed that letting the students practice it with easy reading materials might be useful for them, especially at this intermediate level. Most importantly, these topics were intended to give the students more personalized or customized writing practice closely related to their own choice of books each week. The full list of questions used for the 10-minute writing activity is provided in Appendix F.

**Book recording form.** The treatment group students were required to briefly record their weekly reading progress (e.g., book titles, levels, pages, and time spent on reading) on a book recording form (see Appendix G). This book recording form was designed by adapting Ken Schmidt’s reading recording form, reproduced in Bamford and Day’s (2004) book, *Extensive Reading Activities for Teaching Language* (p. 78). One of the purposes of using this form was for the instructor to keep track of how much time individual students read each week, along with
how many pages and what kinds of books they read. Another underlying reason lay in the assumption that keeping records of their progress could be a factor in motivating “self-initiated reading,” which is at the heart of ER. The recording form was also useful to compare the amount of time that the different groups spent on their weekly assignments.

One or two books per week, or 1.5–2 hours per week, were recommended, and extra credit (2 points) was given to those who maintained this throughout the semester. Students were also encouraged to set their own reading goal as a personal motivational factor at the beginning of the semester. Fifteen percent of the course grade for the treatment group was for ER in and outside of class and the 10-minute writings, and the same amount of the course grade was for freewriting activities in the control group. Moreover, in order to make sure that both groups spent a comparable amount of time on homework per week, the control group students were asked to briefly record their assignments and keep track of the time they spent on them in a weekly assignment log (see Appendix H).

**Teacher journal.** As the instructor, I kept a journal throughout the semester to reflect on the ER activities implemented in the treatment class. Analyzing students’ reading behaviors in detail was not the primary purpose of the study, but regularly writing about the students’ in-class ER and their pair or whole class discussions, along with recording intriguing events and my own reactions and ideas, improved my observations of the progress of the class. In addition, reading students’ 10-minute writings enabled me to indirectly observe the degree of student enjoyment and engagement with ER. Together, reading these short writings and keeping my own journal provided me with insights that helped me make better-informed decisions or judgments about ER-related class activities (e.g., formulating topics for the 10-minute writing) for the students as the semester progressed. Writing a journal regularly about which things went well or did not go
well was a learning process for me as an instructor; it was useful for designing class activities in accordance with students’ attitudes, and it contributed to the progress of the class.

**Rating Procedures**

In order to examine the impact of ER on students’ writing improvement over one semester, the students’ pre- and post-tests were analyzed employing three assessment tools: holistic rating, analytic rating, and objective rating. Holistic and analytic ratings were used to evaluate the overall effectiveness of the written communication, focusing on whether a message was conveyed clearly and efficiently in the essay. For objective rating, various measures were used to specifically examine complexity, accuracy, and fluency of individual essays, in order to gain a more complete picture of students’ writing performance.

All pre- and post-tests collected in both control classes and treatment classes were mixed together and coded with a new identification number based on a random distribution. To ensure blind rating, any information that might identify students or pre- and post-tests was removed from the essays. Next, all essays were typed in Microsoft Word documents to avoid any influence of handwriting, either good or bad, especially on the holistic judgment, which largely relies on first impressions (Hayes et al., 2000). Spelling and grammatical errors were preserved, however.

**Rater training for holistic and analytic ratings.** To grade student essays holistically and analytically, three experienced ESL instructors, in addition to the researcher, were recruited as raters. As all the raters were either current or former instructors in the ELI, pursuing a graduate degree in applied linguistics, they were familiar with the context as well as the expected proficiency level of students in ELI 73.
Due to the large number of essays, not all were scored by all of the raters. Of the 168 coded essays—the pre and post essays of 84 students—raters 1 and 3 each scored 66.67%, and raters 2 and 4 each scored 50% of the essays. To ensure inter-rater reliability, approximately 20% of the essays were scored by all four raters.

A two-hour training session was held for the raters to learn the rating protocol. The purposes of this session were (a) to guide the raters to clearly understand the rating procedures, (b) to give the raters opportunities to familiarize themselves with the rubrics so as to reach acceptable inter-rater reliability, and (c) to finalize the scoring criteria based on a shared understanding among raters. Because it is important to familiarize raters with the criteria and terms used in the descriptors to increase inter-rater reliability (Weigle, 1994), sample papers, which served as “anchor papers,” were prepared to help the raters practice their holistic and analytic judgments before they did any actual rating. These anchor papers had been written by previous ELI 73 students on the same prompt. During the training, the raters practiced scoring by applying the provided criteria to the anchor papers. Based on the discussion and suggestions from the raters, the holistic and analytic criteria were slightly modified (e.g., rewordings for clarification).

The training emphasized that the raters should refer to the criteria each time they worked on rating the essays in order for each rater’s scoring to remain consistent (i.e., intra-rater reliability). The raters were also asked to score the essays first holistically and then analytically. Because holistic scoring is strongly influenced by first impressions, analytic scoring could contaminate raters’ holistic judgment if it were done first. In addition, the raters were encouraged to leave comments on the areas where they were not certain. That is, if they found themselves to be on the borderline between scores, they were instructed to choose the one that they were
leaning toward and leave comments to remind themselves later about their reasons for their decision when discussing any discrepancies in scores.

All materials, including the anchor papers, the rubrics, and detailed guidelines along with the assigned set of student essays, were provided to the raters as a packet. Each rater scored the assigned essays individually. It took between 10 and 20 hours for each rater to complete the rating, which they did over a period lasting from one to three weeks. They were asked to record their scores on a provided Excel spreadsheet.

A holistic scoring rubric was created by adapting two resources: (a) the ELI in-house writing hallmarks for the writing placement and diagnostic tests and (b) the TOEFL Internet Based Test rubrics developed by Educational Testing Service (https://www.ets.org/s/toefl/pdf/toefl_writingrubrics.pdf). The completed rubric was based on a six-point scale, with one being incomplete and six being very good, or excellent. Before the rubric was put into use for the actual data, it was tested using the anchor papers and further refined reflecting the raters’ suggestions. Each possible score (1–6) was accompanied by a detailed description of four to five qualities that an essay needed to possess to achieve that score. As this was the scale for holistic scoring, meeting two to three qualities sufficed for the essay to receive the corresponding score (see Appendix I).

An analytic scoring rubric was adapted from Jacobs et al.’s (1981) ESL Composition Profile, which has been widely used in L2 writing studies. This rubric breaks down essays into five sub-skills of writing—content, organization, vocabulary, language use, and mechanics. The five sub-skills in the original version are differently weighted in proportion to their relative importance, with content being the most heavily weighted (30%) and mechanics being the least (5%). However, the present study assigned equal weighting to all categories (ranging from 7 to
20 points) because no clear rationale for the different weighting is provided by Jacob et al. (Kondo-Brown, 2002; McNamara, 1996) and because the present study is more interested in the individual categories, and comparing student abilities across the five categories, than in the sum. The category of mechanics, however, which only concerns spelling and punctuation errors, was assigned a score range from two to five. This decision was made because mechanics was assumed to provide little information that might distinguish student ability, given that spelling and punctuation are less important at this level (see Appendix J).

**Rater training for objective ratings.** In order to capture the multiple dimensions of L2 writing performance, the student essays were also analyzed using a variety of objective measures to examine three different constructs: complexity, accuracy, and fluency (CAF). Unlike the holistic and analytic ratings, the objective measures analysis was done by only two raters (rater 1 and rater 4), because objective rating is less susceptible to an individual rater’s subjective judgment. Therefore, rater 1 scored about 75% of the essays, and rater 4 scored about 50% of the essays based on a random distribution.

Detailed guidelines of the CAF measures are provided in Appendix K. The guidelines were borrowed from Polio (1997); however, a few tricky cases found in the anchor papers were added to the guidelines along with example sentences. The rater training for objective measures was held on a separate day from the training for holistic and analytic rating but followed the same training protocol. Moreover, the same anchor essays were used for the training to help the raters become familiar with and practice the CAF measures selected for the study.
Data Analysis

**Quantitative analysis.** To answer the first research question, the essays were holistically and analytically rated. Holistic scores gathered from the four raters were first averaged for each student’s pretest and posttest. To ensure that writing performances of the treatment group and the control group were not significantly different from each other, a $t$-test was conducted with the pretest scores to ensure group equivalence. Then, an independent samples $t$-test was conducted with the students’ posttest scores to compare the two means to determine whether the difference of posttest scores was statistically significant at the alpha level of .05. Descriptive statistics were used to learn to what extent the two groups were different from each other. Effect sizes were also reported referring to Cohen’s $d$.

For analytic rating, descriptive statistics were reported first. Next, line graphs and box plots were created to visually show the distribution of the data and, more specifically, to display to what extent the two groups were different in terms of the five categories—content, organization, vocabulary, language use, and mechanics. Central tendency and dispersion were described for each category. Furthermore, a Rasch measurement using FACETS was computed with the analytic scores. Given that four different raters scored essays across five different categories, the FACET analysis was done to provide more details on student ability, rater severity, and intra-rater reliability.

To address the second research question, various measures were employed to analyze the complexity, accuracy, and fluency of the student essays. In particular, ratio-based measures were chosen as they are much more indicative than frequency-based measures (Wolfe-Quintero et al., 1998). Two measures were chosen for evaluating syntactic complexity. First, the ratio of clause to T-unit (C/T) was employed, as this measure has shown a linear relationship to students’
proficiency (Wolfe-Quintero et al., 1998). Second, the ratio of clauses to sentences (C/S) was also used, based on Ishikawa’s (1995) argument that sentence-level complexity is more informative, especially in regard to lower-level students’ syntactic maturity. It should be noted that the definition and scope of clauses varies across the previous studies; the present study followed Polio’s (1997, p. 139) specific guidelines for determining clauses.

Accuracy (i.e., the ability to be free from errors) was measured by calculating error-free T-units per T-unit (EFT/T), which has shown a stable correlation with proficiency in a number of previous studies (e.g., Casanave, 1994; Hirano, 1991). Error-free clauses per T-unit (EFC/T) and error-free clauses per sentence (EFC/S) were also employed, as these have been found to be useful for measuring changes over a relatively short period of time (Ishikawa, 1995).

Written fluency was analyzed by counting the total number of words (WT) of the timed pre- and post-tests. Words per minute (WPM) was therefore additionally taken into account. Moreover, words per T-unit (W/T), words per clause (W/C), and words per sentences (W/S) were calculated, given that several empirical studies have found these measures to have a linear relationship to proficiency.

To answer the third research question, descriptive statistics were used to analyze the extensive reading survey and to explore students’ feelings about and perceptions of the usefulness of ER and the 10-minute writing. The survey answers were further analyzed qualitatively, along with the interview data.

Qualitative analysis. Thirteen students from the treatment group voluntarily participated in interviews to share their thoughts and opinions about ER based on their one semester experience. Students who particularly liked ER, so called bookworms, and students who did not show much interest in ER were included. Interviewees were also recruited from the control
group to explore their general thoughts about ELI 73 as well as their reading and writing habits.

Five students from the control group voluntarily participated.

The interviewees were assured that neither their names nor any other identifying information would be revealed in this dissertation or any other products of the research. One student from the treatment group felt nervous about having her interview recorded, so I took notes during the interview. All other interviews were audio recorded, and the recordings were transcribed for data analysis.

Open-ended interview questions were designed to fulfill the purpose of the interview; however, as mentioned, I tried to embed these questions in our conversation, so that interviewees could talk about their feelings naturally. The interviewees were encouraged to reconstruct their experience and inner voice by means of storytelling, as the interview aimed to “find out what their experience is and the meaning they make of it” (Seidman, 2006, p. 128). I asked follow-up questions when I felt clarification, concrete details, or examples were necessary.

As a process of coding the data, I first marked interesting passages as I read them. Next, I grouped them into several categories that emerged from the interviews. The grouped categories were then further developed into thematic connections, which are presented and discussed in the results chapter.
CHAPTER IV

RESULTS

This chapter reports the findings of the study. The quantitative reports include the findings of the statistical analyses used for examining students’ writing performance based on the three different ratings (i.e., holistic, analytic, and objective). Graphs and tables are also used to visually present the differences between the treatment group and the control group. The qualitative reports summarize the findings of the interviews, which aimed to explore students’ thoughts and perceptions about the usefulness of ER in their English learning, particularly in reading and writing.

Writing Performance: Holistic Ratings

In order to examine students’ overall writing improvement, holistic scores gathered from the four raters were averaged for each student’s pretest and posttest. Interrater correlations among the raters were examined first with the Pearson product-moment correlation coefficient, which found an acceptable range of intercorrelations, from .756 to .810 (see Table 3). Because each essay was rated by two raters, the Spearman-Brown prophecy formula\(^4\) was used to estimate the approximate reliability of all four raters combined, and it suggested a high reliability \((r_{xx} = .925)\). The intrarater reliability of each rater will be discussed later in this chapter along with the analytic rating results.

An independent samples \(t\)-test was also conducted with the pretest essay scores to confirm that the two groups had similar writing ability before any type of instruction or treatment

---

\(^4\) Brown (2005, p. 187): \(r_{xx}= (\text{full-test reliability}) = n \times r / (n - 1)r + 1. r = \text{the lowest estimate of interrater correlation. } n = \text{number of times the test length is to be increased.}\)
took place. There was no significant difference between the control \((M = 3.30, SD = .76)\) and the treatment group \((M = 3.01, SD = .75)\); \(t(82) = 1.710, p = .091\).

Table 3

<table>
<thead>
<tr>
<th></th>
<th>Rater 1</th>
<th>Rater 2</th>
<th>Rater 3</th>
<th>Rater 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rater 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rater 2</td>
<td>.756*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rater 3</td>
<td>.789*</td>
<td>.789*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rater 4</td>
<td>.775*</td>
<td>.795*</td>
<td>.810*</td>
<td></td>
</tr>
</tbody>
</table>

*Note. *\(p < .01\)*

Table 4 summarizes the descriptive statistics for the holistic rating. The mean of the treatment group increased from 3.30 to 4.18, and that of the control group increased from 3.01 to 3.31. Considering the possible score range, from 1 (Incomplete) to 6 (Excellent), the mean scores of the pretest indicate that students in the two groups generally performed poorly. The two positively skewed data sets from the pretest also confirm students’ relatively poor writing ability. Notably, the most frequently observed score, the mode, for both groups was 3 (poor).

However, as evidenced by the skewness, the data distribution of the two groups shifted to a negative skewing, which confirms that students on the whole made progress over the semester and did better on the posttest. Furthermore, as is clearly shown in Table 4, the treatment group performed better than the control group on the posttest. Both mode and median values increased from 3 to 4 in the treatment group, while those values increased less, from 3 to 3.50, in the control group. In addition, no student in the treatment group but five students (12.50%) in the control group received a score of 2 (very poor) on the posttest. No student in either group received an extreme score of 1 (Incomplete) or 6 (Excellent) on either test.
Table 4

Descriptive Statistics for Holistic Rating

<table>
<thead>
<tr>
<th></th>
<th>Treatment Group</th>
<th>Control Group</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( n = 44 )</td>
<td>( n = 40 )</td>
</tr>
<tr>
<td>( M )</td>
<td>3.30</td>
<td>3.01</td>
</tr>
<tr>
<td></td>
<td>(55.00)</td>
<td>(50.17)</td>
</tr>
<tr>
<td>Mode</td>
<td>3.00</td>
<td>3.00</td>
</tr>
<tr>
<td></td>
<td>(50.00)</td>
<td>(50.00)</td>
</tr>
<tr>
<td>Median</td>
<td>3.00</td>
<td>3.00</td>
</tr>
<tr>
<td></td>
<td>(50.00)</td>
<td>(50.00)</td>
</tr>
<tr>
<td>( SD )</td>
<td>0.76</td>
<td>0.75</td>
</tr>
<tr>
<td></td>
<td>(12.67)</td>
<td>(12.50)</td>
</tr>
<tr>
<td>Std. Error</td>
<td>0.11</td>
<td>0.12</td>
</tr>
<tr>
<td></td>
<td>(11.33)</td>
<td>(13.33)</td>
</tr>
<tr>
<td>Skewness</td>
<td>0.34</td>
<td>0.41</td>
</tr>
<tr>
<td></td>
<td>(-0.26)</td>
<td>(-0.36)</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>-0.20</td>
<td>-0.54</td>
</tr>
<tr>
<td></td>
<td>(-0.74)</td>
<td>(-0.06)</td>
</tr>
<tr>
<td>Range</td>
<td>3.00</td>
<td>2.75</td>
</tr>
<tr>
<td></td>
<td>3.50</td>
<td>3.00</td>
</tr>
<tr>
<td>Min</td>
<td>2.00</td>
<td>2.00</td>
</tr>
<tr>
<td></td>
<td>3.00</td>
<td>2.00</td>
</tr>
<tr>
<td>Max</td>
<td>5.00</td>
<td>4.75</td>
</tr>
<tr>
<td></td>
<td>5.50</td>
<td>5.00</td>
</tr>
</tbody>
</table>

Note. Possible score ranges are 1–6. Numbers in parentheses represent percentages.

Graphics are used to visually represent different gains across the two tests of the two groups. Line graphs display the overall improvement by simply comparing the pretest scores and the posttest scores. In addition, boxplots (or “box-and-whisker plots”) are used to clearly display the data spread along with an indication of central tendencies; they are especially useful for comparing data sets such as those for the pretest and posttest of the two groups in this study.

Boxplots characterize data in terms of the 25th (lower quartile, \( Q_1 \)), 50th (median, \( Q_2 \)), and 75th (upper quartile, \( Q_3 \)) quartiles. The middle 50% of the data values (IQR; interquartile range = \( Q_3 – Q_1 \)) falls between the 25th percentile and the 75th percentile. The central bold line drawn inside the box indicates the median of the distribution. The whiskers, reaching to the largest (upper whisker) and smallest observed score (lower whisker), give additional information about the spread of the data. Outliers, observed outside the upper and lower whiskers, are marked with
dots. Asterisks represent values that are even further out. These outliers and anomalies are values that lay more than one and a half times the length of the box from either end of the box, that is, values at either above \( Q_3 + 1.5 \times IQR \) or below \( Q_1 - 1.5 \times IQR \). If the distribution of the boxes and whiskers around the median is relatively symmetrical, the data are normally distributed. Conversely, if the distribution is asymmetric, with the median cutting through the boxes unequally, the data set is skewed (lopsided). Boxplots are particularly useful in this study to clearly show the degree of dispersion and skewness of the data set. In the following discussion, line graphs and boxplots display the data in terms of percentages.

![Figure 1. Average Writing Scores based on Holistic Rating](image)

The line graph in Figure 1 presents the students’ writing performance over the semester. The steeper rise of the treatment group line indicates a greater gain on their part. That is, the treatment group’s average writing ability was scored at around 55.43% on the pretest but increased to 69.38% on the posttest, whereas the control group’s average score increased from 49.80% to 55.79%. The overall gain in the mean score (from pretest to posttest) made by the
treatment group \((M = 0.88, 14.77\%)\) is much greater than that made by the control group \((M = 0.30, 4.90\%)\).

Figure 2. Boxplots for Holistic Rating

In Figure 2, boxplots display the two groups’ data distribution across the two tests. The medians are the same as 50.00% on the pretest. However, the boxplots of the treatment group show that almost 50% of the pretest data is clustered in the upper box (i.e., 50th –75th percentiles), while the control group’s median cuts through the center of the boxes. The same pattern was revealed with the posttest data. The treatment group’s median is placed higher than the control group’s median (58.33%), and the boxes of the two groups rarely overlap.

Moreover, an independent samples \(t\)-test showed that the difference in the posttest scores between the treatment group \((M = 4.18, SD = .68)\) and the control group \((M = 3.31, SD = .80)\) was statistically significant, \(t (82) = 5.411, p = .000, 95\%\ CI [0.554 to 1.197]\). Levene’s test
indicated that the assumption of homogeneity of variances was tenable ($p = .332$). The magnitude of the effect of the intervention, ER, was explored by calculating Cohen’s $d$ (Cohen, 1988), which defines an effect size of .2 as small, .5 as medium, and .8 as large. The effect size for this study ($d =$ mean of the treatment group – mean of the control group/pooled standard deviation) turned out to be large ($d = 1.172$). It is noteworthy that a relatively short period of in- and out-of-class ER exerted a large effect on writing performance. Taking the findings for the holistic rating together, one semester of ER benefitted students’ overall writing performance, and the effect was large.

**Writing Performance: Analytic Ratings**

To help explain which specific aspects of writing the ER treatment had a positive effect on, this section presents the statistical analysis of the analytic ratings of the student essays, in which the essays received scores on five categories—content, organization, vocabulary, language use, and mechanics. Table 5 displays means, standard deviations, and 95% confidence intervals of these five categories of analytic rating. Given that the category of mechanics was on a different scale (1–5), percentages are also provided so that improvement across the five categories is readily comparable.

As shown in the table, the treatment group did slightly better than the control group on the pretest across the categories, and the mean difference between the two groups became even greater on the posttest. The smaller standard deviations of the treatment group in each category also suggest that their scores cluster more closely to the average than the scores of the control group.
Table 5

*Descriptive Statistics for Analytic Rating*

<table>
<thead>
<tr>
<th></th>
<th>Treatment Group $(n = 44)$</th>
<th>Control Group $(n = 40)$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pre (%)</td>
<td>Post (%)</td>
</tr>
<tr>
<td><strong>Content</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$M$</td>
<td>13.57</td>
<td>80.63</td>
</tr>
<tr>
<td>$SD$</td>
<td>2.11</td>
<td>8.37</td>
</tr>
<tr>
<td>95% CI lower bound</td>
<td>12.93</td>
<td>78.08</td>
</tr>
<tr>
<td>95% CI upper bound</td>
<td>14.22</td>
<td>83.17</td>
</tr>
<tr>
<td><strong>Organization</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$M$</td>
<td>13.52</td>
<td>79.74</td>
</tr>
<tr>
<td>$SD$</td>
<td>2.07</td>
<td>8.65</td>
</tr>
<tr>
<td>95% CI lower bound</td>
<td>12.89</td>
<td>77.11</td>
</tr>
<tr>
<td>95% CI upper bound</td>
<td>14.15</td>
<td>82.38</td>
</tr>
<tr>
<td><strong>Vocabulary</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$M$</td>
<td>13.48</td>
<td>77.56</td>
</tr>
<tr>
<td>$SD$</td>
<td>1.58</td>
<td>7.78</td>
</tr>
<tr>
<td>95% CI lower bound</td>
<td>13.00</td>
<td>75.19</td>
</tr>
<tr>
<td>95% CI upper bound</td>
<td>13.96</td>
<td>79.92</td>
</tr>
<tr>
<td><strong>Language Use</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$M$</td>
<td>13.55</td>
<td>76.25</td>
</tr>
<tr>
<td>$SD$</td>
<td>1.65</td>
<td>7.34</td>
</tr>
<tr>
<td>95% CI lower bound</td>
<td>13.04</td>
<td>74.02</td>
</tr>
<tr>
<td>95% CI upper bound</td>
<td>14.05</td>
<td>78.48</td>
</tr>
<tr>
<td><strong>Mechanics</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$M$</td>
<td>4.17</td>
<td>92.27</td>
</tr>
<tr>
<td>$SD$</td>
<td>0.41</td>
<td>9.12</td>
</tr>
<tr>
<td>95% CI lower bound</td>
<td>4.05</td>
<td>89.50</td>
</tr>
<tr>
<td>95% CI upper bound</td>
<td>4.30</td>
<td>95.04</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$M$</td>
<td>58.29</td>
<td>79.35</td>
</tr>
<tr>
<td>$SD$</td>
<td>6.90</td>
<td>7.21</td>
</tr>
<tr>
<td>95% CI lower bound</td>
<td>56.19</td>
<td>77.13</td>
</tr>
<tr>
<td>95% CI upper bound</td>
<td>60.39</td>
<td>81.58</td>
</tr>
</tbody>
</table>

*Note.* The score range is 1–5 for mechanics and 7–20 for all other categories.

In accordance with the holistic gains, the treatment group’s average gain ($M = 9.16$, 10.77%) across the pre and posttests was much more than that of the control group ($M = 2.37$, 2.78%). More specifically, the treatment group made gains in the order, from highest to lowest, of content ($M = 2.56, 12.76\%$), organization ($M = 2.43, 12.15\%$), vocabulary ($M = 2.03, 10.15\%$),
mechanics ($M = 1.70, 8.86\%$), and language use ($M = 0.44, 8.52\%$). The control group showed the same pattern but with smaller gains; their biggest gain was in content ($M = 0.88, 4.38\%$), followed by organization ($M = 0.58, 2.91\%$), vocabulary ($M = 0.53, 2.65\%$), mechanics ($M = 0.30, 2.63\%$), and language use ($M = 0.13, 1.50\%$). These different degrees of gain in the five categories are presented individually, along with graphics in order to visually represent the differences of the two groups.

Figure 3. Average Scores for Content

Moreover, the Pearson correlation coefficient showed that the holistic scores and the sum of the analytic scores were strongly correlated for the pretest, $r (82) = .84$, and for the posttest, $r (82) = .90, p = .00$, which confirms that students who gained higher scores on the holistic rating tended to receive higher analytic total scores.

In addition, the intercorrelations among the analytic categories were examined. As shown in Table 6, all the categories were associated with each other from a moderate to a high degree. In particular, the correlation between content and organization was highest in both groups, indicating that students’ abilities in these two areas may be somehow strongly correlated. The
correlation between vocabulary and language use was the second highest. Mechanics showed the weakest correlations with the other categories.

Table 6

*Intercorrelations among the Five Analytic Categories*

**Treatment Group**

<table>
<thead>
<tr>
<th></th>
<th>Content</th>
<th>Organization</th>
<th>Vocabulary</th>
<th>Language Use</th>
<th>Mechanics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Content</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Organization</td>
<td>.921*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vocabulary</td>
<td>.804*</td>
<td>.803*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Language Use</td>
<td>.666*</td>
<td>.694*</td>
<td>.813*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mechanics</td>
<td>.273</td>
<td>.369</td>
<td>.458*</td>
<td>.461*</td>
<td></td>
</tr>
</tbody>
</table>

**Control Group**

<table>
<thead>
<tr>
<th></th>
<th>Content</th>
<th>Organization</th>
<th>Vocabulary</th>
<th>Language Use</th>
<th>Mechanics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Content</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Organization</td>
<td>.868*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vocabulary</td>
<td>.737*</td>
<td>.700*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Language Use</td>
<td>.572*</td>
<td>.508*</td>
<td>.796*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mechanics</td>
<td>.543*</td>
<td>.548*</td>
<td>.656*</td>
<td>.621*</td>
<td></td>
</tr>
</tbody>
</table>

*Note. *p < .05

**Content.** Figure 3 clearly illustrates that, in terms of content, the two groups demonstrated similar writing performance on the pretest (TR: 67.87%; CT: 63.31%), but the gap between the groups became greater on the posttest, with the treatment group outperforming the control group by 12.84% (TR: 80.63%; CT: 67.79%).
Similarly, the boxplots in Figure 4 display the two groups’ performances in terms of content across the pre and posttests. The IQRs of the two pretest boxes largely overlap (with their medians at 67.50% for the treatment group and 62.50% for the control group), indicating that the level of the students’ writing performance at the beginning of the semester was similar. Moreover, the asymmetry of the pretest boxes around their medians in both groups indicates that the data is not normally distributed, being skewed to the right. In contrast, the treatment group clearly outperformed the control group on the posttest, with these boxes showing only a minimal overlap of the IQRs, suggesting that it is likely that there is a difference between the two groups. Except for one outlier, the spread of the treatment group’s data is also more condensed around the median than that of the control group.

**Organization.** Figure 5 displays the average scores of the students’ writing organization. The line graph confirms more improvement for the treatment group than the control group. The
treatment group improved by 12.15% (from 67.59% to 79.74%), while the control group demonstrated a smaller gain, 2.91% (from 63.81% to 66.72%).

Figure 5. Average Scores for Organization

Figure 6 illustrates that both groups had a wide range of scores on the pretest; the boxplots overlap, and the two groups have similar medians (TR: 66.88%; CT: 65.00%). Conversely, the boxes do not overlap on the posttest, showing a large median difference (TR: 80.00%; CT: 65.00%). The distribution of the treatment group’s data shows a negative skewing with one outlier below the lower whisker, while the control group is positively skewed, as evidenced by a longer whisker and a box on the upper side. It is also worth noting that although the control group students improved on the posttest, as demonstrated by the slight increase in their data range, their median remains the same as for the pretest (65.00%).
Vocabulary. A similar trend was observed in the area of vocabulary (see Figure 7). The pretest showed little difference between the two groups, with means of 67.41% (TR) and 64.16% (CT). However, on the posttest, while the treatment group scored an average of 77.56%, the control group’s mean was 66.81%.

The boxplots illustrating this difference are shown in Figure 8. On the pretest, the boxes of both groups overlap with the other’s medians (TR: 67.50%; CT: 65.00%). Both groups’ data are positively skewed, but the control group shows a wider range of IQR around the median. In contrast, somewhat of an opposite trend is observed on the posttest. Notably, the two boxplots for the posttest scores barely overlap, with a large difference between the medians (TR: 77.50%; CT: 67.50%). Also, compared to the pretest, the treatment group demonstrates a wider range of IQR, which is negatively skewed, while the control group shows a narrower IQR with nice symmetry of whiskers and boxes around the median.
Figure 7. Average Scores for Vocabulary

Figure 8. Boxplots for Vocabulary
**Language use.** Students in both groups improved least in language use, although their proficiency was at a similar level with other categories on the pretest. While the treatment group gained 8.52%, from 67.73% to 76.25%, the control group only improved by 1.5%, from 12.61% to 12.91%, as illustrated by the flattened line graph in Figure 9.

![Figure 9. Average Scores for Language Use](image)

In Figure 10, the boxplots also represent the different degrees of improvement of the two groups. For the pretest, the boxes of the two groups overlap with the other’s median (TR: 65.00%; CT: 62.50%), with both being positively skewed. However, on the posttest, the distribution of the data seems close to a normal distribution. Notably, the box of the treatment group appears much higher than that of the control group, and they barely overlap each other. A large median difference is observed as well (TR: 76.88%; CT: 65.00%).
Figure 10. Boxplots for Language Use

Figure 11. Average Scores for Mechanics
Mechanics. As seen in Figure 11, compared to the other four categories discussed above, in mechanics, the treatment group and the control group demonstrated relatively high average ability on the pretest (TR: 83.41%; CT: 79.75%), and they achieved even higher scores on the posttest (TR: 92.27%; CT: 82.38%). In Figure 12, the boxplot of the treatment group shows that almost all the data are clustered in the 50% to 75% range on the pretest. The treatment group also showed a large improvement across the two tests, with their median changing from 80.00% to 95.00%. On the other hand, the control group’s boxplot of the pretest represents a wide range of IQR around the median, showing large data variability. However, on the posttest, the data are clustered in the 50% to 75% range, exhibiting an almost identical shape to that of the treatment group on the pretest. That is, there is no change in their median, 80.00%; however, about 50% of the data falls in the upper quartile. Two outliers are observed below the 25th percentile.

Figure 12. Boxplots for Mechanics
**FACETS Analysis of Analytic Rating**

The analytic rating results were additionally analyzed using a multi-faceted Rasch measurement. This measurement has been found useful in assessing rater behavior (e.g., Kondo-Brown, 2002; McNamara, 1996; Winke, Gass, & Myford, 2012), as rater performance or rating criteria can threaten the validity of writing assessment (Winke et al., 2012). Rater training has become an essential part of writing assessment, providing one viable solution to this issue; however, even with rater training, concerns have still been raised because “an emphasis on rater consensus may force raters to ignore their own experience and expertise in judging writing” (Weigle, 1998, p. 263). Proponents of the Rasch measurement argue that, while they recognize the importance of rater training, raters cannot be trained to have the same degree of severity. This approach acknowledges rater variation as an unavoidable element in the rating process (Weigle, 1998). Moreover, the Rasch measurement has been argued to be particularly useful in that it also allows the effect on the scores of variables such as person ability, rater behavior, and categories to be examined simultaneously (Brown & Edmonds, 2012, p. 76). Because this study employed different raters and analytic rating criteria, FACETS analysis was thought to be useful to show their impact on students’ essay scores. The analysis was done using FACETS version 3.71.4 (Linacre, 2014).

The three facets included in the analysis were ability of students, severity of raters, and the five analytic categories. These facets were reported on a logit scale, which refers to “an interval which allows locations (of items and persons) to be determined either individually or in relation to other items, persons, or locations” (Brown & Edmonds, 2012, p. 77). Therefore, all facets can be represented by using a common logit scale. The average is set at zero logits, with a standard deviation of 1.
FACETS analysis also provides fit statistics that show how well persons, raters, or categories adhered to an expected pattern of responses, indicating whether the data fit the Rasch model. If items or persons respond in an expected manner, it is considered *infitting*, and if otherwise, it is termed *misfitting*. Brown and Edmonds (2012) noted that misfitting indicates “(a) examinees with abilities outside the possible range of scores or other examinee issues (like fatigue, inattention, etc.), (b) raters with insufficient training or attention to the task, or (c) categories with poorly written descriptors” (p. 77). Generally, fit statistics of 1.5 or greater are considered too much unpredictability in raters’ scores, whereas .5 or less indicates not enough variation in the scores (Weigle, 1998, p. 276). Another way to determine what degree of fit is acceptable is calculated as the mean plus/minus twice the standard deviation, $M \pm (2 \times SD)$, which the present study employed.

Figure 13 summarizes the FACETS analysis of the present study in a variable map. From left to right, this map, also called a vertical ruler, graphically represents how student ability varied, how severely or leniently raters behaved, and how difficult each category was. The first column in the map indicates the measurement scales in the logit scale (range: -2 to 2 logits) as a common reference for all the facets.

**Examinee measurement report.** The second column shows estimates of the 84 students’ writing ability. Higher-scoring students appear at the top, and lower-scoring students appear at the bottom. Each asterisk represents two students while each dot represents one student. The students are spread out with logit values from 1.53 to -1.70 ($M = .99, SD = .72$), showing a narrow total range of 3.23 logits. On the one hand, considering that all the students were in the same level of writing course (i.e., ELI 73), this relatively narrow range was anticipated. On the other hand, the range also indicates that although the students were placed into the same level
course based on their placement test results, their writing proficiency still varied somewhat.

According to the FACETS student separation statistics, the separation index was 3.68, which means that this sample of students could be separated into 3 statistically distinct levels of writing ability. The reliability of this index was .93, analogous to Cronbach’s alpha, which suggests that students were reliably separated according to their writing ability. Based on this significant finding, the null hypothesis that all the students’ writing abilities were equal is rejected. Based on the fit criteria (0.99 ± [.72 × 2 = 1.44]), five students were misfitting, that is, slightly over the upper-control limit.

\[ \text{Figure 13. Variable Map from the FACETS Analysis} \]
**Rater measurement report.** The third column illustrates the degrees of severity, or leniency, of the four raters in their assessments of the students’ essays. More severe raters appear higher in the column while more lenient raters appear lower. That is, raters 1 and 3 were slightly more lenient than raters 2 and 4. Table 7 shows logit values, errors, and fit statistics for the four raters. All the raters turned out to be relatively lenient with a slight variation among them, ranging from -21 to -76 logits; the difference between the most severe and the least severe rater is small at .55 logits ($M = -48, SD = .26$). Infit statistics show that all the raters were stable in terms of their severity, meaning that they were self-consistent across the essays they scored. Fit values for all the raters were within the range of two standard deviations around the mean, ranging from .59 to 1.46 ($1.02 \pm [.22 \times 2 = .44]$). This indicates that no rater was misfitting.

Table 7

**Rater Measurement Report for Four Raters**

<table>
<thead>
<tr>
<th>Rater</th>
<th>Severity (logits)</th>
<th>Error</th>
<th>Infit Mean Square</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rater1</td>
<td>-0.76</td>
<td>0.03</td>
<td>0.75</td>
</tr>
<tr>
<td>Rater2</td>
<td>-0.30</td>
<td>0.03</td>
<td>1.18</td>
</tr>
<tr>
<td>Rater3</td>
<td>-0.64</td>
<td>0.03</td>
<td>0.94</td>
</tr>
<tr>
<td>Rater4</td>
<td>-0.21</td>
<td>0.03</td>
<td>1.23</td>
</tr>
<tr>
<td>$M$</td>
<td>-0.48</td>
<td>0.03</td>
<td>1.02</td>
</tr>
<tr>
<td>$SD$</td>
<td>0.26</td>
<td>0.00</td>
<td>0.22</td>
</tr>
</tbody>
</table>

*Note. Reliability of separation index = .98; fixed (all same) chi-square: 195.2, df: 2; significance: $p < .00$*

The reliability of separation index was high (.98), and the chi-square of 195.2 with 3 df was significant at $p < .00$. The null hypothesis that all raters showed equal degrees of severity is therefore rejected. The ideal situation would be that the raters demonstrated the same degree of severity with a reliability close to 0.00 (Weigle, 1998; Winke et al., 2012). Even though there exists variance in terms of harshness (or leniency) among the raters in the present study, the distribution of rater severity is much narrower (range = 0.55 logits) than the distribution of
student ability (range = 3.23 logits), which indicates that individual differences in rater severity had a relatively small effect on students’ writing scores.

The fourth column represents the student performance on the pretest and the posttest, which indicates that it was more difficult for the students to gain high scores in the pretest than in the posttest. The fifth column shows that the test was more difficult for the control group than the treatment group, which concurs with the control groups’ lower scores in the holistic and analytic ratings.

**Rating criteria measurement report.** The sixth column compares the relative difficulty of the five categories of the analytic rating. As with rater severity, the most difficult category appears at the top and the least difficult at the bottom. It was more difficult for students to receive high scores in the category appearing highest in the column (i.e., language use) than the category appearing lowest in the column (i.e., mechanics). In other words, *language use* was the most harshly scored category, *mechanics* the most leniently scored.

Table 8 provides more details about the measurement report for the five categories. The difficulty logit scores show a total spread of 2.47 logits, from the most difficult category, language use, to the least difficult category, mechanics. As shown in the variable map, language use and mechanics are located approximately 1 logit score apart from the mean (0), whereas content, organization, and vocabulary are clustered together slightly higher than the mean. The reliability of separation index was very high (1.00), and the chi-square of 847.8 with 4 df was significant at $p < .00$. This indicates that not all categories are equally difficult, being distinct from each other. In addition, the measurement errors were very low across the five categories, which indicates that each rater was quite accurate in rating essays. No category turned out to be
misfitting; fit values for all categories were within the acceptable range of two standard
deviations around the mean, ranging from .87 to 1.11 (0.99 ± [.06 × 2 = .12]).

Table 8

<table>
<thead>
<tr>
<th>Category</th>
<th>Difficulty (logits)</th>
<th>Error</th>
<th>Infit Mean Square</th>
</tr>
</thead>
<tbody>
<tr>
<td>Content</td>
<td>0.18</td>
<td>0.03</td>
<td>1.04</td>
</tr>
<tr>
<td>Organization</td>
<td>0.25</td>
<td>0.03</td>
<td>1.08</td>
</tr>
<tr>
<td>Vocabulary</td>
<td>0.08</td>
<td>0.04</td>
<td>0.93</td>
</tr>
<tr>
<td>Language Use</td>
<td>0.98</td>
<td>0.04</td>
<td>0.97</td>
</tr>
<tr>
<td>Mechanics</td>
<td>-1.49</td>
<td>0.09</td>
<td>0.93</td>
</tr>
<tr>
<td>M</td>
<td>0.00</td>
<td>0.04</td>
<td>0.99</td>
</tr>
<tr>
<td>SD</td>
<td>0.90</td>
<td>0.02</td>
<td>0.06</td>
</tr>
</tbody>
</table>

*Note.* Reliability of separation index = 1.00; fixed (all same) chi-square: 848.7. *df:* 4, significant: *p* < .00.

The seventh column in the variable map displays the rating scale that the raters used in
scoring student essays. This column informs us, for example, that students with an ability
measure of 1 logit were likely to receive scores of 16 for content and organization, 15 for
vocabulary, 17 for language use, and 4 for mechanics.

Taking all these elements of the FACETs analysis together, it can be concluded that all
four raters showed consistent performance in scoring the essays, as evidenced by the stable infit
scores and confirmed by their almost negligible measurement error values. This indicates that the
raters were able to distinguish individual students’ writing performance. However, the variation
among the raters may indicate that they exercised slightly different degrees of severity, which
could have affected students’ analytic scores. More detailed rater training along with the norming
session could have minimized differences among the raters.
Writing Performance: CAF Analysis

This section reports the study’s findings on student writing performance based on the complexity, accuracy, and fluency (CAF) analyses. Along with the descriptive statistics, boxplots visually highlight the differences between the treatment group and the control group with regard to the several measures employed for exploring CAF changes in the students’ writing.

Complexity. Complexity was examined by means of the ratio of clause to T-unit (C/T) and the ratio of clause to sentence (C/S). As seen in Table 9, the control group has slightly higher means on both measures on the pretest, which may suggest that the control group students were able to produce more complex sentence structures than the treatment group at the beginning of the semester. With regard to C/T, the mean of the treatment group showed only a small increase, from $M = 1.52$ to 1.59, and the control group’s mean slightly decreased, from $M = 1.58$ to 1.53. In terms of C/S, while the treatment group’s mean increased from 1.66 to 1.78, the control group’s remained the same at 1.74.

Table 9

Descriptive Statistics for Complexity Measures

<table>
<thead>
<tr>
<th></th>
<th>Treatment ($n = 44$)</th>
<th>Control ($n = 40$)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pre</td>
<td>Post</td>
</tr>
<tr>
<td>C/T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M</td>
<td>1.52</td>
<td>1.59</td>
</tr>
<tr>
<td>SD</td>
<td>0.23</td>
<td>0.25</td>
</tr>
<tr>
<td>95% CI lower bound</td>
<td>1.45</td>
<td>1.51</td>
</tr>
<tr>
<td>95% CI upper bound</td>
<td>1.59</td>
<td>1.66</td>
</tr>
<tr>
<td>C/S</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M</td>
<td>1.66</td>
<td>1.78</td>
</tr>
<tr>
<td>SD</td>
<td>0.28</td>
<td>0.25</td>
</tr>
<tr>
<td>95% CI lower bound</td>
<td>1.57</td>
<td>1.70</td>
</tr>
<tr>
<td>95% CI upper bound</td>
<td>1.74</td>
<td>1.85</td>
</tr>
</tbody>
</table>

Note. C/T = clauses per t-unit; C/S = clauses per sentence

Reflecting the descriptive statistics results, all the boxplots in Figure 14 largely overlap with each other with only a minimal difference between the medians in the pretest and in the
posttest; the treatment group changed from a median of 1.50 to 1.51, and the control group changed from a median of 1.54 to 1.53. One finding worth pointing out is that while both groups’ medians remained almost the same across the tests, the IQR apparently became narrower (.34 vs. .23) on the posttest in the control group, which indicates less data variability around the median. However, several outliers are observed.

Figure 14. Boxplots of C/T for Complexity

Similarly, the boxplots for C/S overlap with each other (see Figure 15), suggesting little difference across the tests. On the pretest, the treatment group data have a close to normal distribution with a similar length of whiskers above and below the box, while the control group clearly demonstrates a positive skewing with one outlier above the upper whisker. The treatment group’s median (1.64) is lower than that of the control group (1.72). However, on the posttest, the median of the treatment group (1.66) was higher than that of the control group (1.65).
Moreover, it is interesting to note that the mean value is the same across the tests ($M = 1.74$) in the control group, but the median value, as illustrated in Figure 15, decreased from 1.72 to 1.65. This may indicate that the value ($max = 3.63$) far above the whisker could have inflated the mean on the posttest.

![Boxplots of C/S for Complexity](image)

**Figure 15.** Boxplots of C/S for Complexity

**Accuracy.** Three measures were used to assess accuracy in the student essays: error-free T-units per T-unit (EFT/T), error-free clauses per T-unit (EFC/T), and error-free clauses per sentence (EFC/S). As shown in Table 10, the treatment group exhibited a slightly higher accuracy level on the pretest, and the gap between the two groups became greater on the posttest. While the treatment group made progress across the three measures, the control group’s accuracy either remained the same (i.e., EFT/T) or demonstrated a very small improvement (i.e., EFC/T and EFC/S).
Table 10

**Descriptive Statistics for Accuracy Measures**

<table>
<thead>
<tr>
<th></th>
<th>Treatment (n = 44)</th>
<th>Control (n = 40)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pretest</td>
<td>Posttest</td>
</tr>
<tr>
<td><strong>EFT/T</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M</td>
<td>0.39</td>
<td>0.50</td>
</tr>
<tr>
<td>SD</td>
<td>0.19</td>
<td>0.18</td>
</tr>
<tr>
<td>95% CI lower bound</td>
<td>0.34</td>
<td>0.45</td>
</tr>
<tr>
<td>95% CI upper bound</td>
<td>0.45</td>
<td>0.56</td>
</tr>
<tr>
<td><strong>EFC/T</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M</td>
<td>0.76</td>
<td>0.94</td>
</tr>
<tr>
<td>SD</td>
<td>0.33</td>
<td>0.34</td>
</tr>
<tr>
<td>95% CI lower bound</td>
<td>0.67</td>
<td>0.84</td>
</tr>
<tr>
<td>95% CI upper bound</td>
<td>0.86</td>
<td>1.05</td>
</tr>
<tr>
<td><strong>EFC/S</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M</td>
<td>0.84</td>
<td>1.05</td>
</tr>
<tr>
<td>SD</td>
<td>0.36</td>
<td>0.36</td>
</tr>
<tr>
<td>95% CI lower bound</td>
<td>0.73</td>
<td>0.94</td>
</tr>
<tr>
<td>95% CI upper bound</td>
<td>0.95</td>
<td>1.16</td>
</tr>
</tbody>
</table>

*Note. EFT/T = error-free T-units per T-unit; EFC/T = error-free clauses per T-unit; EFC/S = error-free clauses per sentence.*

In Figure 16, the boxplots for the pretest overlap with each other’s medians (TR: 0.38; CT: 0.31), both being skewed to the right. On the posttest, the treatment group shows a normal distribution with the median (0.51) cutting the box into two pieces almost equally. The control group has a lower median (0.33) than the treatment group, but the data cluster more around the median with a narrower IQR (.13) than in the treatment group (.21). The data are skewed to the right due to one outlier above the upper whisker.
Another finding worth noting is that the clause-based measures, that is, EFC/T and EFC/S, showed greater gains of accuracy than the T-unit-based measure, EFT/T. Aligning with the descriptive statistics, the information presented in Figure 17 and Figure 18 clearly shows that not only did the treatment group produce a higher ratio of error-free clauses than the control group across the two tests, but their gains were bigger as well (EFC/T = 0.18; EFC/S = 0.21) when compared to their gains in EFT/T (0.11). In addition, overall, there seem to be more variability in the data sets for EFC/T and EFC/S than for EFT/T, as evidenced by longer boxes and whiskers.
Figure 17. Boxplots of EFC/T for Accuracy

Figure 18. Boxplots of EFC/S for Accuracy
Fluency. Fluency was assessed by employing the following four measures: W/T (words per T-unit), W/C (words per clause), W/S (words per sentence), and words per minute (WPM). As displayed in Table 11, the control group had higher means for W/T, W/C, and W/S than the treatment group on the pretest and the posttest.

Table 11

<table>
<thead>
<tr>
<th></th>
<th>Treatment Group</th>
<th>Control Group</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(n = 44)</td>
<td>(n = 40)</td>
</tr>
<tr>
<td></td>
<td>Pre</td>
<td>Post</td>
</tr>
<tr>
<td>W/T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M</td>
<td>13.08</td>
<td>14.57</td>
</tr>
<tr>
<td>SD</td>
<td>2.62</td>
<td>2.35</td>
</tr>
<tr>
<td>95% CI lower bound</td>
<td>12.28</td>
<td>13.85</td>
</tr>
<tr>
<td>95% CI upper bound</td>
<td>13.88</td>
<td>15.28</td>
</tr>
<tr>
<td>W/C</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M</td>
<td>8.71</td>
<td>9.24</td>
</tr>
<tr>
<td>SD</td>
<td>1.81</td>
<td>1.24</td>
</tr>
<tr>
<td>95% CI lower bound</td>
<td>8.16</td>
<td>8.87</td>
</tr>
<tr>
<td>95% CI upper bound</td>
<td>9.26</td>
<td>9.63</td>
</tr>
<tr>
<td>W/S</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M</td>
<td>14.28</td>
<td>16.32</td>
</tr>
<tr>
<td>SD</td>
<td>3.18</td>
<td>2.46</td>
</tr>
<tr>
<td>95% CI lower bound</td>
<td>13.32</td>
<td>15.57</td>
</tr>
<tr>
<td>95% CI upper bound</td>
<td>15.25</td>
<td>17.06</td>
</tr>
<tr>
<td>WPM</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M</td>
<td>4.68</td>
<td>6.65</td>
</tr>
<tr>
<td>SD</td>
<td>1.53</td>
<td>1.60</td>
</tr>
<tr>
<td>95% CI lower bound</td>
<td>4.21</td>
<td>6.15</td>
</tr>
<tr>
<td>95% CI upper bound</td>
<td>5.15</td>
<td>7.14</td>
</tr>
<tr>
<td>Word Total</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M</td>
<td>280.91</td>
<td>398.73</td>
</tr>
<tr>
<td>SD</td>
<td>92.85</td>
<td>97.37</td>
</tr>
<tr>
<td>95% CI lower bound</td>
<td>252.68</td>
<td>369.12</td>
</tr>
<tr>
<td>95% CI upper bound</td>
<td>309.14</td>
<td>428.33</td>
</tr>
</tbody>
</table>

Note. W/T = the number of words per T-unit; W/C = the number of words per clause; W/S = the number of words per sentence; WPM = words per minute
However, the treatment group gained more than the control group in W/T (TR: 1.49; CT: 1.26) and W/S (TR: 2.04; CT: 1.78). In Figure 19, although several outliers are observed, the boxes and whiskers seem symmetrical around the medians, and the length of the IQRs also look quite similar across the two tests in both groups. From the pretest to the posttest, the medians increased from 12.62 to 14.14 in the treatment group and from 14.31 to 15.36 in the control group.

**Figure 19.** Boxplots of W/T for Fluency

The gain for W/C of the treatment group (0.53) is smaller than that of the control group (1.47). Looking at the data distribution of W/C of the treatment group, several outliers are seen above the upper whiskers on the pretest as well as on the posttest, which could have inflated the mean score on the pretest (see Figure 20). In addition, the data set, except these outliers, appears to cluster around the medians in the treatment group (pretest: 8.46 and posttest: 9.20), while the
control group’s data set displays wider IQRs around the medians for both the pretest (9.16) and the posttest (9.99). Regarding W/S, Figure 21 shows that the differences of the medians between the two groups on the pretest (TR: 13.88; CT: 15.80) became greater on the posttest (TR: 15.33; CT: 17.71). As with W/C, the control group exhibits wider IQRs than the treatment group for both the pretest and the posttest, indicating greater data variability.

Figure 20. Boxplots of W/C for Fluency
Figure 21. Boxplots of W/S for Fluency

Contradictory findings were observed for WPM. Comparing the total number of words per essay on the posttest, the treatment group wrote on average 398.73 words and the control group, 324.25 words: a difference of 74.48 words. As the test was 60 minutes long, this difference indicates that the treatment group wrote on average 1.25 words more than the control group per minute. In Figure 22, it is clearly shown that, for the treatment group, the pretest boxplot is positively skewed with one outlier whereas the posttest boxplot is normally distributed. There is no overlap of the boxes between the pre and posttests, with a large gap between the medians (4.56 vs. 6.54), indicating that the treatment group improved a lot on the posttest. In contrast, the control group boxplots overlap each other across the two tests with a smaller difference of medians (4.15 vs. 5.16).
Figure 22. Boxplots of WPM for Fluency

Taking all these fluency measures together, the two different trends suggest that the treatment group produced more t-units, clauses, and sentences in total during the given time, but the lengths of these units were shorter than those of the control group.

Findings from the Extensive Reading Survey

Before the extensive reading survey was analyzed, the amount students read was examined by referring to their book recording forms. The first and last weeks were excluded; the treatment group did not engage in ER due to course instruction, pre- and post-test administration, and course evaluations. The arrangement was the same for the control group. Therefore, students’ reading amount was calculated for 14 weeks.

As seen in Table 12, the treatment group students spent a total of 29.84 hours for ER for 14 weeks, which is approximately 2.13 hours per week. This is the sum of 7 hours of in-class
reading (30 minutes x 14 weeks = 420 minutes) and 22.84 hours (1370.36 minutes) of out-of-
class reading. Whereas all the students regularly joined in the in-class ER during every class, the
amount that individual students read outside of class varied, from a minimum of 0.75 hours to a
maximum of 2.5 hours per week. The control group students spent about 1.5 hours per week for
the course assignments.

Table 12

<table>
<thead>
<tr>
<th>Amount of Time Spent and Number of Books Read for ER</th>
<th>Per Week</th>
<th>14 Weeks</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total time spent (hours)</td>
<td>2.13</td>
<td>29.84</td>
</tr>
<tr>
<td>In-class</td>
<td>0.50</td>
<td>7.00</td>
</tr>
<tr>
<td>Out-of-class</td>
<td>1.63</td>
<td>22.84</td>
</tr>
<tr>
<td>Number of books</td>
<td>1</td>
<td>14</td>
</tr>
<tr>
<td>Number of pages</td>
<td>73.87</td>
<td>1108</td>
</tr>
</tbody>
</table>

Each student read, on average, 14 books during the treatment period, or roughly one book
per week. More specifically, they read 73.87 pages per week, which amounts to 1108 pages in
total. Because the books the students read were diverse in terms of level (e.g., ranging from 2 to
6) and genres (e.g., graded readers vs. young readers), the page numbers were additionally
counted to provide a more accurate picture of their reading amount. Nevertheless, it should be
kept in mind that the page count is still not a precise amount as higher level books contain more
words per page. Students who read more books of higher levels would end up reading more
words than students who read a similar number of pages of lower level books.

Table 13 displays means and standard deviations of the Extensive Reading survey. The
column on the far right shows the percentage of students who either agreed (4) or strongly agreed
(5) on each item. The survey, consisting of 19 items in total, showed a high reliability according
to Cronbach’s alpha (α = .898). The discussions of the survey findings fall into two categories:
general perceptions of ER and the perceived usefulness of ER.
**Attitude toward ER.** Students’ general attitudes toward ER are considered in terms of the degree of enjoyment, comfortableness, and confidence they expressed. First, on the items (1, 2, and 8) asking about their enjoyment of reading, most students responded favorably overall. For example, about 93% of the students responded that they enjoyed reading in the classroom (Item 1), and this was ranked the highest among all the items, with a mean value of 4.41. However, the out-of-class ER experience was apparently relatively less pleasurable for students (Item 2, $M = 3.89, 68.18\%$). Another finding worth noting is that, thanks to ER, 84.09% of the students claimed to enjoy reading books more than before (Item 8, $M = 4.18$). Item 12 inquired into the students’ desire to continue ER, and 77.27% ($M = 4.02$) of the students showed willingness to carry on by themselves beyond this course.

Regarding the items (5 and 6) about the comfortableness, or easiness, of reading books, most students answered that they felt comfortable when reading (Item 6, $M = 4.23, 86.36\%$) and that understanding stories became easier as the semester progressed (Item 5, $M = 4.30, 88.64\%$). In addition, 88.64% of the students responded that they had gained more confidence about reading English books (Item 19, $M = 4.21$).

As opposed to their positive attitudes toward reading in general, students found the 10-minute writing less enjoyable (Item 7, $M = 3.73$). Only slightly more than half of the students (59.09%) responded positively; 40.91% students neither agreed nor disagreed. No one expressed disagreement, however.
Table 13

**Descriptive Statistics of the Extensive Reading Survey**

<table>
<thead>
<tr>
<th>Attitude toward ER</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Enjoyment</strong></td>
<td><strong>M</strong></td>
<td><strong>SD</strong></td>
<td><strong>%</strong></td>
</tr>
<tr>
<td>1. I enjoyed reading books in class.</td>
<td>4.41</td>
<td>0.62</td>
<td>93.18</td>
</tr>
<tr>
<td>2. I enjoyed reading books at home.</td>
<td>3.89</td>
<td>0.89</td>
<td>68.18</td>
</tr>
<tr>
<td>7. I enjoyed writing about the stories that I read.</td>
<td>3.73</td>
<td>0.73</td>
<td>59.09</td>
</tr>
<tr>
<td>8. I think I enjoy reading books now more than before.</td>
<td>4.18</td>
<td>0.76</td>
<td>84.09</td>
</tr>
<tr>
<td>12. I think I will keep reading English books by myself after this course.</td>
<td>4.02</td>
<td>0.93</td>
<td>77.27</td>
</tr>
<tr>
<td><strong>Comfortableness</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. I think understanding stories got easier as the semester progressed.</td>
<td>4.30</td>
<td>0.67</td>
<td>88.64</td>
</tr>
<tr>
<td>6. I felt comfortable while reading books.</td>
<td>4.23</td>
<td>0.68</td>
<td>86.36</td>
</tr>
<tr>
<td><strong>Confidence</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10. I think reading books helped me feel more confident about writing.</td>
<td>3.82</td>
<td>0.79</td>
<td>68.18</td>
</tr>
<tr>
<td>19. I have more confidence now about reading English books.</td>
<td>4.21</td>
<td>0.63</td>
<td>88.64</td>
</tr>
<tr>
<td><strong>Usefulness of ER</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Reading</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. I think reading books helped me improve my reading comprehension ability.</td>
<td>4.30</td>
<td>0.76</td>
<td>86.36</td>
</tr>
<tr>
<td>13. I think reading books was useful for improving my reading speed.</td>
<td>4.30</td>
<td>0.70</td>
<td>90.91</td>
</tr>
<tr>
<td><strong>Writing</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. I think reading books was useful for me to improve my overall written language.</td>
<td>4.07</td>
<td>0.85</td>
<td>77.27</td>
</tr>
<tr>
<td>9. I think writing activities were helpful for me to understand the stories better.</td>
<td>3.86</td>
<td>0.80</td>
<td>70.45</td>
</tr>
<tr>
<td>14. I think reading books was useful to learn sentence structures.</td>
<td>4.02</td>
<td>0.76</td>
<td>75.00</td>
</tr>
<tr>
<td>15. I think reading books helped me increase my grammar knowledge.</td>
<td>3.73</td>
<td>0.92</td>
<td>59.09</td>
</tr>
<tr>
<td>16. I think I learned some useful expressions while reading.</td>
<td>4.07</td>
<td>0.66</td>
<td>86.36</td>
</tr>
<tr>
<td><strong>Vocabulary</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11. I was able to guess the meaning of unknown words using the context in the story.</td>
<td>4.18</td>
<td>0.62</td>
<td>88.64</td>
</tr>
<tr>
<td><strong>Speaking &amp; Listening</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17. I think reading books helped me improve my English speaking ability.</td>
<td>3.43</td>
<td>0.90</td>
<td>43.18</td>
</tr>
<tr>
<td>18. I think reading books helped me improve my listening ability.</td>
<td>3.07</td>
<td>1.00</td>
<td>29.55</td>
</tr>
</tbody>
</table>

*Note. N = 44. The rightmost column gives the percentages of students who responded “agree” or “strongly agree.”*
Usefulness of ER. The survey revealed different degrees of perceived usefulness of ER in different areas of English skills. In line with their positive attitudes toward reading discussed above, the students’ ratings of the reading-related items showed the highest means. For example, many students agreed that reading books was particularly useful for improving their comprehension (Item 3, 86.36%) and reading speed (Item 13, 90.91%); these items’ ratings had the same mean value of 4.30. More than three quarters of the students (77.27%) perceived ER to be helpful for improving their overall written language (Item 4, $M = 4.07$), and specifically for learning sentence structures (Item 14, $M = 4.02$, 75%) and new expressions (Item 16, $M = 4.07$, 86.36%). However, students seemed less certain about whether ER helped them to learn English grammatical knowledge (Item 15, $M = 3.73$). On this item, only 59.09% students responded favorably; 31.82% students felt unsure and 9.09% students disagreed. The students also displayed uncertainty about the effectiveness of ER on speaking and listening; these items received much lower mean scores (Item 17, $M = 3.43$ and Item 18, $M = 3.07$) compared to the areas of reading and writing.

Open-ended questions. The survey included five open-ended questions in order to (a) seek the students’ detailed thoughts and ideas and (b) elicit useful information for developing interview questions. The five questions asked about students’ likes and dislikes in regard to ER, the integration of ER and the 10-minute writing, and the usefulness of ER in their language learning. Students were also encouraged to provide any suggestions they might have about ER for future ELI 73 students. Table 14 provides a summary of student responses on the five questions.
Table 14

Summary of Responses to Open-ended Questions on ER and the 10-minute Writing

Student responses

| Students liked | • Reading easy materials  
|               | • That the books were easy to carry (e.g., small & light)  
|               | • Having freedom to choose what to read  
|               | • Having more opportunities to read English books  
|               | • Feeling relaxed with no pressure  
|               | • Learning about new cultures and history  
|               | • Short length of the books  
|               | • Talking about books with classmates  
|               | • Having a variety of book selections and the easiness of checking out books  
| Students disliked | • Certain topics for the 10-minute writing activity  
|                  | • Filling in the Book Recording Form  
|                  | • Keeping track of their reading (e.g., date & time)  
|                  | • Lack of interesting books in the class library  
| Students found ER useful in improving or learning | • Reading  
|                  |   o Speed  
|                  |   o Comprehension (understanding)  
|                  |   o Skills to focus on content more than words  
|                  | • Writing  
|                  |   o Organization  
|                  |   o Summarizing (writing in chronological order)  
|                  |   o Useful expressions and idioms  
|                  |   o Sentence structures  
|                  |   o Grammar  
|                  | • Vocabulary  
|                  |   o New words  
|                  |   o Skills to deal with unknown words  
|                  |   o Spellings  
|                  | • Speaking  
|                  |   o Useful expressions or idioms for conversation  
|                  | • Listening  
|                  |   o Listening ability  
| Thanks to ER, students | • Improved their attitude toward reading in English  
|                       | • Feel more at ease about reading and writing in English  
|                       | • Feel more motivated to read and write in English (e.g., writing personal journals in English)  
| Students suggested | • Including more interesting books in the class library  
|                     | • Reducing the amount of the 10-minute writing (e.g., once a week)  

110
The responses to the open-ended questions were quite analogous to the survey results, but also included some comments of additional interest. For example, one of the reasons students said they liked ER was that it actually gave them more opportunity to read in English. Interestingly, the attractive display of books at the front of the classroom and the easiness of checking out or exchanging books were also motivating factors for students.

Although the items about speaking and reading had low mean scores on the survey, many students commented that ER might be useful for improving speaking and listening skills because reading could improve their knowledge of expressions and idioms. Some noted that learning new expressions from their reading indeed had helped them communicate better in English. In addition to these positive remarks, some students expressed negative feelings about the book recording form due to the difficulty of keeping track of their reading. They tended to forget about recording their reading dates and times. Also, while most students seemed to feel that the class library had a large number of interesting books, some students wished for a greater variety of books to meet specific topical interests.

Findings from Student Interviews

The purpose of the interviews was to investigate students’ reactions and opinions in depth, while supplementing the extensive reading survey findings. All the interviews were carried out in a library after the semester was over. Each interview lasted from 30 minutes to 60 minutes. Thirteen students from the treatment group agreed to voluntarily participate in the interview. Five students from the control group were also recruited to gain some idea of their general reading and writing habits as well as the amount they read during the semester.
This section reports the treatment group interviewees’ voices in detail with excerpts taken from the transcripts of individual interviews. The interview findings are closely connected to the survey findings, and they are discussed according to four themes that emerged from both the surveys and the interviews: (a) students’ attitudes toward ER, (b) perceived usefulness of ER in language learning, (c) perceived usefulness of integrating ER and writing activities, and (d) negative remarks on ER and the 10-minute writing. The student responses were first transcribed exactly. For the excerpts that appear in this dissertation, however, some grammatical mistakes have been corrected in order to deliver the interviewees’ opinions clearly and without ambiguity. In making such corrections, care was taken to avoid distorting the meaning.

Students’ perceptions about ER. The analysis of students’ attitudes toward ER generated the following three subthemes: enjoyment, comfortableness, and confidence.

Enjoyment. The majority of students responded that ER was an enjoyable experience for them. This pleasurable component seemed particularly related to the freedom to self-select reading materials, the content of the books, and the short length of the books. Even those who claimed to hate reading in the past commented that they started to enjoy it. Students’ remarks included:

“I liked it [ER] because I can read what I want to read. I particularly liked crime and horror stories, so I mostly read these books throughout the semester.”

“I really didn’t like reading, even in Japanese. I just hated reading. I was required to read textbooks this semester in other classes and I hated it. I just couldn’t concentrate on reading…In ELI 73, I started reading biography, like Princess Diana. They were okay and interesting. And then I read Wonder. It was really interesting. I really liked it. It just changed my mind about reading.”

“If I hadn’t read these short books, I would have hated reading in English. Reading books motivated me to enjoy English, and made me keep interested in English.”
“The only bad thing about ER is when you find a super interesting book, you are not able to stop. I had a few books that I read from cover to cover in one sitting. When I started, I couldn’t stop.”

It is interesting that to some students the length of the book was an important motivating factor, while to others it did not matter as long as they liked the story. Two contradictory remarks were:

“I liked it [ER] because I could finish one story quickly and move onto other stories.”

“It’s not the size of the book. It’s about the content. If I like the story, I don’t care about the length.”

One student mentioned that he did not choose to read longer stories (e.g., young readers), not because of the length but because of his reading habits. He said:

“I didn’t want to pick a thick book because once I start read something, I really don’t want to end it in the middle. I need to know the ending regardless of how much I like the story.”

**Comfortableness.** Another aspect that students mentioned many times during the interviews was their comfortable feeling while reading, which was closely associated with the easy reading materials provided in the classroom library. The easiness of the reading seemed to alleviate their fear of reading in English, while continuously facilitating their reading. Their comments included the following:

“It was enjoyable because it was not stressful reading. I could enjoy the stories because it was easy to understand.”

“The books I chose were pretty easy to read. It was not stressful at all. When I read academic textbooks, it’s always stressful. It takes a long time to just finish a couple of pages. And I feel frustrated because I can’t understand clearly after spending hours and hours.”

“I felt comfortable reading 20 minutes everyday. I usually read books before I go to bed.”

“I read many academic texts for other classes, so I liked ER because I could read novels like short stories.”
The interviews clearly uncovered that reading a book within their language ability relieved their stress about reading, which in turn served as one of the motivating factors for further reading. One student mentioned how frustrated he had been when reading a novel above his reading ability. He recalled:

“I kind of had a fear of starting reading. I read the book, *The Hobbit*, before and my reading was super slow. I had to find every other word in a dictionary to understand the story. It took me a lot of time, so I gave up. But now it’s like I can read a whole book comfortably without looking in dictionary although those books (short stories) are generally easier than *The Hobbit.*”

**Confidence.** The interview comments revealed that students felt a sense of achievement in the fact that they could finish English books. Most of the students read an English book from cover to cover for the first time in this class. Easily finishing one book increased their confidence level about their reading ability, which encouraged their subsequent reading. In addition, being able to handle a higher level of books as the semester progressed served as a confidence booster. Positive remarks included the following:

“I never tried to read English books before because I thought reading is difficult and boring. But the books in the class library were not difficult to understand. I could finish many books this semester.”

“I felt good when I finished the thick book, *Wonder*. I never thought I would finish this book when I first started it. But actually I forgot about how long this book was once I really enjoyed the story. And I was surprised that you told me that this book is higher level than the short stories. I am proud of myself.”

“I think level 5 was comfortable for me. Level 5 has enough details, and I can still understand and enjoy the story. Once I tried level 5, I didn’t want to go back to easier ones because I know I can understand and enjoy it.”

It is worth noting that several students showed their desire to continue reading even after the semester was over. Their raised confidence as well as their interest in particular stories led them to purchase English books for their own ER.
“I bought Coco Chanel biography. I read *Women in Business* this semester and one of the women in the book was Coco Chanel. So I wanted to know more about her life. I am going to read this during the winter break.”

“It’s my first time to go to the book store here. It was kind of fun to see many books. I bought one novel and it was my first time to buy an English book, not for my major…I don’t have time now but I think I can read it later during the break.”

“I read *Mind Reader 1* this semester, and it was really interesting although some words were difficult. I wanted to read the second one, but the class library didn’t have it. So I bought *Mind Reader 2* and I’ve just started reading it.”

**The perceived usefulness of ER in language learning.** Individual students indicated their perception that some areas of language learning had benefitted from engaging in ER for one semester. Their interview comments are in line with the survey findings.

*Reading rate.* As students’ reading amount increased as the semester progressed, their reading rate seems to have increased naturally. The interview also revealed that improved reading skills, such as getting used to focusing more on general ideas rather than details or words, played a role in increasing reading rate. Some comments were:

“I think reading books got faster as I read more books…It’s a story, so I can understand the content easily, and that’s why my reading speed got faster.”

“I think it’s [reading rate] growing every day, every time I read. I guess I wasn’t that fast when I started. I am used to it [reading] now.”

“I think my speed has been improved. Even though the books were easy, I read kind of slowly at the beginning of the semester. I didn’t want to miss any word and I tried to understand everything. As you said, I tried to understand the general idea about the story and tried not to focus on specific details. And I think it helped improve my reading speed.”

*Writing speed and writing anxiety.* Several interesting points emerged with regard to writing. Most of the remarks were related to the nature of the 10-minute writing, which was stress-free and fast writing. Some comments included the following:
“It [the 10-minute writing] is a good practice. So I think it’s good to maintain it. I had a problem with my TOEFL test because I couldn’t write too much. If you write and write and write, you can learn how to do it well.”

“I think I can write faster now because we practiced writing every week. I only could write half a page but now I can write almost a full page within the same time. I didn’t think too much. I just wrote.”

“I tried not to care too much what I was writing but just write. I tried to improve. I set the timer on for 10 minutes. I tried to write more and more. And I got faster.”

Writing regularly through the 10-minute writing seemed to have also contributed to decreasing their nervousness about writing in English. A few students noted that writing about a familiar topic, that is, their books, lessened their writing anxiety. Two students mentioned,

“I like that activity. If it were academic writing, I think it would be really boring. The good thing is when I did that writing, I didn’t feel nervous because that was for fun. I didn’t have any stress about writing because I knew what I needed to write. I didn’t feel like I had to write well. My hesitation to write is gone.”

“I think I felt comfortable when doing the 10-minute writing because I didn’t have to focus on grammar or organization. I usually feel very nervous when I have to write a paper and I can’t write much. I think I can write more than before. And I feel a little bit more confident about writing in English.”

*Expressions and sentence structures.* Students chose learning new or useful expressions as well as learning sentence structures as one of the benefits of reading books. It is intriguing that some students in fact wrote down new expressions for self-learning and further actually used them in their 10-minute writing. Several comments were,

“I liked it when I see new interesting expressions in the story. Sometimes I don’t know what they exactly meant, but I could usually understand their meaning based on the context. If I learn new expressions, I sometimes wrote them down [in a notebook] so I can use them later.”

“In the conversations between characters, I sometimes picked up expressions like idioms.”

“Reading was helpful to see how sentences were written. I can write simple sentences but I have difficulties writing complex sentences. I think books showed me how to write complex sentences.”
“When I found interesting expressions or sentence structures that I didn’t know, I tried to use them in my writing so that I can remember how to use it.”

**Vocabulary learning.** Another area that students brought up often was vocabulary learning. In addition, most students responded that they did not consult a dictionary while reading, as advised. They commented that they were able to deal with unknown words, either skipping or guessing the meaning from the context:

“I think I learned a lot of words because of reading in this class. The books I read were easy but there were some words that I still don’t know. I was mostly able to imagine the meaning of the words because the books usually have many descriptions and help my guessing.”

“For me, extensive reading was most useful for learning vocabulary. It was interesting to see how the same word can be used in a different situation.”

“I did not look up words in a dictionary. At first, it was hard because I wanted to know all the unknown words in the book. But I got used to not checking the meaning of every word that I don’t know. And I think I could still follow the story. Sometimes I could guess the meaning.”

“In *Walk Two Moons*, I saw some words that I hadn’t known, but in the context, they made sense. So I just kept reading.”

However, this skipping or guessing skill may not have been successful all the time, or at least not for some students. One student mentioned that she sometimes could not figure out the meaning of unknown words even after encountering them several times. Similarly, another student reflected that he usually guessed the meaning, but he tended to forget the word when it came up again in the story, which often frustrated him. Their comments were as follows:

“I tried to imagine the words based on the context. Usually I could guess the meaning, but the problem is I can’t remember the word. I guess one time is not enough [to learn the meaning of the word].”

“I tried to skip some words that I don’t know. They kept appearing in the story, but I still couldn’t understand it.”
Several students mentioned that they occasionally used a dictionary when they felt the word seemed important to understand the story or when they wanted to learn the exact meaning. They commented:

“\(\text{I used a dictionary sometimes when they looked important to follow the story.}\)”

“When I couldn’t figure out the meaning in the context, I looked in a dictionary to check the meaning. I think I did it often at the beginning of the semester, but I think I didn’t use a dictionary much later.”

“I often used a dictionary because I wasn’t sure whether my guess was right. I think if I check the meaning, and if I see the word again in the story, it was helpful for me to understand the story and I could remember better.”

**Learning about culture or history.** A few students commented that they liked ER because they could learn some cultural information or history through reading. Student comments include,

“I chose Martin Luther King Jr. for my first book [for ER]. I’ve heard his name so I just grabbed the book. It was short but I learned a lot about him, like he was a leader of the Civil Rights Movement or things that he did for the equality of African-American people. It was really interesting. …I searched him to know about his life more and I became interested in this topic. So after I finished that book, I read Nelson Mandela and then I read Cry for Freedom. It (Cry for freedom) was a novel, not a real story, but the background was similar and it was about human equality.”

“I read Seasons and Celebrations. The book was introducing many holidays of western countries, like Easter or Thanksgiving. I’ve heard most of them but some are new to me. The book was explaining what people do or eat on these holidays. It was interesting to learn about different holidays from my country.”

**The perceived usefulness of integrating reading and writing.** The following excerpts are comments from some of the interviewees who responded positively to the questions on the effectiveness of incorporating ER into L2 academic writing classes. They mostly recognized the usefulness of integrating the two skills of reading and writing, but most of their comments were about how reading helps writing, rather than vice versa. However, there seemed to be a general
consensus that the input they got from reading aided them to understand how to use words and expressions properly in a specific context, and using them in their own writing further helped them practice their gained knowledge.

“If you read a lot of different genres, your writing gets better... I think it’s kind of forced to get better, because if you never have read a single book, how could you possibly know how to write? So I think it was a good activity to combine the two.”

“My major is computer science. I don’t have a chance to read novels. I usually read dry books like academic stuff. I think if I really want to be a good writer, I should read a lot of books because I can learn how I can express my feelings or how to use language efficiently.”

“I like reading books because I can learn new words. English is not my native language, so I don’t know how to use words naturally. But if you read a book, you get input a lot. I can practice it in my writing. I think combining the two was good.”

“I think the writing exercise was good. When you are reading books and writing about them at the same time, you can apply them [words and expressions] in your own way to fit into your sentence. Sometimes it just happened, not maybe trying. ‘Oh this word is good for describing this situation.’”

One student gave a specific example of how his knowledge of using a certain transition word was enhanced by encountering the word many times during reading:

“I think when you study in a foreign language you have to find things people use. I guess I improved writing skills a lot, like contrast clauses, by reading books. We don’t use similar kinds of clauses in Finnish language. So ‘however’ was like I didn’t even know when to use it. I took one English class before coming here in Finland and that was probably the first time I heard people using ‘however.’ But you can see it [‘however’] a thousand times when you read and learn how to use it.”

Several students also commented that they liked this type of light assignment (i.e., ER and the 10-minute writing) because it was easier to do than other assignments, such as reading a textbook or academic articles, or writing a research paper. While heavy course assignments sometimes made them lose interest or postpone their assignment until the last minute, thus often depriving
them of the joy of reading in English, students perceived this assignment combination as enjoyable and feasible.

“I think if it [the class] were about just writing and writing, it would have been so tiring and too much. So I thought this class was balanced, and I liked it.”

“I liked the combination of reading and writing because they don’t take much time to do. It was manageable for me. I usually postpone my assignment if I don’t want to do it. And I got really busy one day before the deadline.”

“Compared to other assignments in my literature class, like reading difficult textbooks, I think the assignment here (i.e., ER + 10-minute writing) was easy. It didn’t take me much time.”

**Negative comments about ER and writing activity.** As reported above, most students expressed positive attitudes about their one-semester experience of ER and the 10-minute writing assignments, although some students made negative comments about its usefulness. Such negative remarks were mostly related to reading materials, writing activities, and time-related issues.

**Difficulty finding good reading materials.** Some students seemed to have trouble finding books that interested them and at the same time were within their reading level. They tended to lose interest or feel frustrated easily when failing to find appropriate books. Lack of enough books in the class library was identified many times as a factor that negatively influenced students’ motivation to read further. Some of the comments were as follows:

“I don’t know how to find good and interesting books. If I read a super difficult one, I don’t want to read. It’s not easy to find appropriate books for me.”

“The collection of book is pretty wide but I feel like some of the books I chose were boring. I guess the interesting ones were already taken. There are probably some books I’ve never seen.”

“I tried level two and three first, and I moved on to a bit thicker one. When I read *All I Want*, (graded reader) it was easy to understand. I just started to read some part of *Liar*
(young readers), but there were many idioms, which I didn’t understand. So I stopped reading it and went back to short stories.”

Although none of the students expressed resistance to ER being part of their writing course, two students doubted whether reading easy materials could benefit writing ability, especially academic writing:

“I liked ER and I think I learned some useful expressions or sentence structures from reading short stories. But I am not sure those readings can directly help me write academic papers. I think reading academic texts will be more useful for me to learn academic words.”

“I think I have more confidence about English writing, and I feel like my writing is improved, but I don’t know about academic writing skills…Like organization or quoting, I think teaching (lecture) was more helpful.”

**Writing activity.** Students generally did not feel the 10-minute writing to be a burden, partly because, as some of the interview excerpts above illustrated, they did not feel pressure to produce an accurate piece of writing. However, some students pointed out that the 10-minute writing was more tiresome than ER itself. Others mentioned that the suggested writing topics were sometimes unsuitable for their books, so they had to turn to summary writing. Their comments were:

“I like reading but sometimes I thought writing every week was boring. I sometimes had to write almost the same thing because I did not read much that week. One writing per week might be better.”

“I sometimes wished I didn’t have to write but read. I mean, I really enjoyed reading but sometimes I didn’t want to write. It’s kind of boring.”

“Writing activity was okay, not too much burden for me. But sometimes the topic was troublesome, and in that case I just wrote summaries.”

**Time issues.** As anticipated, one of the most common negative comments about ER was about the difficulty of finding time to read. In particular, those who did not enjoy reading as
much as their classmates usually said that it was difficult for them to maintain their reading practice. They felt that reading about 15 – 20 minutes per day was a burden for them as they were busy and had other priorities:

“I don’t have time to read. And I tend to lose focus when I am busy.”

“I think I would read more if I had time. I am too busy with other assignments. It’s hard to find time for reading.”

“Reading was fun but I tended to forget about reading because I was usually busy doing other homework. If I have time, I want to read more.”

**Summary of the Findings**

One of the most striking findings of the study is that the students who had the ER experience exhibited greater improvement on their posttest than the students with traditional instruction. Analytic rating further revealed the different degrees of effectiveness of ER on various aspects of writing, from content and organization to mechanics. In terms of CAF, although gains were less evident in both groups compared to the noticeable gains based on the holistic and analytic ratings, the two groups displayed a different trajectory across the two tests. That is, the treatment group students gained minimally in complexity, but improved more in accuracy and fluency. On the other hand, the control group students improved in fluency, but gained negligibly in accuracy with a slight decrease in complexity.

To reiterate the findings of the extensive reading survey, the treatment group students showed favorable reactions toward ER, largely due to interesting content and the comfortable reading experience. They also responded positively about the effectiveness of ER on their target language learning, including reading rate, vocabulary, expressions, and cultural knowledge. The findings from the interviews accorded with this result, yet they also revealed that the students
enjoyed ER itself more than the 10-minute writing. Nevertheless, the students acknowledged the advantages of engaging in both, especially in improving reading and writing abilities. It is intriguing that many students suggested that they were willing to continue reading by themselves beyond the semester; however, they inevitably had time-related concerns about the feasibility of continuing on their own.
CHAPTER V
DISCUSSION

This chapter provides an interpretation of the findings. It discusses each research question in detail, highlighting the key findings of the study and drawing connections to other relevant research and to pedagogic practice.

The Impact of ER on Overall Writing Performance

The first research question inquired into whether ER could positively affect L2 writers’ overall writing performance. The holistic rating showed that both groups improved after receiving one semester of writing instruction in ELI 73. However, the gain of the treatment group was much greater than that of the control group ($M = 14.67\%$ vs. $5.00\%$). The mean scores of the posttest indicate that the treatment group achieved slightly above average writing proficiency ($M = 4.18$), while the writing of the majority in the control group still remained relatively poor ($M = 3.31$). Referring to the holistic rubric, the treatment group students’ essays can be characterized as having a clear thesis and good organization, but also some points that are not fully elaborated. Their range of vocabulary and/or syntactic structure may be limited. The control group’s essays generally exhibited more serious problems, including thesis underdevelopment, lack of exemplification, and/or limited vocabulary.

These holistic findings can reasonably be taken to indicate that ER facilitated students’ overall writing performance. As Cohen’s $d$ (1.17) confirmed the large magnitude of the effect of ER, the students exposed to a more print-rich environment in the form of ER increased their overall written communication skills more than the students taught in a more traditional way. From a pedagogic perspective, this is promising in that only one semester of ER experience with
small portions of the in-class and out-of-class time allotted to ER brought about such an improvement in the quality of written production.

In an attempt to understand the effectiveness of ER more comprehensively, the students’ essays were further assessed by breaking them down into the five categories of content, organization, vocabulary, language use, and mechanics. Corresponding to the holistic findings, the descriptive statistics for these analytic ratings indicated that both groups made gains on the posttest. In addition, the gap between the two groups’ writing performance became larger on the posttest as the treatment group made more gains in each category than the control group.

Looking into each category, students in both groups made the greatest improvement in the areas of content and organization. Several reasons might account for these gains. One of the primary foci of ELI 73 was to help students improve in these two areas. That is, students not only received explicit instruction on how to write a good thesis statement and how to devise a clear organization but also had ample opportunities throughout the semester to practice these skills by writing multiple drafts they had to submit as their major course assignments. Feedback from course instructors and peers also emphasized these two areas more than any other skills, such as grammar or mechanics. This writing instruction and focused practice in the form of multiple revisions may largely explain their achievement.

It is, however, noteworthy that the progress made by the treatment group was much more evident than that made by the control group in both content (TR: 12.76%; CT: 4.38%) and organization (TR: 12.15%; CT: 2.91%). It is difficult to clearly pinpoint how or which characteristics of ER in fact generated this substantial improvement in the treatment group; however, the gains can be partially explained by the influence of the ER materials with which the students engaged over the semester. In other words, being constantly exposed to literature-based
texts, which usually have cohesive flow in a story’s unfolding plot, may have facilitated their skills for expressing their ideas in a nice flow conducive to clear communication. Several students’ responses to the open-ended survey questions and in the interviews suggested that they found reading short stories useful for getting a sense of how to weave their sentences together in a more coherent manner by, for example, using transitional phrases (e.g., however) or appropriate pronouns. Just as Tudor and Hafiz (1989) found that ER helped their students produce more semantically acceptable sentences, ER may have helped the participants in the current study learn how to appropriately use new words and expressions in the right context, thus producing more natural sounding sentences, and leading to more unambiguous and successful communication.

The similar degree to which content and organization improved is also worth noting. It may derive from their close association in the development of writing. For example, it is less likely that essays supported by well-reasoned content suffer from disorganization because the development of ideas or logical sequencing of content is usually connected to the way we organize our ideas into sentences or paragraphs. In this respect, however, it is also possible that the raters were subject to a “halo” effect, and failed to make a clear distinction between these two areas. Teasing these two aspects completely apart was probably challenging for the raters.

Of the five categories, vocabulary was the area in which the treatment group students demonstrated the least proficiency on their pretest ($M = 13.48, 67.41%$), showing a limited range of vocabulary along with frequent errors in choice and usage. This low level of vocabulary was displayed in the control group as well ($M = 12.83, 64.16%$). However, on the posttest, the treatment group attained a score of $77.56\%$ ($M = 15.51$) for vocabulary knowledge whereas the control group only reached $66.81\%$ ($M = 13.36$). This result came as no surprise, as a strand of
research has reported ER’s substantial benefits for vocabulary learning (e.g., Elley & Mangubhai, 1981; Hafiz & Tudor, 1990; Horst, 2005; Lai, 1993; Nation, 2015; Pigada & Schmitt, 2006; Tudor & Hafiz, 1989; Webb & Macalister, 2013). Although the present study did not assess lexical diversity or lexical density, the analytic scores for vocabulary do indicate that the treatment group students demonstrated an adequate range of vocabulary with no meaning obscured on the posttest, albeit some errors of word choice were still persistent. On the other hand, the control group maintained a somewhat limited range of vocabulary, with meaning obscured more often.

Research has consistently shown that the larger a student’s vocabulary size, the likelier it is that the student can understand texts easily. Hu and Nation (2000) specifically noted that knowledge of 98% of a text’s vocabulary is prerequisite for L2 learners to adequately comprehend the text. Several corpus studies (e.g., Hu & Nation, 2000; Nation & Wang, 1999; Webb & Macalister, 2013) have shown that graded readers are particularly useful in this regard in that they require a small vocabulary size to reach this 98% coverage due to the controlled and repeated usage of words in the texts. Consequently, ER offers an optimal environment for incidental learning of words by allowing students to encounter the same words over and over (Nation, 2015; Nation & Waring, 2013). As Nation stated (2015, p. 138), “Where a word is repeated, the occurrences of the word after the first meeting provide an opportunity for a combination of guessing from context clues and retrieval of the meaning of the word gained from previous meetings.”

However, it is important to keep in mind that knowing a word means not only understanding its meaning but also being aware of how to use it in a proper context (Laufer, 1997; Nation, 1990, 2001, Pigada & Schmitt, 2006); depth as well as breadth of vocabulary
knowledge is crucial in reading (Hudson, 2007). The fact that graded readers provide not just frequent repetitions but plentiful opportunities for learners to meet words in various and meaningful contexts makes them highly likely to enrich students’ vocabulary knowledge (Nation, 2015).

Taking the aforementioned points into account, one possible interpretation of the treatment group’s greater gains in vocabulary knowledge may relate to the cumulative encounters they had by means of ER, which might have enabled the students to learn and strengthen their knowledge of the form and meaning of new or partially known words. Another plausible reason might be their dictionary use. Dictionary use is usually discouraged in ER because it slows down reading, thus impeding fluent reading (Day & Bamford, 1998). In line with Grabe and Stoller (1997), Nation (2015), however, noted that as long as students are trained to use it “as a way of confirming a guess rather than replacing a guess” (p. 138), it can assist vocabulary learning. Students in the present study were encouraged to reduce their dictionary use, but it was not entirely banned. As the interviews revealed, students occasionally consulted a dictionary when they could not guess the meaning of unknown words even after several encounters in the text or when they thought the words held important meanings to understand the story. By meeting the word after confirming the meaning with the help of a dictionary, students could reinforce their learning.

In addition to meaning-based knowledge gains, Pigada and Schmitt (2006) found that ER can increase not only grammatical knowledge but also comprehensive knowledge of words. In fact, one of the most surprising features of the current study’s results was the improvement in language use. Despite the fact that most students have already acquired sufficient knowledge to write essays at the academic level without much difficulty in grammar when admitted to this
university, language errors were pervasive on the pretest, from basic errors to more serious ones. These errors often hindered their ideas from being clearly delivered to the raters, resulting in poor to fair performance on the pretest in both groups. However, on the posttest, the treatment group achieved the range of good to average \( M = 15.25, 76.25\% \); occasional minor grammar errors appeared in essays, but meanings were not obscured. Their sentence construction seemed relatively simple, but effective. Conversely, the control group remained at almost the same level as their pretest, improving only minimally \( M = 12.91, 64.56\% \). Their essays contained more serious problems in sentence construction as well as in grammar, which could have undermined the raters’ comprehension. As Tsang (1996) noted, meeting appropriate models of language through constant reading may aid ER treatment groups either to gain new knowledge or to consolidate their previous structural or grammatical knowledge. In particular, repetitive exposure to good language use in the form of meaning-focused reading could have also facilitated student learning.

Another speculation based on this finding is that the comfortable level of the reading texts may have exerted positive effects on their command of grammatical knowledge. As emphasized by many researchers, providing relatively easy materials to students is critical in that complex or less accessible materials would impede “the active assimilation of the linguistic input, thus preventing input from becoming intake” (Tudor & Hafiz, 1989, p. 176). The results of the current study may support their argument, if the results are taken to indicate that the cognitively less demanding reading enabled students to bring their awareness to language features while reading. Given that many students responded on the survey that they recognized new expressions \( 86.36\% \), sentence structures \( 75.00\% \), or grammatical features \( 59.09\% \) while reading, and
that they thought they gained knowledge in these areas due to ER, this interpretation seems persuasive.

Finally, the treatment group improved more than the control group in the category of mechanics, that is, spelling, capitalization, punctuation, and paragraphing. The pretest essays for both groups contained only occasional errors of this sort, which did not obscure meaning (TR: 83.41%; CT: 79.75%). This was anticipated because mechanics are not a major variable at this level; most of the students had already acquired a fair amount of mechanical knowledge. Spelling errors were minimal, and the most common punctuation errors were related to comma splices and quoting. On the posttest, the treatment group demonstrated near mastery of knowledge in mechanics ($M = 4.61$, 92.27%), and the control group also displayed good knowledge of mechanics ($M = 4.12$, 82.38%). A series of studies (e.g., Lai, 1993; Pigada & Schmitt, 2006; Polak & Krashen, 1988; Tudor & Hafiz, 1989) has provided evidence that spelling appears to be enhanced by pleasure reading. However, the explicit instruction about paragraphing that both groups received may have been more beneficial than ER, as Tsang (1996) noted.

To sum up, the analytic scores of the treatment group increased substantially more between the pretest and the posttest than those of the control group. It seems justifiable to suggest that ER appeared to provide natural input that fed students’ improvement in ways that explicit instruction could not have done, which aligns with Hedgcock and Ferris’s (2009) claim:

Certain aspects of language can only be acquired through extensive and authentic exposure to the L2 (i.e., through reading, listening, and interaction), and once learners arrive at a certain stage of L2 acquisition (intermediate to advanced levels), it is likely that their continued progress in that language will largely result from such natural exposure and not classroom instruction. (p. 214)
Moreover, the constant exposure to input along with writing practice as well as instruction that the treatment group had in the current study may have contributed to enhancing their abilities across all five areas. That is, their knowledge about writing could have been strengthened by the combination of input and explicit instruction, and this knowledge could have been further enhanced as students were able to confirm it while reading.

This section has discussed students’ categorical as well as overall writing improvement, as reflected in their essay tests. In the next section, students’ gains in writing analyzed with more objective measures are interpreted in relation to possible benefits of ER.

**The Impact of ER on CAF**

The second research question explored any possible effects of ER on students’ writing performance in terms of complexity, accuracy, and fluency. Overall, a small amount of gain was achieved in both groups. This may indicate that one semester of writing instruction, regardless of the intervention of ER, is insufficient to result in a great deal of improvement in CAF. However, in line with Leki, Cumming, and Silva’s (2008) claim that the three elements of CAF do not develop hand in hand and that fluency develops faster than the other two, the present study’s results demonstrated the tendency for the elements to change over time to different degrees. Students made the most evident progress in fluency with the least in complexity in both groups.

To reiterate the findings on complexity, the control group demonstrated a higher degree of complexity than the treatment group on the pretest. On the posttest, however, the control group showed less complexity, whereas the treatment group’s complexity increased minimally. This finding is opposed to that of Tudor and Hafiz (1989), in a study in which students who engaged in ER displayed decreased complexity. The authors postulated that the participants were
“modeling their written production on the simplified language of the graded reading materials” (p. 174). In the current study, the increase in the treatment group’s clauses per sentence, albeit almost negligible, may have stemmed from their writing shorter sentences and connecting them with coordinators (e.g., and, but, so) and using relative clauses more frequently. Such behavior could also be the result of their assimilation of the input, that is, an effect of their continuous exposure to graded readers that contain relatively simple, short sentences.

In contrast to the minimal changes in complexity, more distinct trends were revealed in terms of accuracy. The treatment group had a slightly higher level of accuracy on the pretest, and they maintained their superiority on the posttest. The control group’s accuracy level remained almost the same across the two tests. L2 writing research has shown that errors in articles, verb tense, or prepositions are the most common across proficiency levels (Larsen-Freeman, 1983); moreover, after teaching intermediate level classes for several years, I also have noted that these errors are not only the most frequent, but that they are persistent in spite of the students having received explicit instruction in the relevant grammar rules in the past. Therefore, as discussed in the findings of language use from the analytic rating, the notable improvement in accuracy in the treatment group seems to be a reasonable justification for suggesting that constant exposure to good language models through ER could have helped the students increase or consolidate their grammatical knowledge.

The fluency measures yielded intriguing results. In the simple count of words produced within the time limit (60 minutes), the treatment group surpassed the control group by producing more words on the posttest (TR: 398.73 words; CT: 324.25 words). This increase of writing readiness aligns with findings by Hafiz and Tudor (1990) and Lai (1993). However, the three other fluency measures revealed that the treatment group students produced fewer words per T-
unit, clause, and sentence than the control group. Reading regularly may have aided the students by improving their overall command of English, resulting in more word production with ease, but, as mentioned earlier, the relative shortness of sentences in their main input, the graded readers, could have also led to similar simplicity in the students’ output. In addition, students’ WPM indicated that the treatment group produced 1.25 words per minute more than the control group. However, this finding needs to be interpreted cautiously as individual students’ actual writing time was not taken into account. It is unclear whether all students used 60 minutes for writing. Therefore, WPM may not provide an accurate picture of students’ reading fluency or differences in the two groups.

Several important points are worth noting in relation to the CAF findings. First, as clause-based measures have been argued to be more effective than T-unit based measures for lower level learners (Ishikawa, 1995; Norris & Ortega, 2009; Wolfe-Quintero, 1998), they were also found to be more informative in detecting changes in the intermediate level students’ production in the present study. For example, in terms of accuracy, longer T-units were harder to make error-free, and therefore did not give much insightful information when comparing students’ writing ability. In line with Ishikawa’s (1995) study, where the most significant changes in student essays were in regard to clause-based measures, the current study also suggests that the clause, a relatively shorter and simpler construction, may be more sensitive as a unit of measurement when distinguishing among students at a single level or when there is minimal gain.

Second, a possible reason that no clear trends emerged may be related to individual differences. Although the students were in the same level of writing course based on their placement test results, the holistic and analytic ratings revealed that their writing proficiency varied, which may imply different strengths and weaknesses of individual students. It is also
plausible that some learners acquire certain features earlier than other learners; that is, that they
do not make equal progress (Bardovi-Harlig & Bofman, 1989). In other words, because of their
individual differences, some learners may have been more strongly influenced by ER in
particular aspects of writing, resulting in different levels of performance.

Larsen-Freeman (2006, p. 601) noted that even when students receive the same
instruction, “diverging patterns of development” can emerge, probably due to the way individual
students use their limited resources. She found that while a group exhibited steady improvement
in terms of CAF, individual students did not show a smooth progression of development, and
some students even became worse. Skehan (2009) explained this kind of less uniform
development by suggesting a trade-off hypothesis: “committing attention to one area, other
things being equal, might cause lower performance in others” (p. 511). Because learners have
limited attention spans and working memories, there can be, for example, a trade-off effect
between form (complexity and accuracy) and fluency. It is also conceivable that as students
strive to write more accurately, monitoring their production more carefully, the fluency and
complexity of their writing may be inhibited slightly (Skehan, 1989).

Looking into the overall CAF changes in the control group, the students increased their
fluency, producing more words along with longer clauses and sentences on their posttest, but
their complexity slightly decreased. Accuracy remained almost the same across the two tests.
Drawing on the trade-off hypothesis, this finding may indicate that the students generally tried to
write more, while not engaging in much risk-taking behavior by producing less complex
sentence constructions. For the treatment group, on the other hand, their complexity remained
almost the same, but their accuracy and fluency improved more evidently. These students may
have tried to write as much as they could, while still attending to accuracy at the same time.
However, as the increased ratio of clause per sentence on the posttest indicates, there is the possibility that the students simply produced more choppy clauses but not syntactically sophisticated and complex constructions. It is difficult to draw any solid conclusions at this point because the gains were minimal and may have been the influence of individual differences. Nevertheless, the diverging patterns of changes in CAF between the two groups may suggest some impact of ER on students’ writing performance.

Third, to expand the previous argument, another point that deserves further discussion is whether these varying degrees of changes in CAF should be interpreted as progress. As Ishikawa (1995, p. 65) noted, change in CAF “is not necessarily progress and may not reflect improvement,” partly because ideal measures that can be consistently applied across proficiency levels as well as to individual progress have not been established. Similarly, Casanave (1994) raised the question of how “progress” should be understood, because there can be so much diversity within a single group of students. In her study, almost half of the students did not progress, but their risk-taking behavior did increase. Along the same line, in the present study, because L2 learners at this level may frequently use trial and error in the process of developing and refining their written language, more complexity with less accuracy cannot be viewed as superior to the opposite tendency. As argued by researchers, the use of highly complex sentence constructions is not necessarily an index of better performance (Ortega, 2003), and highly accurate grammar production does not guarantee effective written communication (Pallotti, 2009). In particular, care must be taken when evaluating students’ writing performance based on a one-time writing assessment, as in this study. Therefore, a more in-depth analysis that could offer detailed pictures of individual students’ developmental pathways would be of great interest.
Although the objective measures employed for the study may represent the students’ explicit knowledge of L2 to some extent and reflect their attention while writing the essay, clear-cut impacts, or benefits, of ER cannot be claimed in the present study due in part to less-established objective measures and the possibility of unexamined individual differences (e.g., variation in risk-taking behavior or different developmental trajectories of individual students). However, it is worth noting the different characteristics of the writing produced by the two groups.

**Students’ Reactions to and Perceptions of the Usefulness of ER**

The third research question addressed students’ reactions to ER and their perceptions of its usefulness. Supporting previous findings (e.g., de Burgh-Hirabe & Feryok, 2013; Judge, 2011; Takase, 2007), the majority of students showed positive attitudes toward ER and found ER to be useful for many aspects of their target language learning. This study provides evidence that, as Day and Bamford’s (1998, p. 30) bootstrap hypothesis posits, students’ positive experience of ER leads to confidence in their reading ability and stimulates favorable attitudes toward reading, which, as a consequence, leads to subsequent reading.

The interview responses clearly indicated that ER was fun for the students. A few students commented that it relieved the stress that heavier reading caused them, and that they actually turned to ER when they were tired of reading academic texts (e.g., textbooks or journal articles). Their pleasurable experience of ER appeared to be augmented by being compared to the experience of academic reading, which they typically found difficult due to its conceptually demanding content and complex linguistic features. In addition, several interviewees who claimed to hate reading said they got hooked on reading once they found the right book. It is
remarkable that one book could actually change these students’ perceptions of reading—a finding that is in line with previous studies that have demonstrated the power of a single positive experience of ER (e.g., Cho & Krashen, 2002; Mason & Krashen, 1997).

Moreover, as also emerged in Nishino’s (2007) study, the students’ enjoyment in the present study was closely associated with the reading materials. Many interviewees mentioned that the freedom to choose their own books and the content of specific books were the most important factors in their positive attitudes, which in turn stimulated them to read further. For most of them, as long as the story kept them interested, they did not limit the scope of their book selection to a particular topic or genre. There were, however, some students who did show a strong preference for specific topics (e.g., romance, crime) or genres (e.g., fiction, nonfiction). This finding aligns with Day and Bamford’s (1998) expectancy value model, which emphasizes the importance of reading materials in influencing students’ decision to read.

Another finding worth discussing is that the ease and comfort of the reading were linked to students’ positive reactions toward ER to a great extent. Because they were able to easily comprehend the texts and this became even easier as the semester progressed, the students seemed to accept the reading as a stress-free regular routine. Moreover, handling English books without much difficulty boosted their confidence in reading. Being able to finish English books and reading without consulting a dictionary often clearly led them to feel a sense of achievement, in contrast to reading challenging academic texts. Students’ increased or recovered confidence from reading easy texts seems to have also played a role in triggering a cycle of reading.

As opposed to the many positive reactions to ER, the interviews also revealed some negative attitudes about ER. First, although the library contained around 250 books, which was thought to be sufficient to accommodate students’ diverse topical interests, some students
mentioned that they had a hard time finding books that interested them, which often discouraged them from further reading. This situation was at least partly due to the fact that many students had similar interests, and so certain books were checked out more often than others. Building a well-equipped library involving multiple copies of each book as well as a wide range of books is clearly necessary, as providing reading materials “capable of generating and maintaining learners’ spontaneous interest” (Hafiz & Tudor, 1990, p. 11) is key to successful ER implementation.

In other negative comments, a few students recalled that at the beginning of the semester they were unsure about ER’s effectiveness for their English learning because of the relative easiness of the reading materials provided. They did not perceive ER as a waste of time, but had doubts of its usefulness due to the lack of feeling a sense of learning. As one student remarked, while appreciating the pleasure side of ER, he or she wished to be learning more from the reading. This reaction is understandable considering these students’ goals as international students and the ways they have studied English in the past. To them, reading self-selected material for pleasure was not easily accepted as a helpful way to achieve their academic goals. Such feelings might be due to what Day and Bamford (1998) noted as an underlying belief of learners: “No reading pain, no reading gain” (p. 92). For most students, the ER was apparently a very different approach to reading, and because it did not involve looking up words or word-by-word translation, they seemed to feel they were not learning much. However, most students’ initial doubtful thoughts seem to have faded away as they came to enjoy reading more and to recognize the value of ER themselves.

According to the survey and the interviews, the students on the whole did not find the 10-minute writing activity as enjoyable as reading per se, but they still seemed to feel that it was a useful activity for their writing practice. The students particularly enjoyed the fact that the
writing activity allowed them to write freely about their books and did not require them to produce accurate writing, which relieved the anxious feelings about writing that some students suffered. This positive reaction toward writing activities was also reported by the control group interviewees who engaged in freewriting over the semester, which shared similar features with the 10-minute writing. The students appreciated the features of freewriting, in which students can freely explore and write about the topics, without giving much attention to grammar or organization, and they found it to be useful for their fluency development. However, it is noteworthy that while the control group interviewees pointed out that writing about topics given by their teacher or peers was less inspiring, the treatment group reported that writing about a self-selected book was one of the favorable elements of the ER. It appears that the 10-minute writing based on students’ natural reflections on their own reading not only relieved their nervous feelings about writing in English but also added enjoyable elements to the writing practice.

In addition, some students in the treatment group mentioned that they sometimes had no time to read between the two class periods in one week, which made it hard to find something new to say in their 10-minute writing assignment, resulting in repetitive writing. Another demotivating factor they reported pertained to writing topics. A variety of topics was formulated in an effort to encourage students to think about the stories from different angles or perspectives (e.g., writing different endings, analyzing characters) as well as to generate diverse writing opportunities (e.g., summarizing, describing, creative writing). However, the suggested topics sometimes did not quite fit the stories they were reading in a given week.

Corresponding to their favorable reactions to ER overall, the extensive reading survey and the interviews revealed students’ positive perceptions of the usefulness of ER for their
English learning. Students perceived ER as most useful in increasing their reading comprehension and reading speed. These two benefits of ER have also been reported in a number of other studies (e.g., Al-Homoud & Schmitt, 2009; Beglar et al., 2012; Beglar & Hunt, 2014; Bell, 2001; Elley & Mangubhai, 1983; Hitosugi & Day, 2004; Iwahori, 2008). The students in the current study thought ER to be useful for writing as well, especially for learning expressions or sentence structures, as well as grammatical features to a lesser degree. They probably perceived their improvement in reading clearly since they were able to read books faster and with more ease as the semester progressed, but they may not have experienced a similar “sense of learning” in their writing. Krashen’s (1984, p. 20) assertion that learners gain writing ability subconsciously while they are involved in voluntary and pleasure reading may partially explain these findings.

In contrast to the findings regarding reading and writing, the survey showed that the students perceived ER to be less helpful for increasing their speaking and listening ability. However, it is noteworthy that a good number of students felt that ER did enhance their speaking (43.18%) skills. Some students mentioned during the interview that they learned words or expressions from their books and had actually used them in their daily lives. This finding is relevant to Hedgcock and Ferris’s (2009, p. 251) assertion that “whereas reading does not directly address pronunciation, accent, or formal speaking skills, both formal speeches and informal interactions are certainly facilitated by having access to an extensive vocabulary and a grasp of varied syntactic and morphological structures.” The natural and authentic conversations in some of the graded readers may have provided a good model for students to learn from. The five minutes of pair discussion about their reading in each class was also reported as helpful for practicing speaking as well as listening. Students who were self-conscious about their English
and reluctant to speak in class commented that they felt comfortable about these discussions because they were talking about something they knew well.

**Benefits of Incorporating ER into an EAP Writing Class**

One major benefit generated from incorporating ER into this writing course was the increased amount of reading and writing. Although there were individual differences, the amount of reading and writing that the students were accustomed to engage in on a weekly basis was generally small in both groups at the beginning of the study period.

According to their self-reported reading amounts, they read on average 3.87 hours per week, for both academic and personal purposes. One of their major reading activities involved reading postings on social network services (e.g., Facebook, Twitter) or reading news articles online, but none of the students read English books on a regular basis. Many students reported that they neither read much nor had interest in reading, and this seemed to shape their negative perceptions of reading. They found reading frustrating, and their frustration was particularly associated with failing to clearly understand the texts or encountering endless unknown words in the texts. In particular, the interviews revealed that most of the reading instruction had been based on intensive reading during their previous formal English education, requiring students to read difficult passages and translate them into their first language. Apparently, reading for pleasure was not even an option for them, because getting high scores on high-stakes reading tests, such as university entrance exams, was their priority, and their school curricula were designed accordingly.

In addition, the amount they read per week in their first semester at this university, except for ER, was low, with a mean of 6.27 hours per week, and also varied greatly from student to
student. Their reading involved reading textbooks, PowerPoint slides, and handouts for academic purposes as well as reading news articles on the internet for personal purposes. In this respect, engaging in ER for one semester seemed not only to help students increase the amount of their reading but also to explore another purpose of reading: reading for pleasure. Students chose having such a chance to read English books in and outside of the classroom as one of the advantages of ER, as they would not have thought of doing such reading by themselves. At least to a few students, this experience served as a turning point to shift their perceptions about reading from negative to positive as they became more interested in reading.

It was the same case for writing in English, of which the students had had even less experience. Many students had never taken an English writing course before, or if they had, the courses centered around grammar instruction or translation. Their lack of experience and confidence may explain their poor performance in writing (Kubota, 1998) on the pretest as well as an absence of interest in writing.

As discussed earlier, both groups of students reported that it was helpful to engage regularly in writing practice, in the 10-minute writing for the treatment group and in the freewriting activities for the control group. However, in accord with Casanave’s (2004) argument that journal writing was advantageous for deepening students’ thinking and building their fluency, the 10-minute writing, which is almost a type of journal writing, seemed to be more beneficial than simple freewriting, in which students have to write on topics initiated by their teacher or classmates. As Zamel (1992) noted, the benefits of journal writing lie in its characteristics of promoting a dynamic reading experience as well as fostering writing development; likewise, writing about books led the students to think in more depth and produce more detail in their writing, which I believe worked as one of the significant factors in the
treatment group’s greater gains in fluency. Many students in the current study reported that they had tried or used words or expressions in their writing that they learned or picked up while reading, which supports Hedgcock and Ferris’s argument (2009, p. 215) that “reading provides materials to write about, linguistic tools with which to express ideas, and rhetorical models to learn from.” I suggest that, regardless of whether their usage of such items was correct or incorrect, trying out the new knowledge gained from reading provided room for learning.

Noting that one of the challenges EAP students face is their lack of experience in reading and writing, Grabe and Zhang (2013) argued that in order to help students acquire academic literacy, teachers should “provide many opportunities for students to practice reading/writing tasks (even if short tasks) so that they (a) build confidence and fluency and (b) receive consistent feedback on their writing” (p. 20). Reading easy books and writing short personal responses in the present study seemed to some students to be an informal and even frivolous task, but it clearly generated more reading and writing practice, and also strengthened the relationship of the two literacies for the participants. The accompanying 10-minute writing activity not only generated ample opportunities for practicing writing per se but also gave learners more chances to express their personal engagement with their reading. Students often related the stories to themselves, reflecting on their own experiences, which consequently brought about individual students’ unique meaning-making processes. As the interviews and the 10-minute writing revealed, some students became interested in a certain topic after they read and wrote about it, and this interest led them to further research about the topic (see Appendix L for sample 10-minute writing essays). This supports Grabe and Kaplan’s (1996) argument that personal writing such as journal writing is useful for generating ideas and topics as well as finding sources of interest.
Another finding worth noting is students’ reactions about reading literary texts for ER. They seemed to enjoy the fact that they could experience various types of texts (e.g., literary and non-literary texts) in class, which may have helped them broaden their knowledge about different kinds of texts. This finding concurs with what Hirvela (2001b) found about using literary texts in EAP contexts. He observed students’ favorable reactions to literary texts in an EAP writing course, where the course assignments included writing response papers to both literary and non-literary texts. One of his interesting findings is that students showed a greater level of enjoyment when reading literary texts than non-literary texts, and this in turn led to easier response writing. Hirvela (2001b, p. 345) pointed out the advantage of using both types of texts for writing as follows:

Providing an array of reading experiences through both literary and non-literary text types will enable students to gain additional practice in using different reading strategies and create more opportunities for EAP teachers to discuss a host of such strategies. At the same time, writing about both literary and non-literary texts opens students up to writing in a variety of repertoires.

Given that EAP students will need to read different kinds of texts as they move from course to course in a university, helping students become familiar with diverse reading and writing assignments with a good balance may be an important matter for teachers in EAP settings to consider.

Because one of the goals of this study is to determine the potential for incorporating ER into EAP writing classrooms, examining students’ needs and goals at this level seemed necessary. On the background survey, students in both courses were asked which specific aspects of writing they aimed to improve; their identified goals were varied, including overall writing skills, grammar, coherence, organization, vocabulary, fluency (speed and length), expressions, and more writing practice. Given these diverse goals and needs, the implementation of ER
coordinated with the 10-minute writing played a positive role in helping students come closer to their personal goals. For example, as discussed, the analytic scores showed that the treatment group performed better in almost all of the measured areas, which provides strong support for implementing ER in EAP writing classes.

I believe that the findings of this study have provided some implications as to how such meaningfully combined reading and writing activities promoted students’ language development, thereby also strengthening the relationship between reading and writing for these learners. It is noteworthy that the self-initiated and pleasant characteristics of ER facilitated an active role for the students as readers by leading them to approach the text from their individual perspectives. As evidenced by a few earlier studies (e.g., Hafiz & Tudor, 1990; Lai, 1993; Tsang, 1996; Tudor & Hafiz, 1989), it is appealing that this pleasure-based leisure type of reading can lead to the improvement of writing. However, as Belcher and Hirvela (2004) pointed out a decade ago, we still need more studies to investigate whether students actually become better L2 readers and writers as a result of integrated reading and writing instruction. I will discuss these implications in more depth in the following chapter.
CHAPTER VI
CONCLUSION

In addition to providing a brief conclusion to the study, this chapter discusses the limitations of the study and makes recommendations for future research. Given the nature of this classroom-centered study, this chapter addresses pedagogic implications in some depth, pointing out challenging issues surrounding ER along with some viable and practical suggestions.

Conclusion

Research has shown incontrovertible evidence that ER promotes target language development in many different aspects, including reading comprehension, reading rate, and vocabulary learning, as well as affective domains such as attitude and motivation. The diverse benefits of ER in the writing domain that the present study found provide empirical evidence for Hedgcock and Ferris’s claim (2009): “Extensive reading naturally exposes readers to naturally-occurring phrasal and clausal patterns, repeated and alternate uses of lexical items and their spellings, and a range of other graphological features such as paragraphing, punctuation, and capitalization conventions” (p. 214). It is illuminating that ER can assist students to improve their communicative competence in written communication, and it benefits areas of writing ranging from content to mechanical knowledge. However, there is clearly a paucity of research in the writing domain. Much more research needs to be done to provide a conclusive picture of the value of ER for the development of writing.

Although the data were limited in some respects, several factors that led this implementation of ER to be successful are worth highlighting. First, the curriculum was carefully designed to encourage students’ engagement with reading, by weaving it harmoniously with
writing practice and interaction with classmates (i.e., reading-writing-sharing). These sequential and recursive pedagogic activities carried out throughout the semester appeared to prompt students’ constant engagement with texts. Second, the students’ target language level seemed appropriate. Although these students were admitted to the University with an acceptable TOEFL score, their reading and writing experience in English was very limited. Because of this, many students had difficulties dealing with academic literacies in their first semester at the University. Therefore, regularly engaging in the integration of reading and writing through ER seems to have been a good practice for increasing both literacies in quantity and in quality. Last, as the students themselves recognized the value of ER, they became more actively engaged in ER, which was probably the most salient factor that resulted in a positive outcome of the study. Most importantly, the students mostly welcomed this new approach.

In relation to this point, another significant contribution of the present study is to show the feasibility of bringing ER to EAP contexts. I hope this study will draw the attention of ER practitioners and L2 educators to the possibility of utilizing ER in higher education, where many students struggle in their transition to academia. ER is usually conducted or encouraged in EFL settings as a way to increase the learners’ exposure to the target language. Considering the challenges that ESL learners face, however, ER can be a valuable approach to help them to better deal with the substantial amount of reading and writing they encounter in an academic setting. Although there may be limitations on the extent to which ER can result in improvement in L2 writing or to which it can be usefully integrated with other pedagogic practices, it is worthwhile to have this evidence that learning can occur naturally while the learners enjoy reading in and of itself with no intention to learn.
Limitations

While this study found promising benefits of ER, they need to be viewed with a certain degree of caution. Limitations regarding the experimental design need to be addressed first. The students were living in an English-speaking milieu at the time of data collection, so it is not clear to what extent their gains in writing can be solely attributed to the impact of ER. That is, individual students were exposed to varying degrees of input in their daily lives, which might have affected the outcomes of the study. For example, as the participants came from various academic fields, their reading and writing workload in their other courses could not be controlled. This methodological limitation might suggest that it is difficult to posit a meaningful relationship between ER and writing. However, according to the students’ self-report, the control and treatment groups had similar average amounts of reading and writing across one semester. Moreover, considering that all students in both groups experienced very similar class instruction and activities except for the ER the treatment group engaged in, ER appears to be a good indicator of the students’ better performance in the treatment group.

Creating a true experimental design is difficult in real educational settings (Grabe, 2004), and therefore efforts were made to keep extraneous variables to a minimum by utilizing the same course syllabus and lesson materials. There were nonetheless still elements that could not be fully controlled in this quasi-experimental design. The influence of different instructors no doubt created different classroom interactions that could have affected the outcomes. For example, individual teachers’ teaching styles as well as teaching skills may have influenced students’ gains in writing. Teachers’ personality or rapport with students also could have affected students’ involvement in class, which could contribute to both their improvement and their overall attitude toward the course.
Another limitation is the small number of students that participated in the study. Although it is noteworthy that several studies have found meaningful results with a relatively small number of participants for ER groups (e.g., Tudor & Hafiz, 1989: \( n = 16 \); Tsang, 1996: \( n = 36 \)), the small sample size may undermine the generalizability of the findings. Moreover, a majority of the participants were from Asian countries, and thus might not constitute a representative sample of diverse ESL students. Therefore, the study cannot guarantee the generalizability of the findings to other ESL programs involving different context variables such as participants’ nationalities and proficiency levels or the purposes of programs.

**Pedagogic Implications**

Given the exploratory nature of this study and its contextual limitations as discussed above, any teaching implications based on the results of this study should be cautiously interpreted. However, the findings still suggest some general implications for the implementation of ER and its effectiveness in the EAP classroom.

**Essential elements for a successful ER program.** Day and Bamford’s (2002) ten principles may delineate the ideal situation for taking full advantage of ER, but there is much variation in the ways in which ER has been implemented from study to study depending on context or pedagogic aims. Some studies followed principles that are only feasible in their settings; other studies did not report sufficient information as to how ER was implemented in their specific context (Waring & McLean, 2015). Although slight differences exist among researchers (see Nation & Waring, 2013; Waring & McLean, 2015), there seems to be general agreement that two core principles (# 3 and 4) at least are necessary for an implementation to be considered ER. That is, the baseline is that students read a great deal of self-selected materials.
In particular, Nation and Waring (2013) argued that reading materials are the most important element in ER. Given that challenging books may frustrate L2 readers and negatively affect students’ motivation to read further (Lyutana, 2011; Stoller, 2015) and that the present study also provides evidence that easy access to interesting books at a comfortable level or at an acceptable difficulty level is essential for students, the importance of reading materials should not be underestimated. Therefore, a well-equipped class library involving a large quantity of interesting books embracing varying proficiency levels should be ensured, as it is the fundamental condition to constantly provoke students’ genuine interest in reading (Hedgcock & Ferris, 2009).

Another important element that played a great role in keeping students interested in reading in the present study was a comfortable and inviting reader community. There are many cases in previous studies in which ER was assigned as an out-of-class activity in the form of homework assignments (e.g., Hafiz & Tudor, 1990; Robb & Susser, 1989; Tudor & Hafiz, 1989). However, in support of arguments made by other ER advocates (e.g., Macalister, 2010), setting a certain amount of time aside for in-class reading was crucial in the present study, not only for encouraging students as members of the whole-class reader community, but also for promoting their interest in reading. Moreover, in-class ER offered an opportunity for students to check their own reading progress as well as to learn what kinds of books their classmates were reading, which particularly seems to have motivated those who did not have much interest in reading, serving as peer pressure in a positive way. Engaging in pair or group discussion, sharing their stories, and interacting with their classmates further facilitated students’ comprehension and sparked their interest in stories read by their classmates.
The survey in fact revealed that the students found in-class ER more enjoyable than out-of-class ER. A couple of reasons can account for this. One benefit of in-class ER lies in the fact that students did not need to make an effort to find time to read. A certain amount of time, a reader community, and plenty of books were already set up for them, as opposed to out-of-class ER in which reading was entirely up to their free choice. As the interviews revealed, for some students who struggled to keep up with all the assignments in their other courses in their first semester, finding time for ER was challenging. ER seemed to be readily replaceable, especially when they had other priorities such as reading academic articles and writing papers. This may indicate that merely having a favorable attitude toward ER in and of itself may not bring about subsequent reading (Yamashita, 2004) or the development of a reading habit (Rodrigo et al., 2014). Therefore, from a pedagogic perspective, creating an environment in which all students could see the value of reading, thereby appreciating their reading, is essential for promoting the development of a life-long reading habit beyond the classroom.

Equally important as establishing a positive and noncompetitive classroom environment, the active participation of a teacher turned out to be critical (Day & Bamford, 1998, 2002). Even though the importance of the teacher’s role has been little emphasized in the literature (Macalister, 2010), the present study illustrates the power of the teacher-as-role-model in helping students grasp the excitement of reading (Stoller, 2015), which is a fundamental impetus for the successful implementation of ER. As discussed, one of the ultimate goals of ER is to help students develop learner autonomy so that they can eventually develop a life-long reading habit (Rodrigo et al., 2014; Takase, 2007). Yamashita (2015) stressed the commitment of teachers in this process. In particular, when students are first introduced to ER or until they develop a certain degree of learner autonomy, I believe that teachers’ guidance and effort in helping students see
the merit of ER is of great importance. For instance, along with an occasional consultation, the teacher’s constant guidance customized to individual students’ reading pace and interests supported them to be able to take charge of their reading as the semester progressed. The book recording form also served as a useful tool for teachers to check what individual students were reading or how much they were involved in ER. In sum, access to a great deal of materials, an inviting reader community, and the teacher’s constant participation were obviously crucial factors that led to the successful implementation of ER in the present study.

Creating reading and writing opportunities through ER. Another pedagogic implication that emerges from this study is that some types of pedagogic combinations enhance L2 learners’ language development. Highlighting the role of ER in increasing confidence and positive attitudes toward reading as well as fluency, Day and Bamford (1998, pp. 44–45) claimed that the potential role of ER in preparing students for academic reading should not be undervalued. They stressed that without these skills, it is less likely that students can deal with challenging academic texts and develop the critical thinking skills that are required in higher education.

Extending Day and Bamford’s assertion, Grabe (2003, p. 249) noted that the interconnection between reading and writing can be made more effective by combining ER with constant writing practice. Hedgcock and Ferris (2009, p. 255) also encouraged the coordination of ER with reading-writing courses. Acknowledging the challenges, they nevertheless argued that when reading provides a topical basis for writing and the material is self-selected, students will be more motivated to write about it. Given these benefits, the successful integration of ER into an EAP writing classroom in this study demonstrated how combining two pedagogic practices can create a synergistic effect. Teaching students to become successful L2 writers and
helping them meet the demands of academic writing conventions and disciplinary expectations is a long-term process, and ER per se may not directly guide ESL students into academic writing. Nevertheless, as Grabe (2001, p. 37) stressed, “Many activities and skills for combining reading and writing need to be practiced if they are to be strong foundations for the more complex tasks required in academic disciplines.” Furthermore, as Spack (1993, p. 185) noted, one of the important roles of a teacher is to “establish mediating links to fill a gap between what students bring to the academic community and what the academic community will ultimately expect of them.” I believe that constant practice of combined reading and writing, on the basis of the increased amount of both, can play a crucial role in building a solid foundation of academic literacy for students at this level, who are transitioning to an academic environment but not quite ready to deal with highly demanding reading and writing. General reading and writing proficiency should be established as a prerequisite in order for students to reach the level of proficiency required in academia.

Controversial and challenging issues in the implementation of ER. Utilizing easy materials is one of the central tenets of ER, but the usefulness of it has been somewhat controversial. Although proponents of ER generally support students’ use of graded readers for gaining reading ability, new vocabulary, and pleasure in reading, critics have questioned the quality of graded readers by asserting that graded readers are poorly written and thus not qualified for learning. They criticize graded readers on the grounds that simplified texts distort the language in the original. Furthermore, the adapted or linguistically manipulated texts written for pedagogic purposes lack the richness of the original, thus depriving L2 learners of the experience of subtlety or authenticity in their reading.
This concern may be true for some poorly written graded readers (Day, 2004; Nation & Wang, 1999). However, in response to the viewpoint that authentic text is natural, interesting, and relevant, while simplified text is not pedagogically sound due to its unnaturalness, Day and Bamford (1998) argued that authentic (i.e., not adapted) and simplified materials should not be seen as mutually exclusive. Moreover, Day (2004), in his exhaustive discussion of authentic materials, pointed out that there is neither agreement on what authentic means nor clear empirical evidence that using authentic materials has any benefit (p. 110). Rather, authentic materials can be too difficult for L2 learners to digest, and thus can make students perceive reading as a laborious instead of a pleasant experience. They can cause learners to take a word-by-word approach to reading (Day, 2004), leading to confusing learning to read with reading to learn, which may be less likely to result in the development of reading skills (Nuttall, 1996). Therefore, Day (2004) strongly asserted that the issue of materials should be discussed in terms of appropriateness; using appropriate materials for students’ linguistic level and further applying this concept of appropriateness to class activities, tasks, and exercises should be of interest to teachers. Similarly, Bassett (2011) pointed out that a graded reader should be judged as “a story in its own right and in its appropriacy for its target readership” (p. 28). In this regard, graded readers should be understood as well-adapted materials to suit the needs of a particular community: slow and less proficient L2 readers in general. In addition, because of the nature of literary texts, graded readers better facilitate the active involvement of readers and writers compared to information-based texts, which allow learners to take more passive roles (Bassett, 2011). In support of this view, a series of studies has shown the superiority of graded readers over original texts in enhancing vocabulary development due to the graded readers’ plentiful repetition of frequent words (e.g., Horst, 2005; Hu & Nation, 2000; Nation, 2009; Webb &
The students in the present study in fact did not seem to care about “authenticity” but did care much about whether the content was interesting enough to command their attention. While the majority of students preferred to read graded readers because they are usually short and easy to understand, some preferred to read books written for young readers because they usually contain more detail.

The quantity and quality of individual students’ ER practice were expected to be different because of the demands of university life on individual students’ time, their course loads, and their motivation. As in previous studies (e.g., de Burgh-Hirabe & Feryok, 2013; Grabe, 2001; Judge, 2011; Macalister, 2008) as well, one of the participants’ most commonly reported concerns about ER was the time commitment. Advocates of ER contend that the primary goal of ER in a curriculum should be “to convince students of its value so that they will continue reading extensively on their own once the class is over” (Hedgcock & Ferris, 2009, p. 220). However, during the interview, a common response to the question about students’ willingness to do ER beyond this course was “if I had more time.” As discussed earlier, having a positive attitude toward ER did not seem sufficient to bring about subsequent reading, especially when students had other priorities. As one way to encourage students’ engagement, researchers (Ferris & Hedgcock, 2005; Grabe, 2001; Green, 2005) have argued that the amount of reading and writing tasks can be negotiated in accordance with students’ specific goals and needs. Teacher guidance as to how to develop strategies to make time for pleasure reading in a busy schedule is also necessary (Ferris & Hedgcock, 2005).

Time constraints are often discussed from the teachers’ perspective as well, particularly in higher education contexts where there is less flexibility of the curriculum and a heavy focus on academic literacy skills. Silent reading time in class is neither considered direct teaching (Nation,
Moreover, the lack of instant outcomes often leads to ER being dropped from the curriculum before the benefits of ER are recognized, or never even put into practice (Grabe & Stoller, 2002). Macalister (2010), an advocate of ER in higher education, investigated the relationship between teacher cognition and classroom practices of ER. Interestingly, the teachers interviewed in his study displayed strong beliefs about ER, recognizing its benefits and acknowledging its indispensability, but in many cases, their favorable predisposition did not lead to actual practice in their classes.

ER is “a form of implicit learning” (Yamashita, 2015, p. 361); it is natural for it to require a time commitment for benefits to come to the surface (Grabe & Stoller, 2002). In order to raise awareness of this time commitment in implementing ER, especially in an EAP context, teachers’ understanding of the true merit of ER should be supported. Further, within time constraints, the integration of ER into L2 literacy programs should be encouraged by keeping a balance with other forms of instruction (Yamashita, 2015). In this regard, the present study showed how the integration of ER into writing courses can be a useful pedagogic tool that offers a meaningful connection between as well as improvement in receptive and productive skills.

Suggestions for Future Research

Nevertheless, this study has shown clear evidence of the benefits of ER for the development of university ESL students’ academic writing. Also, despite the substantial impact of ER on different aspects of L2 development reported in numerous studies, such findings need to be interpreted cautiously because many studies involve methodological problems. That is, whether the observed positive effects can be solely ascribed to ER is somewhat debatable due to either the lack of comparable control groups (e.g., Iwahori, 2008; Lai, 1993; Tudor & Hafiz,
or simply additional exposure that experiment groups had (e.g., Robb & Susser, 1989). As addressed in the previous section, not all variables were controlled in the present study either; therefore, its contribution is limited. Conducting a similar study in an EFL context where the target input can be more efficiently controlled may contribute to determining the true impact of ER on writing.

As an extension of this study, it would be interesting to look more closely into the relationship between input and output. For example, knowing how different types of input would affect learners’ written production may bring more insight into their relationship. For example, as some studies (e.g., Nation, 2009; Webb & Macalister, 2013) have reported that graded readers are better than other types of texts for target language development, possible similar or different degrees of writing improvement in association with different types of input (e.g., ER vs. academic texts) could be a topic for future studies to examine. In addition, the current study administered only one writing test to evaluate students’ writing performance; it would be useful if other types of writing (e.g., narrative writing) tests could be administered as well. Because different genres of writing require different writing skills (Hayes et al., 2000), the output would show different strengths and weaknesses. Comparing students’ writing performance in different genres of writing would provide a more holistic picture of the impact of ER on writing.

The holistic and analytic ratings employed in the study were useful for investigating the students’ overall writing improvement. In an effort to investigate the influence of ER more specifically, the objective measures were also used to analyze the triad of CAF. However, looking at the participants’ improvement as a group or merely looking at the increase or decrease of, for example, the ratio of T-units or the ratio of error-free clauses did not provide much information as to how individual students’ writing unfolds due to the influence of ER across the
semester. Therefore, of particular importance in future research will be scrutinizing individual essays by breaking them down at a micro level and tracking changes over time. As macro-level group means are likely to conceal the variability of idiosyncratic individual learners’ progress (Larsen-Freeman, 2006), examining individual essays in detail would offer much broader insights on the writing development of students at this level and its changes with regard to input.

As the study period was only one semester, it would definitely be interesting to see the results of a similar study continued for more than one semester. According to Nakanishi’s (2015) meta analysis, length of exposure is a critical element in ER. Unveiling whether more longitudinal exposure to ER promotes substantial progress in L2 writing and influences different aspects to different degrees would deepen our understanding of the long-term effects of ER. Looking into students’ reading improvement along with their writing development would also provide useful information for discussing the actual feasibility and practicality of implementing ER in an EAP context.
APPENDIX A: Background Survey

1. Name:

2. Age:

3. Email address:

4. Native language(s):

5. How long have you lived in English-speaking countries?

6. How long have you been learning English?

7. Are you an exchange student (Please circle)? Yes (1 semester/ 2 semesters) No

Undergraduate Students:
   a. What do you think you will major in?
   b. What non-ELI classes are you taking?

Graduate students:
   a. What field are you specializing in?
   b. What non-ELI classes are you taking?

8. In which language skills do you feel strongest? (Speaking, Listening, Reading, & Writing)

9. In which language skills do you feel weakest?

10. Briefly describe what you think are your strengths in your English writing abilities.

11. Briefly describe what you think are your weaknesses in your English writing abilities

12. What do you expect to learn in this class? What are your goals?
A. Writing in English

1. Do you enjoy writing in English in general?
   If Yes, why?
   If No, why?

2. How often do you write in English?
   Never-----Seldom-----Sometimes-----Often-----Always (almost everyday)

3. How many pages on average do you write in English per week including personal journals and academic essays?

4. What kinds of writing do you usually do? (circle all that apply)
   emails/ SNS (e.g., facebook) / academic papers/ personal journals/ others (  )

5. How would you assess your overall writing ability in English?
   Very poor-----Poor-----Average-----Good-----Very good

B. Reading in English

1. Do you enjoy reading in English?
   If Yes, why?
   If No, why?

2. How often you read in English?
   Never-----Seldom-----Sometimes-----Often-----Always (almost everyday)

3. How many hours on average do you read per week?

4. What materials do you usually read?

   newspapers / magazines / novels / comics / academic texts/ others (  )

5. Do you read for fun (e.g., read English books for fun, not for academic purposes)?

   If yes, how often? Never-----Seldom-----Sometimes-----Often-----Always
   How many hours on average do you read per week?
   If no, why?

6. How would you assess your overall English reading ability?

   Very poor-----Poor-----Not bad-----Good-----Very good
APPENDIX B: Writing Test

Topic: What is your attitude toward writing?
Time: 60 min
Instruction: Support your opinion with your own experience, observations, and/or background reading, and information from the quotations below.

Lopez-Ortiz (2005)

“Is the purpose of writing (publicly) for the entertainment of the reader, for the education of the reader, or for the communication between reader and writer that can develop? Or is it a wonderful combination of all three? You can’t say that you write without consideration of the reader. That, I think, is an impossibility.”

Blume (2001)

“A lot of my readers ask me for ‘writing tip.’ I wish it were that easy! There are no hard and fast rules for writing, and no secret tricks, because what works for one person doesn’t always work for another. Everybody is different. That’s the key to the whole business of writing-your individuality.”

Raibert (1985)

“My formula for good writing is simple: once you decide that you want to produce good writing and that you can produce good writing, then all that remains is to write bad stuff, and to revise the bad stuff until it is good.”

Fischer (2003)

“All writing is worthwhile, I think, and can be read with profit by someone. It is amazing that human beings, living always with such complexity, can say anything at all. So whatever anyone writes is worth reading, however clumsy it may seem according to canons of style.”

Delillo (1991)

“I write to find out how much I know. The act of writing for me is a concentrated form of thought. If I don’t enter that particular level of concentration, the chances are that certain ideas never reach any level of fruition.”
APPENDIX C: Extensive Reading Survey

Instructions: Please read each sentence carefully and circle the phrase that best describes your feelings toward extensive reading and the 10-minute writing.

1. I enjoyed reading books in class.
   
   Strongly Disagree---Disagree---Not Sure---Agree---Strongly Agree

2. I enjoyed reading books at home.
   
   Strongly Disagree---Disagree---Not Sure---Agree---Strongly Agree

3. I think reading books helped me improve my reading comprehension ability.
   
   Strongly Disagree---Disagree---Not Sure---Agree---Strongly Agree

4. I think reading books was useful for me to improve my overall written language (e.g., vocabulary, expressions).
   
   Strongly Disagree---Disagree---Not Sure---Agree---Strongly Agree

5. I think understanding stories got easier as the semester progressed.
   
   Strongly Disagree---Disagree---Not Sure---Agree---Strongly Agree

6. I felt comfortable while reading books.
   
   Strongly Disagree---Disagree---Not Sure---Agree---Strongly Agree

7. I enjoyed writing about the stories that I read.
   
   Strongly Disagree---Disagree---Not Sure---Agree---Strongly Agree

8. I think I enjoy reading books now more than before.
   
   Strongly Disagree---Disagree---Not Sure---Agree---Strongly Agree

9. I think writing activities were helpful for me to understand the stories better.
   
   Strongly Disagree---Disagree---Not Sure---Agree---Strongly Agree

10. I think reading books helped me feel more confident about writing.
    
    Strongly Disagree---Disagree---Not Sure---Agree---Strongly Agree
11. I was able to guess the meaning of unknown words using the context in the story.

   **Strongly Disagree---Disagree---Not Sure---Agree---Strongly Agree**

12. I think I will keep reading English books by myself after this course.

   **Strongly Disagree---Disagree---Not Sure---Agree---Strongly Agree**

13. I think reading books was useful for improving my reading speed.

   **Strongly Disagree---Disagree---Not Sure---Agree---Strongly Agree**

14. I think reading books was useful to learn sentence structures.

   **Strongly Disagree---Disagree---Not Sure---Agree---Strongly Agree**

15. I think reading books helped me to increase my grammar knowledge.

   **Strongly Disagree---Disagree---Not Sure---Agree---Strongly Agree**

16. I think I learned some useful expressions while reading.

   **Strongly Disagree---Disagree---Not Sure---Agree---Strongly Agree**

17. I think reading books helped me improve my English speaking ability.

   **Strongly Disagree---Disagree---Not Sure---Agree---Strongly Agree**

18. I think reading books helped me improve my English listening ability.

   **Strongly Disagree---Disagree---Not Sure---Agree---Strongly Agree**

19. I have more confidence now about reading English books.

   **Strongly Disagree---Disagree---Not Sure---Agree---Strongly Agree**
APPENDIX C (continued): Open-ended Questions about ER and the 10-minute Writing

1. What are your thoughts about extensive reading (topics, levels, enjoyment, etc.)? What did you like/dislike?

2. Do you think extensive reading was helpful for improving your English? In which skills (reading, writing, speaking, listening) do you think you have benefited the most from ER? Why?

3. Do you think extensive reading was helpful for increasing your knowledge about written language (e.g., vocabulary, expressions, idioms, etc.)? Why?

4. How do you feel about the writing activity (the 10-min writing)? Was it helpful for you to understand or remember stories better?

5. Do you have any suggestions to make ER and the 10-minute writing activity more meaningful/useful for you or for future ELI 73 students?
APPENDIX D: Interview Questions

A. Interview Questions for the Treatment Group

Introductory Question: What do you remember most when you recall ELI 73 this semester?

a. Extensive Reading:
   1. What was your first impression about extensive reading?
   2. What did you like/dislike about the idea of reading easy books in-and out-of-class?
   3. Do you think extensive reading has helped you to enjoy reading more?
   4. Do you think extensive reading has helped you build a reading habit?
   5. Which language skill(s) do you think benefited the most from extensive reading?
   6. Do you think you will keep reading English books by yourself in the future?
   7. Do you recommend extensive reading to future ELI 73 students? Why?

b. 10-minute Writing:
   1. How do you feel about the 10-min writing?
   2. Do you think the 10-min writing was useful for you?

B. Interview Questions for the Control Group

Introductory Question: What do you remember most when you recall ELI 73 this semester?

1. What was the most useful lesson or activity for you in ELI 73?
2. Did you have any lessons or activities connecting reading and writing?
3. What kinds of lessons that ELI 73 did not cover do you think would have been useful for you?
4. What do you think about the idea of reading short stories and using them for writing practice?
APPENDIX E: Consent Forms

Consent Form for the Treatment Group:

Implementing Extensive Reading in a University EAP Writing Class

My name is Jeongyeon Park, and I am a graduate student in the Department of Second Language Studies at the University of Hawai‘i, Mānoa. The primary purpose of this study is to explore the impact of extensive reading on ESL (English as a Second Language) students’ writing performance. The study also aims to investigate students’ reactions to and perceptions about the usefulness of ER implemented in an EAP writing course. You are being asked to participate since you have taken ELI 73 (writing for foreign students) at UHM.

Activities and Time Commitment: Participation in this study includes giving consent to the researcher to use your writing work (e.g., the 10-minute writing) submitted as a class assignment in ELI 73. In addition, you will be asked whether you are interested in participating in a voluntary interview. The interview questions will particularly ask you about your feelings about extensive reading experience across one semester. The interviews should take no more than one hour. With your permission, I would also like to record the interview using a digital audio-recorder for the transcript purpose; however, your name will not be revealed in any of the data reporting.

Benefits and Risks: There will be no direct benefits to you in participating in this study. However, some interview questions will provide you opportunities to think about your feelings toward reading and writing in English as an ESL student. Moreover, the results of this study will contribute to a better understanding of the effectiveness of extensive reading in improving English reading and writing, which possibly lead to a modification of the course to benefit future student writers in ELI 73. I do not anticipate any risks to you participating in this study.

Privacy and Confidentiality: All data from your writing work, interview recordings, and transcripts will be kept in a locked file in the researcher’s office for the duration of the study. The data will be labeled using pseudonyms for analysis and will only be used for the purpose of this study. All personal information and the audio recordings will be destroyed upon completion of the research project.

Voluntary Participation: Participation in this research project is completely voluntary. There is no consequence or penalty for not allowing the researcher to use your writing. In addition, you are free to withdraw your participation at any time without penalty or loss of benefits.

Questions: If you have any questions regarding this study, please contact me via phone at (808) 397-3505 or via email me at pj22@hawaii.edu. If you have any questions concerning your rights as a research participant, you can contact the UH Committee on Human Studies at 808.956.5007 or uhirb@hawaii.edu.

If you agree to participate in this project, please sign the following signature portion of this consent form and return it to Jeongyeon Park at pj22@hawaii.edu.
Signatures for Consent:

I agree to participate in this project and I give consent to the researcher to use my writing work. I understand that I can change my mind at any time, by notifying the researcher.

Name (print) _______________________________
Signature ________________________________ Date _____________

In addition, I agree to participate in audio-recording interviews. I give my permission for the researcher to contact me via email.

Name (print) _______________________________
Email Address ________________________________
Signature ________________________________ Date _____________
APPENDIX E (continued): Consent Forms

Consent Form for the Control Group:

Implementing Extensive Reading in a University EAP Writing Class

My name is Jeongyeon Park, and I am a graduate student in the Department of Second Language Studies at the University of Hawai‘i, Mānoa. The primary purpose of this study is to explore the impact of extensive reading on ESL (English as a Second Language) students’ writing performance. The study also aims to investigate students’ reactions to and perceptions about the usefulness of ER implemented in an EAP writing course. You are being asked to participate since you have taken ELI 73 (writing for foreign students) at UHM.

Activities and Time Commitment: Participation in this study includes giving consent to the researcher to use your writing work (e.g., freewriting) submitted as a class assignment in ELI 73. In addition, you will be asked whether you are interested in participating in a voluntary interview. The interview questions will particularly ask you about your feelings about ELI 73 as well as about your attitude toward and experiences with reading and writing. The interviews should take no more than one hour. With your permission, I would also like to record the interview using a digital audio-recorder for the transcript purpose; however, your name will not be revealed in any of the data reporting.

Benefits and Risks: There will be no direct benefits to you in participating in this study. However, some interview questions will provide you opportunities to think about your feelings toward reading and writing in English as an ESL student. Moreover, the results of this study will possibly lead to a modification of the course to benefit future student writers in ELI 73. I do not anticipate any risks to you participating in this study.

Privacy and Confidentiality: All data from your writing work, interview recordings, and transcripts will be kept in a locked file in the researcher’s office for the duration of the study. The data will be labeled using pseudonyms for analysis and will only be used for the purpose of this study. All personal information and the audio recordings will be destroyed upon completion of the research project.

Voluntary Participation: Participation in this research project is completely voluntary. There is no consequence or penalty for not allowing the researcher to use your writing. In addition, you are free to withdraw your participation at any time without penalty or loss of benefits.

Questions: If you have any questions regarding this study, please contact me via phone at (808) 397-3505 or via email me at pj22@hawaii.edu. If you have any questions concerning your rights as a research participant, you can contact the UH Committee on Human Studies at 808.956.5007 or uhirb@hawaii.edu.

If you agree to participate in this project, please sign the following signature portion of this consent form and return it to Jeongyeon Park at pj22@hawaii.edu.
Signatures for Consent:

I agree to participate in this project and I give consent to the researcher to use my writing work. I understand that I can change my mind at any time, by notifying the researcher.

Name (print)  ____________________________
Signature  ____________________________  Date  ____________

In addition, I agree to participate in audio-recording interviews. I give my permission for the researcher to contact me via email.

Name (print)  ____________________________
Email Address  ____________________________
Signature  ____________________________  Date  ____________
## Appendix F: Suggested Topics for the 10-minute Writing

<table>
<thead>
<tr>
<th>Summarizing:</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Summarize the story so far, including the main themes, events, and characters.</td>
</tr>
<tr>
<td>• Based on what you have read so far, can you guess what will happen next? You can briefly summarize the story first and then develop your ideas from there.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Describing:</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Choose characters in the book that you either liked or disliked, and explain why. You can describe their personality or behavior in relation to some events in the story.</td>
</tr>
<tr>
<td>• What was the most interesting/important event in the story? Please describe.</td>
</tr>
<tr>
<td>• Describe one character that played an important role in the development of the story. Explain how and why.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Creative Writing:</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Write new or different endings for the story. If you were the author, how would you end the story? Why?</td>
</tr>
<tr>
<td>• If you could create a character or a thing to make the story better or more interesting, who or what would that be? And why?</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Analyzing:</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Please choose one character in the story. Then, think of one gift that you want to give that character, and explain briefly why you have chosen that particular gift.</td>
</tr>
<tr>
<td>• Is there any part of the story (e.g., events, characters, plots) that you don’t like? How would you like to change it?</td>
</tr>
<tr>
<td>• If you could give some advice/compliments/encouragement to one of the characters in the story, what would that be? Why?</td>
</tr>
<tr>
<td>• What kinds of information did you learn from the book? Was it totally new to you? Or did you have some background knowledge about the topic?</td>
</tr>
<tr>
<td>• Describe one of the most important events in the story and explain how the main character dealt with or reacted to that specific incident.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Evaluating:</th>
</tr>
</thead>
<tbody>
<tr>
<td>• What were the best and the worst things about the book?</td>
</tr>
<tr>
<td>• If you were to write a new book title, what would that be? Why?</td>
</tr>
<tr>
<td>• Has any of the characters made an important decision in the story? Do you support his or her decision? If YES, explain how that decision has affected the plot (events of story). If NO, what kinds of decision would you make if you were the character?</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Comparing and Contrasting:</th>
</tr>
</thead>
<tbody>
<tr>
<td>• If you already know the ending, did the story end as you expected?</td>
</tr>
<tr>
<td>• If you could change any part of the story (events, characters, etc.), how would you like to change it? And why?</td>
</tr>
<tr>
<td>• Did the book provide the information that you expected? To what extent is the provided information similar or different from your knowledge about the topic?</td>
</tr>
<tr>
<td>• Did you find any interesting cultural information in the book? Is it different or similar to your culture?</td>
</tr>
<tr>
<td>Other:</td>
</tr>
<tr>
<td>---------------------------------------------------------------------</td>
</tr>
<tr>
<td>• Did you learn any new words, phrases, or structures from this week’s reading? Explain the context and provide some new examples of your own.</td>
</tr>
<tr>
<td>• What is the best book you have read so far? Why? Would you recommend this book to your classmates?</td>
</tr>
</tbody>
</table>
## APPENDIX G: Book Recording Form

<table>
<thead>
<tr>
<th>Week</th>
<th>Book Title (Level)</th>
<th>Date (mm/dd)</th>
<th>Reading Time (min)</th>
<th># of Pages Read (pg)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Billy Elliot (3)</td>
<td>8/26</td>
<td>30 min</td>
<td>22 pg</td>
</tr>
<tr>
<td></td>
<td></td>
<td>8/27</td>
<td>20 min</td>
<td>30 pg</td>
</tr>
<tr>
<td></td>
<td></td>
<td>8/28</td>
<td>25 min</td>
<td>35 pg</td>
</tr>
<tr>
<td></td>
<td></td>
<td>8/29</td>
<td>25 min</td>
<td>25 pg</td>
</tr>
</tbody>
</table>

**Enjoyment:**

- **Interesting?**
- **Boring?**

- Enjoyable:
  - I like the way the main character pursued his dream, overcoming bias towards ballet dancing.

**Difficulty:**

- **Easy**
- **Comfortable**
- **Difficult**

- Comfortable
  - Vocabulary level was comfortable to read.
  - Easy to follow the plot (story).

**Short Summary**

This is about one boy who discovers his talent in ballet dancing and pursues his dream as a dancer!
APPENDIX H: Assignment Log

Please record the amount of time you spent on the course assignments, including textbook homework, worksheets, freewritings, etc.

<table>
<thead>
<tr>
<th>Week</th>
<th>HW</th>
<th>Time (min)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### APPENDIX I: Criteria for the Holistic Rating

<table>
<thead>
<tr>
<th>Score</th>
<th>Description</th>
</tr>
</thead>
</table>
| **6 Very good** | An essay at this level largely accomplishes all of the following:  
- Clear development of thesis  
- Well organized argumentation showing logical sequencing with clear transitions  
- Consistency in point of view, using effective explanations, exemplifications, and/or details  
- Wide range of vocabulary and appropriate use of idiomatic expressions with few errors  
- Demonstrates syntactic variety or sentence complexity |
| **5 Good** | An essay at this level largely accomplishes all of the following:  
- Clear development of thesis and argumentation  
- Displays topic unity, using appropriate explanations, exemplification, and/or details with a slight lack of clear connection of ideas in some areas.  
- Wide range of vocabulary but occasional errors in use of idiomatic expressions and/or word form  
- Well-formed and complete sentences with few errors |
| **4 Average or Fair** | An essay at this level may reveal one or more of the following:  
- Addresses thesis but some points are not fully elaborated  
- Generally well organized using exemplifications, and/or details but contains some redundancy, digression, or unclear connections of ideas  
- Somewhat limited range of vocabulary and/or some errors in word choice and/or word form  
- Relatively well-formed but limited range of syntactic structures and/or repetition of simple sentences |
| **3 Poor** | An essay at this level may reveal one or more of the following:  
- Underdeveloped thesis and unclear argument with only simple topic description/ restatement  
- Redundancy of ideas with a lack of supporting evidence, exemplifications, and/or details  
- No major digression from the topic but contains irrelevant specifics  
- Notably limited vocabulary and/or frequent errors in word choice and/or word form  
- Incomplete sentences or frequent errors in sentence formation |
<table>
<thead>
<tr>
<th>2</th>
<th>Very Poor</th>
<th>An essay at this level may reveal one or more of the following:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>• Serious underdevelopment in response due to lack of understanding of the topic or task</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Major digression from the topic and/or no development of thesis and argument</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• No explanations, exemplifications, or details to support thesis</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Very limited vocabulary</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Serious and numerous problems with sentence formation or word choice that interferes with comprehension</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>1</th>
<th>Incomplete</th>
<th>An essay at this level</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>• Merely copies words or quotes from the instructional material</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Shows little or no knowledge of the topic or task</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Is not enough to evaluate and/or insufficient amount of essay</td>
</tr>
</tbody>
</table>
# APPENDIX J: Criteria for Analytic Rating

<table>
<thead>
<tr>
<th>SCORE</th>
<th>LEVEL</th>
<th>CRITERIA</th>
</tr>
</thead>
</table>
| 20-18 | Excellent to Very Good | a. Knowledgeable  
b. Substantive  
c. Thorough development of thesis  
d. Relevant to assigned topic |
| 17-14 | Good to Average | a. Some knowledge of subject  
b. Adequate range  
c. Limited development of thesis  
d. Mostly relevant to topic, but lacks detail |
| 13-10 | Fair to Poor    | a. Limited knowledge of subject  
b. Little substance  
c. Inadequate development of topic |
| 9-7   | Very Poor       | a. Does not show knowledge of subject  
b. Non-substantive  
c. Not pertinent  
d. Or not enough to evaluate |
|       |                | a. Fluent expression  
b. Ideas clearly stated/supported  
c. Succinct  
d. Well-organized  
e. Logical sequencing  
f. Cohesive |
| 20-18 | Excellent to Very Good | a. Somewhat choppy  
b. Loosely organized but main ideas stand out  
c. Limited support  
d. Logical but incomplete sequencing |
| 17-14 | Good to Average | a. Non-fluent  
b. Ideas confused or disconnected  
c. Lacks logical sequencing and development |
| 13-10 | Fair to Poor    | a. Does not communicate  
b. No organization  
c. Or not enough to evaluate |
| 9-7   | Very Poor       | a. Sophisticated range  
b. Effective word/idiom choice and usage  
c. Word form mastery  
d. Appropriate register |
| 20-18 | Excellent to Very Good | a. Meaning not obscured  
b. Adequate range  
c. Occasional errors of word/idiom form, choice, usage |
| 17-14 | Good to Average | a. Meaning confused or obscured  
b. Limited range  
c. Frequent errors of word/idiom form, choice, usage |
| Language Use          | 9-7    | Very Poor | a. Foreign words  
b. Little knowledge of English vocabulary, idioms, word form  
c. Or not enough to evaluate |
|----------------------|--------|-----------|-------------------------------------------------------------|
|                      | 20-18  | Excellent to Very Good | a. Effective complex constructions  
b. Few errors in agreement, tense, number, word order/function, articles, pronouns, prepositions |
|                      | 17-14  | Good to Average | a. Effective but simple constructions  
b. Minor problems in complex constructions  
c. Several errors of agreement, tense, number, word order/function, articles, pronouns, prepositions but meaning not obscured |
|                      | 13-10  | Fair to Poor | a. Major problems in simple/complex constructions  
b. Frequent errors of negation, agreement, tense, number, word order/function, articles, pronouns, prepositions  
c. Sentence fragments, run-ons, deletions  
d. Meaning confused or obscured |
|                      | 9-7    | Very Poor | a. Virtually no mastery of sentence construction rules  
b. Dominated by errors  
c. Does not communicate  
d. Or not enough to evaluate |
| Mechanics            | 5      | Excellent to Very Good | a. Demonstrate mastery of conventions  
b. Few errors of spelling, punctuation, and capitalization, |
|                      | 4      | Good to Average | a. Occasional errors of spelling, punctuation, but meaning not obscured |
|                      | 3      | Fair to Poor | a. Frequent errors of spelling and punctuation  
b. Meaning confused or obscured |
|                      | 2      | Very Poor | a. No mastery of conventions  
b. Dominated by errors of spelling, punctuation, capitalization,  
c. Or not enough to evaluate |
APPENDIX J (continued): The Extended Profile Criteria

### Content

<table>
<thead>
<tr>
<th>Descriptor</th>
<th>Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Knowledgeable</strong></td>
<td>Is there understanding of the subject? Are facts or other pertinent information used? Is there recognition of several aspects of the subject? Are the interrelationships of these aspects shown?</td>
</tr>
<tr>
<td><strong>Substantive</strong></td>
<td>Are several main points discussed? Is there sufficient detail? Is there originality with concrete details to illustrate, define, compare, or contrast factual information supporting the thesis?</td>
</tr>
<tr>
<td><strong>Thorough development of thesis</strong></td>
<td>Is the thesis expanded enough to convey a sense of completeness? Is there a specific method of development (such as comparison/contrast, illustration, definition, example, description, fact, or personal experience)?</td>
</tr>
<tr>
<td><strong>Relevant to assigned topic</strong></td>
<td>Is all information clearly pertinent to the topic? Is extraneous material excluded?</td>
</tr>
</tbody>
</table>

### Organization

<table>
<thead>
<tr>
<th>Descriptor</th>
<th>Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>** Fluent Expression**</td>
<td>Do the ideas flow, building on one another? Are there introductory and concluding paragraphs? Are there effective transition elements—words, phrases, or sentences—which link and move ideas both within and between paragraphs?</td>
</tr>
<tr>
<td><strong>Ideas clearly stated/supported</strong></td>
<td>Is there a clearly stated controlling idea or central focus to the paper (a thesis)? Do topic sentences in each paragraph support, limit, and direct the thesis?</td>
</tr>
<tr>
<td><strong>Succinct</strong></td>
<td>Are all ideas directed concisely to the central focus of the paper, without digression?</td>
</tr>
<tr>
<td><strong>Well-organized</strong></td>
<td>Is the overall relationship of ideas within and between paragraphs clearly indicated? Is there a beginning, a middle, and an end to the paper?</td>
</tr>
<tr>
<td><strong>Logical sequencing</strong></td>
<td>Are the points logically developed? Is this development indicated by appropriate transitional markers?</td>
</tr>
<tr>
<td><strong>Cohesive</strong></td>
<td>Does each paragraph reflect a single purpose? Do the paragraphs form a unified paper?</td>
</tr>
</tbody>
</table>
### Vocabulary

<table>
<thead>
<tr>
<th>Descriptor</th>
<th>Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sophisticated range</strong></td>
<td>Is there facility with words and idioms: to convey intended information, attitudes, feelings? to distinguish subtleties among ideas and intentions? to convey shades and differences of meaning? to express the logic of ideas? Is the arrangement and interrelationship of words sufficiently varied?</td>
</tr>
<tr>
<td><strong>Effective word/idiom choice and usage</strong></td>
<td>In the context in which it is used, is the choice of vocabulary accurate? idiomatic? effective? concise? Are strong active verbs and verbals used where possible? Are phrasal and prepositional idioms correct? Do they convey the intended message? emphasis? antonyms? homonyms? Are denotative and connotative meanings distinguished? Is there effective repetition of key words and phrases? Do transition elements mark shifts in thought? pace? emphasis? tone?</td>
</tr>
<tr>
<td><strong>Word form mastery</strong></td>
<td>Are prefixes, suffixes, roots, and compounds used accurately and effectively? Are words correctly distinguished as to their function (noun, verb, adjective, adverb)?</td>
</tr>
<tr>
<td><strong>Appropriate register</strong></td>
<td>Is the vocabulary appropriate to the topic? to the audience? To the tone of the paper? To the method of development? Is the vocabulary familiar to the audience? Does the vocabulary make the intended impression?</td>
</tr>
</tbody>
</table>

### Language Use

<table>
<thead>
<tr>
<th>Descriptor</th>
<th>Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Effective complex constructions</strong></td>
<td>Are sentences well-formed and complete, with appropriate complements? Are single-word modifiers appropriate to function? Are they properly formed, placed, sequenced? Are phrases and clauses appropriate to function? Complete? Properly placed? Are introductory <em>It</em> and <em>There</em> used correctly to begin sentences and clauses? Are main and subordinate ideas carefully distinguished? Are coordinate and subordinate elements linked to other elements with appropriate conjunctions, adverbials, relative pronouns, or punctuation? Are sentence types and length varied? Are elements parallel? Are techniques of substitution, repetition, and deletion use effectively?</td>
</tr>
<tr>
<td><strong>Agreement</strong></td>
<td>Is there basic agreement between sentence elements: auxiliary and verb? subject and verb? pronoun and antecedent? adjective and noun? nouns and quantifiers?</td>
</tr>
<tr>
<td><strong>Tense</strong></td>
<td>Are verb tenses correct? properly sequenced? Do modals convey intended meaning? time?</td>
</tr>
<tr>
<td><strong>Number</strong></td>
<td>Do nouns, pronouns, and verbs convey intended quality?</td>
</tr>
<tr>
<td><strong>Word order/function</strong></td>
<td>Is normal word order followed except for special emphasis? Is each word, phrase, and clause suited to its intended function?</td>
</tr>
<tr>
<td><strong>Articles</strong></td>
<td>Are <em>a</em>, <em>an</em>, and <em>the</em> used correctly?</td>
</tr>
<tr>
<td><strong>Pronouns</strong></td>
<td>Do pronouns reflect appropriate person? gender? number? function? referent?</td>
</tr>
<tr>
<td><strong>Prepositions</strong></td>
<td>Are prepositions chosen carefully to introduce modifying elements? Is the intended meaning conveyed?</td>
</tr>
</tbody>
</table>

**Mechanics**

<table>
<thead>
<tr>
<th><strong>Descriptor</strong></th>
<th><strong>Criteria</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Spelling</strong></td>
<td>Are words spelled correctly?</td>
</tr>
<tr>
<td><strong>Punctuation</strong></td>
<td>Are periods, commas, semicolons, dashes, and question marks used correctly? Are words divided correctly at the end of lines?</td>
</tr>
<tr>
<td><strong>Capitalization</strong></td>
<td>Are capital letters used where necessary and appropriate?</td>
</tr>
</tbody>
</table>
APPENDIX K: Guidelines for Objective Measures

These guidelines are from Polio (1997). The italicized parts are new additions, which are included for the purpose of data analysis in the present study.

**T-Units**

a. A T-unit is defined an independent clause and all its dependent clauses.

b. Count run-on sentences and comma splices as two T-units with an error in the first T-unit.

   e.g.,) My school was in Saudi Arabia, it was the best school there.

   T / T
   1 error               error-free

If several comma-splices occur in a row, count only the last as error free.

c. For sentence fragments, if the verb or copula is missing, count the sentence as 1 T-unit with an error. If an NP is standing alone, attach it to the preceding or following T-unit as appropriate and count as an error. If a subordinate clause is standing alone, attach it to the preceding or following sentence and count it as 1 T-unit with an error.

   e.g.,) I like Hawaii. Because it has beautiful nature. = 1 T-unit with an error

d. When there is a grammatical subject deletion in a coordinate clause, count the entire sentence as 1 T-unit.

   e.g.,) First we went to our school and then went out with our friends = 1 T-unit

e. Count both “so” and “but” as coordinating conjunctions. Count “so that” as a subordinating conjunction unless “so” is obviously meant.

f. Do not count tag-questions as separate T-units.

g. Count S-nodes with a deleted complementizer as a subordinate clause as in: I believe that A and (that) B = 1 T-unit

h. But, direct quotes should be counted as:

   John said, “A and B.”

   1 T-unit          1-T-unit:

   A direct quote should be counted as 1 T-unit.

i. Count T-units in parentheses as individual T-units.
Clauses

a. A clause equals an overt subject and a finite verb.
   He left the house and drove away. = 1 clause
   He wanted John to leave the house. = 1 clause
   He went to the bookstore and he bought one magazine. = 2 clauses
   He went to the bookstore and he bought one magazine that he needed. = 3 clauses

b. Only an imperative does not require a subject to be considered a clause.

c. In a sentence that has a subject with only an auxiliary verb, do not count that subject and verb
   as a separate clause (or as a separate T-unit).
      e.g.,) John likes to ski and Mary does too; John likes to ski, doesn’t he?; John is happy and
      Mary is too)

d. A direct quote should be counted as 1 clause.

Sentences

a. Group of words punctuated by the writer (Hunt, 1965).

b. Count a direct quote as one sentence.

Error Guidelines (when counting Error free T-units or Error free clauses)

a. Do not count spelling errors (including word changes like “there/ their”).

b. Be conservative about counting comma errors; don’t count missing commas between clauses
   or after prepositional phrase. Comma errors related to restrictive/ non-restrictive relative clauses
   should be counted.

c. Base tense/reference errors on preceding discourse; do not look at the sentences in isolation.

d. Don’t count British usages as errors (e.g. “in hospital,” “at university,” collective nouns as
   plural).

e. Count article errors.

f. Don’t count errors in capitalization.

g. Count errors that could be made by native speakers (e.g. between you and I)

h. Disregard an unfinished sentence at the end of the essay.

i. Do not count words mistakenly divided or undivided (e.g., them selves, Incontrast.)

j. Do not count errors (e.g., spelling errors, years of publications) within the quote.
k. Do not count non-indentted paragraphs as errors.
l. Count errors in a hyphenated word (e.g., well-write = 2 words and 1 error).
APPENDIX L: The 10-minute Writing Samples

Sample 1
Topic: What kinds of lessons did you learn from the story?
Book title: Wonder

The most important lesson I learned from the book entitled “Wonder” is that trying to jump into totally new circumstances makes the person more growing. In this book, main character named August, who has troubles with his face for some reasons, decided to go a mainstream school. Before semester starts, he and his mother went to the school for school tour and he met three kids who was going to be his classmates. August was afraid of getting to know other children because he knows that every child feel scary his extraordinary face. As expected, one of the kids he met at the day teased him about his face. Others, however, didn’t. After he finished the tour, he felt happy a little bit because he met some kids who may be good friends for him.

Sample 2
Topic: If you were to write a new book title, what would that be? Why?
Book title: Crying for Freedom

If I were to write a new title, I would name it “Life for Freedom.” I was interested to learn more about the life of main hero, as the story is based on real events. I was surprised to learn from Wikipedia that the main hero Biko was only 30 years old, when he leaded the resistance movement. I was impressed how this young adult captured the hearts of thousands people. Most adults in his years occupied with physical assets (homes, cars, bank account, etc) and at the end of the life ask themselves what is real purpose in life? Biko chose “Life for Freedom.” He exchanged his life for dreams of thousands South Africans. He was killed at the age of 31 year. Biko paid expensive and terrible price. Finally, he raised the question for all of us. What is our purpose in life? And what price we gonna pay to change this world?
Sample 3

Topic: Please choose one character in the story. Then, think of one gift that you want to give that character, and explain briefly why you have chose that particular gift.

Book title: *Prince William and Kate Middleton*

I want to give a letter from Diana to Prince William. He lost his mother when he was 13. He could spend much time with his mother in his early childhood. Even though Diana tried to share a lot portion of her life with his son, but it was tough thing that losing one’s mother. The letter that I wan to give to Prince William is about encourage his life as a Prince, father, and a son. Diana must have wanted to way lots of lesson of life to his young son. If she can write a letter to his son, she will be proud of his son. Prince William got married beautiful woman, Kate Middleton. Diana might want to say about his role as a father, too. He has a son, and he is waiting for another baby. I think this letter which contains Diana’s lesson for his son will be a precious gift for Prince William.

Book title: *Michael Jackson*

Michael Jackson was very so he could buy everything for himself. The one gift I can think of to give him would be some friends because he was really lonely in the story. He had managers instead of friends. I think the main reason for him being lonely and having a difficult time to make friends is that he didn’t really get a normal childhood. He had to work hard always, whether it was on school or the Jackson 5 band with his brothers. His father made him work too hard and treated him violently regularly. Michael was very talented, but that hard work really paid off, although he was damaged personally for life. I don’t know if Michael would be called the King of Pop today if it wouldn’t have been for his father, Joe. I think Michael could have been a successful pop artist without his father, but not as successful as with his father’s “help.” But Michael would have a more normal life and would maybe not have turned out as weird, he would also have an easier time making friends. I change my mind. I would not give him friends, I would want to give him the ability to make friends and trust them.
REFERENCES


http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Reading-writing+connections:+toward+a+description+for+second+language+learners#0


*Educational Psychologist, 35*(1), 39–50.


Laufer, B. (1997). What's in a word that makes it hard or easy: Some intralexical factors that affect the learning of words. In N. Schmitt & M. McCarthy (Eds.), *Vocabulary:*


*Reading in a Foreign Language, 18*(1), 1–28. Retrieved from

http://nflrc.hawaii.edu/rfl/april2006/pigada/pigada.html


*Reading in a Foreign language, 5*(2), 271–275.


doi:10.1016/j.jslw.2012.09.004


