DEVELOPMENT OF READING FLUENCY DURING AN EXTENSIVE READING COURSE INCORPORATING READING FLUENCY TASKS

A Dissertation
Submitted
to the Temple University Graduate Board

In Partial Fulfillment
of the Requirements for the Degree of
Doctor of Philosophy in Education

By
Jeffrey Huffman
May, 2021

Examining Committee Members

David Beglar, Advisory Chair, Teaching and Learning
Tomoko Nemoto, Teaching and Learning
Katerina Petchko, External Member, National Graduate Institute for Policy Studies
Lance Burrows, External Member, Kinki University
ABSTRACT

Reading approaches in Japan and other EFL settings have typically focused on accuracy at the expense of fluency. However, reading rate and fluency are important components of reading that are currently being neglected. A review of the literature reveals growing interest in investigating the effectiveness of reading rate and fluency enhancement approaches such as extensive reading (ER), speed reading, and oral fluency training. Results of previous research indicate the effectiveness of these approaches for increasing reading speed and fluency as well as comprehension, but the overall body of evidence in this area remains weak, particularly in EFL contexts.

With these gaps in mind, the purposes of this study were to investigate the effectiveness of the ER approach in developing reading fluency, investigate appropriate ways to measure reading fluency in adult EFL readers, investigate whether reading fluency increases are accompanied by reading comprehension increases, and identify the patterns of reading rate development that occur during an ER course.

To this end, a longitudinal quasi-experimental study was conducted with 77 first-year students at a nursing college in Japan, all of whom participated in one 15-week semester of ER with an added fluency training component (ERFT) and one 15-week semester of intensive reading (IR). Silent reading rate, oral reading rate and fluency, reading comprehension, and written receptive vocabulary size were measured at the beginning and end of both semesters. Silent reading rate, oral reading rate, oral reading fluency, and reading comprehension were measured using both simplified and authentic texts. Approximately half of the students took ERFT in the first semester and
IR in the second semester, and the other half did the opposite, so analyses were conducted separately for these two groups.

Repeated-measures ANOVAs showed that for the IR-before-ERFT group, oral reading rate increase was significantly greater for the ERFT treatment (9.79 wcpm) than the IR treatment (3.05 wcpm) \((p < .001)\) for simplified texts; silent reading rate increase was significantly greater for the IR treatment (8.26 wpm) than the ERFT treatment (-2.86 wpm) \((p < .001)\) for authentic texts; and oral reading rate increase was significantly greater for ERFT (15.18 wcpm) than IR (4.37 wcpm) \((p < .001)\) for authentic texts. For the ERFT-before-IR group, silent reading rate increase was significantly greater for ERFT (19.22 wpm) than IR (-2.50 wpm) \((p < .001)\) for simplified texts; oral reading rate increase was significantly greater for ERFT (12.55 wcpm) than IR (4.81 wcpm) \((p < .001)\) for simplified texts; silent reading rate increase was significantly greater for ERFT (21.86 wpm) than IR (-7.64 wpm) \((p < .001)\) for authentic texts; and oral reading rate increase was significantly greater for IR (13.06 wcpm) than ERFT (6.70 wcpm) \((p < .001)\) for authentic texts.

Repeated-measures ANOVAs also showed that for the IR-before-ERFT group, reading comprehension increase for authentic texts was significantly greater for IR (1.00 logits) than ERFT (.17 logits) \((p = .010)\); and vocabulary size increase was significantly greater for IR (.36 logits) than ERFT (-.19 logits) \((p = .001)\). For the ERFT-before-IR group, reading comprehension increase for authentic texts was significantly greater for ERFT (1.05 logits) than IR (.11 logits) \((p = .003)\); and vocabulary size increase was significantly greater for ERFT (.44 logits) than IR (-.07 logits) \((p = .001)\).
No significant correlations were found between changes in any of the reading rate and fluency measures (silent reading rate, oral reading rate, and oral reading fluency) and reading comprehension change. For the ERFT-before-IR group, analysis revealed a significant and moderately high negative correlation ($r = -0.422$) between initial silent reading rate and silent reading rate change (on simplified texts) during the ERFT treatment. Analysis of individual students’ silent reading rate change patterns during the ERFT semester revealed a jagged increase to be the most common (22/77; 28.57%) pattern of rate increase.

These results add to a growing body of evidence that ERFT yields greater reading fluency gains than IR, suggest that both ERFT and IR contribute to reading comprehension and vocabulary size gains, suggest that ERFT yields greater reading fluency gains among students with initially slower reading rates, and provide insight into how individual’s reading rate changes over the course of a semester of ERFT.
ACKNOWLEDGMENTS

I would like to thank a number of individuals for their support, encouragement, and patience throughout the process of planning and conducting this study and writing this dissertation.

First and foremost, I would like to offer my unending gratefulness to Dr. David Beglar, my advisor for this study. He has been a source of information, knowledge, wisdom, encouragement, inspiration, and guidance, both during my dissertation project and long before, indeed for the entirety of my Master’s and doctoral studies. Being blessed with such a long-term, reliable, unwavering mentor and role model, someone who always seems to find the right mix of patience on the one hand and great expectations on the other, has been a godsend to me—I almost certainly would not have been able to complete this study without him.

Next I would like to thank my examining committee members, Dr. Katerina Petchko, Dr. Lance Burrows, and Dr. Tomoko Nemoto. Dr. Petchko and Dr. Burrows offered invaluable advice at the proposal stage that guided the design and methodology of the study during its infancy, and all three gave generously of their time, knowledge, and experience at the oral defense stage as they asked the challenging questions and provided the constructive feedback that helped me improve the quality of the final version of this dissertation.

I also want to offer my appreciation to two colleagues who have supported and encouraged me throughout the process of researching, conducting, and writing this dissertation. Marsha Clark taught me how to teach at the very beginning of my teaching
career, has pushed me in the right direction and opened doors for me at key points in my career, and has frequently been a source of good advice and friendship. Mami Inoue, in her capacity as both colleague and friend, has been a tireless partner in education and research while also serving as a trusted confidant in all aspects of life.

Finally, I wish to thank, from the bottom of my heart, my loving and supportive family. Both near and far away, without their daily and undying support, love, patience, and understanding, I truly could not have made it through this academic and personal journey.
DEDICATION

To RK, who loves, challenges, forgives, and brings joy, all in good measure
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>PAGE</td>
</tr>
<tr>
<td>iii</td>
</tr>
<tr>
<td>vi</td>
</tr>
<tr>
<td>viii</td>
</tr>
<tr>
<td>xiii</td>
</tr>
<tr>
<td>xviii</td>
</tr>
</tbody>
</table>

## CHAPTER

1. INTRODUCTION ................................................................. 1
   - The Background of the Issue ........................................... 1
   - Statement of the Problem ............................................. 5
   - Purposes of the Study .................................................. 7
   - Significance of the Study ............................................. 8
   - Theoretical Perspective .............................................. 10
   - The Audience for the Study ......................................... 13
   - Delimitations ........................................................... 14
   - Organization of the Study .......................................... 16

2. REVIEW OF THE LITERATURE ........................................... 17
   - Reading Fluency ....................................................... 17
   - Text Difficulty, Readability, and Simplification ............... 20
   - Extensive Reading and Reading Fluency Training ............. 21
   - Extensive Reading and Reading Rate Gains ..................... 27
   - Intensive Reading .................................................... 35
Oral Reading Prosody Ratings (Time 1, Simplified) ................. 99
Oral Reading Prosody Ratings (Time 2, Simplified) ............... 102
Oral Reading Prosody Ratings (Time 3, Simplified) ............... 104
Oral Reading Prosody Ratings (Time 4, Simplified) ............... 107
Oral Reading Prosody Ratings (Time 1, Authentic) ............... 107
Oral Reading Prosody Ratings (Time 2, Authentic) ............... 112
Oral Reading Prosody Ratings (Time 3, Authentic) ............... 113
Oral Reading Prosody Ratings (Time 4, Authentic) ............... 116
Reading Comprehension Tests ......................................................... 119
Reading Comprehension Test (Time 1) ................................. 120
Reading Comprehension Test (Time 2) ................................. 126
Reading Comprehension Test (Time 3) ................................. 130
Reading Comprehension Test (Time 4) ................................. 134
Vocabulary Size Tests ................................................................. 139
Vocabulary Size Test (Time 1) ................................................. 142
Vocabulary Size Test (Time 2) ................................................. 145
Vocabulary Size Test (Time 3) ................................................. 147
Vocabulary Size Test (Time 4) ................................................. 152
Summary of Instrument Validation .................................................. 155
5. RESULTS ....................................................................................... 158
  Preliminary Results ......................................................................... 158
  Hypothesis 1: Reading Rate and Fluency Gains ....................... 160
  Hypothesis 2: Reading Comprehension and Vocabulary Knowledge .. 164
  Hypothesis 3: The Relationship Between Silent Reading Rate, Oral Reading Rate, Oral Reading Fluency, and Reading Comprehension ... 167
<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Composition of Participant Classes and Groups</td>
<td>62</td>
</tr>
<tr>
<td>2. List of Measurements</td>
<td>64</td>
</tr>
<tr>
<td>3. Silent Reading Rate Texts</td>
<td>70</td>
</tr>
<tr>
<td>4. Oral Reading Rate and Fluency Texts</td>
<td>76</td>
</tr>
<tr>
<td>5. Oral Reading Prosody Rubric</td>
<td>78</td>
</tr>
<tr>
<td>6. Reading Comprehension Texts</td>
<td>79</td>
</tr>
<tr>
<td>7. Timeline of Measurements and Treatments</td>
<td>85</td>
</tr>
<tr>
<td>8. Rater Measurement Report for the Oral Prosody Rating (Time 1, Simplified)</td>
<td>101</td>
</tr>
<tr>
<td>9. Criteria Measurement Report for Oral Prosody Ratings (Time 1, Simplified)</td>
<td>101</td>
</tr>
<tr>
<td>10. Rater Measurement Report for Oral Prosody Ratings (Time 2, Simplified)</td>
<td>104</td>
</tr>
<tr>
<td>11. Criteria Measurement Report for Oral Prosody Ratings (Time 2, Simplified)</td>
<td>104</td>
</tr>
<tr>
<td>16. Rater Measurement Report for Oral Prosody Ratings (Time 1, Authentic)</td>
<td>111</td>
</tr>
<tr>
<td>17. Criteria Measurement Report for Oral Prosody Ratings (Time 1, Authentic)</td>
<td>111</td>
</tr>
<tr>
<td>18. Rater Measurement Report for Oral Prosody Ratings (Time 2, Authentic)</td>
<td>114</td>
</tr>
</tbody>
</table>
37. Rasch Descriptive Statistics for the Vocabulary Size Test Items (Time 4) ........153
38. English Proficiency (CASEC) ........................................................................158
39. Amount Read (Standard Words) .................................................................159
40. Time Spent Reading (Hours) .......................................................................160
41. Descriptive Statistics for Change in Silent Reading Rate, Oral Reading Rate,
    and Oral Reading Fluency During ERFT and IR Treatments, for IR-Before-
    ERFT Group, for Simplified Texts (N = 38) .............................................161
42. Descriptive Statistics for Change in Silent Reading Rate, Oral Reading Rate,
    and Oral Reading Fluency During ERFT and IR Treatments, for ERFT-
    Before-IR Group, for Simplified Texts (N = 39) ......................................162
43. Descriptive Statistics for Change in Silent Reading Rate, Oral Reading Rate,
    and Oral Reading Fluency During ERFT and IR Treatments, for IR-Before-
    ERFT Group, for Authentic Texts (N = 38) .............................................163
44. Descriptive Statistics for Change in Silent Reading Rate, Oral Reading Rate,
    and Oral Reading Fluency During ERFT and IR Treatments, for ERFT-
    Before-IR Group, for Authentic Texts (N = 39) ......................................164
45. Descriptive Statistics for Change in Reading Comprehension on Simplified
    Texts, Reading Comprehension on Authentic Texts, and Vocabulary Size
    During ERFT and IR Treatments, for IR-Before-ERFT Group (N = 38) ......165
46. Descriptive Statistics for Change in Reading Comprehension on Simplified
    Texts, Reading Comprehension on Authentic Texts, and Vocabulary Size
    During ERFT and IR Treatments, for ERFT-Before-IR Group (N = 39) ......166
47. Descriptive Statistics for the ERFT Treatment for Reading Comprehension
    Change, Silent Reading Rate Change, Oral Reading Rate Change, and Oral
    Reading Fluency Change for Simplified Texts, for IR-Before-ERFT
    Group (N = 38) ......................................................................................168
48. Pearson Correlation Coefficients for the ERFT Treatment for Reading
    Comprehension Change, Silent Reading Rate Change, Oral Reading Rate
    Change, and Oral Reading Fluency Change for Simplified Texts, for the IR-
    Before-ERFT Group (N = 38) ................................................................168
49. Descriptive Statistics for the ERFT Treatment for Reading Comprehension
    Change, Silent Reading Rate Change, Oral Reading Rate Change, and Oral
    Reading Fluency Change for Simplified Texts, for the ERFT-Before-IR
    Group (N = 39) ......................................................................................169
50. Pearson Correlation Coefficients for the ERFT Treatment for Reading Comprehension Change, Silent Reading Rate Change, Oral Reading Rate Change, and Oral Reading Fluency Change for Simplified Texts, for the ERFT-Before-IR Group ($N = 39$) ........................................................................................................ 169

51. Descriptive Statistics for the ERFT Treatment for Reading Comprehension Change, Silent Reading Rate Change, Oral Reading Rate Change, and Oral Reading Fluency Change for Authentic Texts, for IR-Before-ERFT Group ($N = 38$) ..................................................................................................... 170

52. Pearson Correlation Coefficients for the ERFT Treatment for Reading Comprehension Change, Silent Reading Rate Change, Oral Reading Rate Change, and Oral Reading Fluency Change for Authentic Texts, for IR-Before-ERFT Group ($N = 38$) .............................................................................. 170

53. Descriptive Statistics for the ERFT Treatment for Reading Comprehension Change, Silent Reading Rate Change, Oral Reading Rate Change, and Oral Reading Fluency Change for Authentic Texts, for ERFT-Before-IR Group ($N = 39$) ..................................................................................................... 171

54. Pearson Correlation Coefficients for the ERFT Treatment for Reading Comprehension Change, Silent Reading Rate Change, Oral Reading Rate Change, and Oral Reading Fluency Change for Authentic Texts, for ERFT-Before-IR Group ($N = 39$) .................................................................................... 171

55. Descriptive Statistics for the ERFT Treatment for Initial Reading Comprehension, Initial Silent Reading Rate, and Silent Reading Rate Change for Simplified Texts for IR-Before-ERFT Group ($N = 38$) .......................................................................................... 172

56. Pearson Correlation Coefficients for the ERFT Treatment for Initial Reading Comprehension, Initial Silent Reading Rate, and Silent Reading Rate Change for Simplified Texts for IR-Before-ERFT Group ($N = 38$) .......................................................................................... 172

57. Descriptive Statistics for the ERFT Treatment for Initial Reading Comprehension, Initial Silent Reading Rate, and Silent Reading Rate Change for Simplified Texts for ERFT-Before-IR Group ($N = 39$) .......................................................................................... 173

58. Pearson Correlation Coefficients for the ERFT Treatment for Initial Reading Comprehension, Initial Silent Reading Rate, and Silent Reading Rate Change for Simplified Texts for ERFT-Before-IR Group ($N = 39$) .......................................................................................... 173

59. Descriptive Statistics for the ERFT Treatment for Initial Reading Comprehension, Initial Silent Reading Rate, and Silent Reading Rate Change for Authentic Texts for IR-Before-ERFT Group ($N = 38$) .......................................................................................... 174
60. Pearson Correlation Coefficients for the ERFT Treatment for Initial Reading Comprehension, Initial Silent Reading Rate, and Silent Reading Rate Change for Authentic Texts for IR-Before-ERFT Group (N = 38) ................................... 174

61. Descriptive Statistics for the ERFT Treatment for Initial Reading Comprehension, Initial Silent Reading Rate, and Silent Reading Rate Change for Authentic Texts for ERFT-Before-IR Group (N = 39) .................................. 175

62. Pearson Correlation Coefficients for the ERFT Treatment for Reading Comprehension, Initial Silent Reading Rate, and Silent Reading Rate Change for Authentic Texts for ERFT-Before-IR Group (N = 39) ........................................... 175

63. Silent Reading Rate Change Patterns During the ERFT Treatment (N = 77) .................................................................................................................182
LIST OF FIGURES

Figure | Page
-------|------
1. Wright Map for the Oral Prosody Ratings (Time 1, Simplified) | 100
2. Wright Map for the Oral Prosody Ratings (Time 2, Simplified) | 103
3. Wright Map for the Oral Prosody Ratings (Time 3, Simplified) | 105
4. Wright Map for the Oral Prosody Ratings (Time 4, Simplified) | 108
5. Wright Map for the Oral Prosody Ratings (Time 1, Authentic) | 110
6. Wright Map for the Oral Prosody Ratings (Time 2, Authentic) | 113
7. Wright Map for the Oral Prosody Ratings (Time 3, Authentic) | 115
8. Wright Map for the Oral Prosody Ratings (Time 4, Authentic) | 117
9. Abridged Item-Person Map for the Reading Comprehension Test (Time 1) | 123
10. Abridged Item-Person Map for the Reading Comprehension Test (Time 2) | 128
11. Abridged Item-Person Map for the Reading Comprehension Test (Time 3) | 132
12. Abridged Item-Person Map for the Reading Comprehension Test (Time 4) | 137
13. Abridged Item-Person Map for the Vocabulary Size Test (Time 1) | 144
14. Abridged Item-Person Map for the Vocabulary Size Test (Time 2) | 148
15. Abridged Item-Person Map for the Vocabulary Size Test (Time 3) | 151
16. Abridged Item-Person Map for the Vocabulary Size Test (Time 4) | 154
17. Example of a “Gradual Increase” Silent Reading Rate Change Pattern (Student S5) | 177
18. Example of an “Up-Down Increase” Silent Reading Rate Change Pattern (Student S11) | 178
19. Example of a “Jagged Increase” Silent Reading Rate Change Pattern (Student S57) | 179
20. Example of a “Down-Up Increase” Silent Reading Rate Change Pattern
   (Student S16)........................................................................................................ 180

21. Example of a “Plateau Increase” Silent Reading Rate Change Pattern
   (Student S73)........................................................................................................ 181
CHAPTER 1
INTRODUCTION

The Background of the Issue

The earliest forms of written language appeared along with the beginnings of human civilization itself in ancient Sumer and Egypt over 5,000 years ago, and independently in China and Mesoamerica somewhat more recently. Throughout the entire history of civilization, therefore, the recording and transfer of information in written form has been fundamental and integral to all essential activities in society, from government and law to commerce and communication. It logically follows then, that the ability to read has been one of the most important keys to success in nearly any endeavor or profession. Reading has therefore always been the core skill in education, whether that education is available only to the elite of society or required of everyone.

The medium of written language has changed drastically throughout history, from clay tablets and tortoise shells to scrolls, then books, and now electronic media. With these changes have come increases in the dispersal and availability of written information as well as the volume of such information. In modern society, the ability to read is now expected of everyone, and success depends on being able to read, process, and synthesize vast amounts of information quickly and efficiently.

Reading and reading education play a particularly important role for learners of English as a second language. For these learners, English reading ability can unlock doors to vast stores of valuable information not available in their first language—this is true for students engaged in postgraduate academic work and even high-level
undergraduate studies. However, for primary, secondary, and most tertiary students in many non-English speaking countries, the information they need is already available to them in their first language. For these students, I would argue, English reading education takes on an even more important role. It serves as a vehicle to learn the language itself. Reading can be a tool to provide learners of English as a foreign language (EFL), who have limited exposure to English in their everyday life, with vast exposure to the language. It can also serve as a basis for studying formal aspects of the language, such as grammar and vocabulary.

The reading of English text indeed fulfills all of these roles and purposes in EFL countries such as Japan, and it has also become entrenched in the educational and testing systems such that reading ability is in many ways equated or at least used as a proxy for general academic achievement, aptitude, or even intelligence. The development of second language reading ability and reading fluency—which I argue are ultimately the same thing—therefore, is a fundamental and essential pursuit for all students at every level of education in Japan and many other EFL countries around the world.

English reading education in Japan has traditionally been focused nearly exclusively on the development of accuracy at the expense of fluency. In the traditional approach to reading instruction, the yakudoku (literally: “translation reading”) method, as well as its somewhat more meaning-focused cousin, intensive reading (IR), short texts that are far above the students’ proficiency level are used to introduce new vocabulary and grammar as well as to foster text comprehension at a high level of accuracy (Hino, 1988; Laufer-Dvorkin, 1981; Nation, 2009a; Tanaka & Stapleton, 2007). Hino described yakudoku as “the” method of teaching English in Japan, and he defined it as a reading
strategy in which the target-language text is first translated word for word into Japanese, and then the words are rearranged to match Japanese word order. He also cited nationwide surveys showing that 70-80% of Japanese teachers of English were using this method, and that 70% of students were taught to read exclusively with this method, although it should be emphasized that this information is around thirty years old. Law (1995) described yakudoku as having descended from methods of decoding classical Chinese texts, where the goal is on translating the content of the original text rather than learning the skill of reading in the target language. Gorsuch (2000) emphasized the washback effect of the university entrance exam system in Japan, which heavily relies on yakudoku-oriented and accuracy-oriented reading skills. She notes that Japanese high school teachers stick to the most efficient methods of preparing students for entrance exams, which usually involves translating passages of texts, correcting their translations, and studying vocabulary receptively. Nishino (2012) juxtaposes yakudoku with communicative language teaching, noting that although the national university entrance exam (“Center Test”) has reduced its focus yakudoku skills, most individual university exams still focus on receptive translation skills. Law (1995) called for a move toward “communicative reading methodologies” and “a significant improvement in reading speed” (p. 221) as part of the move toward more communicative approaches.

Intensive reading (IR) is an approach described by a number of EFL reading researchers, often in contrast to extensive reading. It does not necessarily involve translation into the L1 as in yakudoku, but it shares the focus on reading short, high-level text and focusing on comprehension at a high level of accuracy, as well as usually involving heavy vocabulary and grammar study. Laufer-Dvorkin (1981) described IR as
intensive study of one to three pages of text per 90-minute lesson, focusing on vocabulary and syntax. Tanaka and Stapleton (2007) described IR in Japan as “concentrating on the teaching of grammar and vocabulary in relatively short texts via post-reading exercises, while often ending up demanding a translation of the passage into Japanese” (p. 116). My own review of five reading textbooks approved by the Japan Ministry of Education, Culture, Sports, Science and Technology (MEXT) for use in high schools—Vivid, Orbit, Power On, Element, and New Legend—revealed that they include numerous post-reading exercises designed to encourage and check detailed and global comprehension of the passage, as well as vocabulary and grammar exercises and varying amounts of translation into the L1. Yamashita (2015) contrasted the focus on learning new linguistic knowledge that occurs during IR, which she defined as “studying a difficult text that includes a substantial amount of unfamiliar vocabulary and grammar,” (p. 170) with the rauding—reading quickly and fluently and understanding the text with little difficulty—that occurs during extensive reading. Nation (2009a) described IR in terms of its focus on comprehension, vocabulary, grammar, and reading strategies, and Richards and Schmidt (2002) described IR in contrast to extensive reading, noting that it involves reading at a slower speed and with a higher degree of understanding. Whether yakudoku, IR, or some combination of the two, under these approaches students are exposed to a small amount of text, and they rarely if ever engage in long periods of uninterrupted reading.

In the past 20 years or so, more teachers have been implementing a complementary approach, known as extensive reading (ER), which is designed to expose students to much larger amounts of text while also increasing their reading rate and overall fluency by having them read texts matched to their own reading proficiency level.
There is an active and growing area of research indicating that this approach improves students’ attitudes toward reading due to ease of reading and a sense of achievement, as well as the fact that they self-select their texts from a library (Arnold, 2009; Judge, 2011; Nishino, 2007; Takase, 2008).

**Statement of the Problem**

The first problem addressed in this study, then, deals with which types of instructional approaches or learning environments are most likely to develop students’ reading fluency. Specifically, in this study I sought to clarify whether ER or IR is more effective in improving reading fluency. A few studies have indicated that reading rate gains are greater with the ER approach (Beglar et al., 2012; Huffman, 2014), but few researchers have looked at the longitudinal development of reading fluency, a multifaceted construct that consists of accuracy and prosody in addition to rate, under different reading approaches, particularly when it comes to adolescent and adult learners in EFL environments. Thus, there is not yet sufficient evidence to warrant a strong claim that ER leads to greater reading fluency development (as opposed to reading rate increases) than traditional approaches.

The second problem to be addressed was similar to the first, but with regards to how the ER and IR approaches affect reading comprehension ability and vocabulary acquisition differently. The IR approaches address these important aspects of reading ability in a straightforward way: logically speaking, texts that are at a difficulty level that is higher than the readers’ ability level would require students to improve their comprehension ability and vocabulary knowledge in order to read them. However, ER
educators and researchers have proposed that students must be exposed to larger amounts of comprehensible input in order to improve their implicit vocabulary knowledge, and also that reading speed and fluency are intimately and inseparably connected to reading comprehension. Therefore, the extent to which ER and IR affect reading comprehension and vocabulary acquisition is a matter of ongoing interest and debate.

The third problem addressed in this study concerns how well different ways of measuring reading rate correlate with reading comprehension. Studies on native-English speaking children reading orally in their first language (L1) as well as EFL students reading silently in their second language (L2) indicate a strong link between both silent and oral reading rate and reading comprehension (Fuchs et al., 2001; Breznitz, 1987, 1988; Jiang et al., 2012), but there are still few studies on this issue in EFL environments, especially in longitudinal contexts, so the claim is still weak for this population. This situation raises the question, particularly for researchers, of whether oral or silent reading should be used to measure reading rate in EFL settings. The problem to be addressed here, therefore, is the lack of empirical data on the extent to which silent and oral reading rates correlate with each other and with reading comprehension for EFL learners.

A final problem addressed by this study is that there is currently little data available on long-term patterns of reading rate development, for example over the course of a semester or a year, for individual students in an ER program. It would be useful to know whether the pattern is highly dependent on individual differences initial reading rate and also whether reading rate develops consistently over the course of a semester for most students or whether it plateaus at a certain point.
Purposes of the Study

In light of the problems discussed above, the purposes of this study were to provide insight into best-practice pedagogical approaches to developing the different aspects of reading fluency—rate, fluency, comprehension, and vocabulary knowledge—in the context of adult EFL readers, to carefully examine the construct of reading fluency and how best to measure it in this population, and to identify patterns in reading fluency development over the course of an academic year.

The first purpose of this study was to further investigate whether ER is more effective in improving reading fluency than traditional approaches to reading instruction such as IR. While the form-focused nature of IR likely remains useful for the explicit learning of vocabulary and grammatical forms in EFL settings as well as for the teaching of basic reading skills to low-proficiency students, a small but growing body of research indicates that ER is more effective in developing reading fluency (Beglar et al., 2012; Huffman, 2014). The first purpose of this study was to put that claim to the test.

The second purpose was to shed light on which reading approaches aid in vocabulary acquisition and improving reading comprehension. Both of these are explicitly targeted, although not extensively, in the IR approach. However, there have been many studies linking ER to gains in reading comprehension and/or proficiency (Elley, 1991; Pichette, 2005; Suk, 2017). The studies that have investigated vocabulary acquisition in ER contexts have shown that it happens slowly, at least in terms of quantity of new vocabulary learned.

In service to the first two purposes as well as the overall general purpose of this study, the third purpose was to thoroughly examine the construct of reading fluency itself
in order to determine which instruments and measurements are appropriate for investigating reading fluency changes in EFL adolescents and adults. Specifically, this includes the question of whether the silent or oral reading mode gives a more accurate picture of EFL adult readers’ underlying reading ability and comprehension. It is possible that different instruments to measure reading fluency are more appropriate for L2 readers and L1 readers.

The final purpose of this study was to investigate and identify patterns in reading rate development that occur during an ER course. The expectation was that the development of reading rate for a given student over the entire course of their English-learning journey would be consistent with the power law of practice (DeKeyser, 1997; Logan, 1988); however, because the data in this study only provide a picture of a small window of the full learning curve, the reading rate change seen during the semester would include ups and downs but see an overall moderate increase. It was also hypothesized that readers whose rates were lower at the initial measurement would experience greater gains than those who had already reached a higher point on their learning curve before entering the treatment period.

**Significance of the Study**

This study provides empirical data regarding to what extent and in what ways the ER and IR approaches are or are not effective, as well as insights into the nature of reading fluency and how it develops, particularly in adult EFL readers.

In the past 20 years, a great deal of research has been devoted to investigating the effectiveness of ER in terms of reading proficiency, vocabulary acquisition, motivation,
and reading rate. However, in a recent survey by the National Reading Panel in the United States, which incorporated the strictest of standards for empirical evidence, no strong evidence was found for the claims of ER’s effectiveness (National Reading Panel, 2000). The current study was quasi-experimental and therefore would not have been included in that survey, but it has the potential to shed light on whether ER should be included as an effective complementary component of reading programs in EFL countries such as Japan. If the results support the hypotheses in this study, it would add weight to the argument that Japanese English programs need to focus more on reading fluency development as opposed to simply using English text as a basis for form-focused instruction. Furthermore, the EFL reading fluency conceptualization work presented here, along with the set of instruments developed for the purpose of this study, will be available to future researchers, thereby contributing to progress in the field of foreign language reading research.

The investigation of the connection between reading rate and comprehension has the potential to provide further justification for a focus on reading fluency development in EFL educational environments, as well as adding clarity to the issue of silent reading rate versus oral reading rate. This has both research and pedagogical implications. On the research side, the results of this study are expected to help establish whether silent or oral reading rate is more appropriate for accurately measuring the reading rate of EFL adult readers. On the pedagogical side, if it is found that oral reading correlates more highly with comprehension than silent reading, this would presumably lead teachers toward including an oral reading training component in reading courses.
Finally, identifying patterns of reading fluency development in a longitudinal ER program has implications for the design of ER courses. For example, if most fluency gains occur in the first third of a semester-long course and taper off after that, teachers might decide to concentrate fluency development activities such as timed readings at the beginning of the course and then phase them out. If relatively even fluency gains continue throughout the semester, however, teachers could evenly space their fluency development activities throughout the course. A third possibility is that if fluency development patterns are highly individualized and perhaps dependent upon initial fluency level at the start of the program, teachers might want to incorporate a degree of automaticity and individual choice into which activities students spend their time on during class.

**Theoretical Perspective**

This study took as its starting point the current conception of reading fluency in the L1 reading research community, as summarized by Kuhn et al. (2010). In this view, reading fluency (operationalized in most L1 research as oral reading fluency) consists of accuracy, automaticity, and prosody. Accuracy is often described and operationalized as accurate word recognition and decoding in an oral reading context. This description assumes that if a reader pronounces a word correctly, he or she also accurately comprehends it. This statement is not always true, even in L1 environments, and in L2 learners this assumption is further complicated by the possibility that readers sometimes understand a word but are not able to pronounce it accurately. That is to say, the decoding of words might sometimes skip the aural route and go directly from visual recognition to
comprehension of meaning. In order to explore this phenomenon in this current study, I developed and included measurements of both oral and silent reading fluency.

Kuhn et al. (2010) relied primarily, as do I, on Logan’s (1988, 1997) definition of automaticity, namely, that a process becomes automatic when it is fast, effortless, autonomous, and occurs without conscious effort. Speed increases concurrently with accuracy, due to practice, and follows the power law in which gains are largest at the beginning and diminish with further practice. Effortlessness is related to ease and occurs when the multiple sequential steps that beginners use to decode words are collapsed into a single step, allowing fluent readers to focus their cognitive resources on higher-level comprehension processes. When reading becomes automatic, it occurs without intention and becomes an obligatory act, meaning it is impossible for readers not to read text on which their eyes and attention are focused. The final characteristic of automaticity in reading is the lack of conscious awareness and effort. That is to say, the slow, deliberate process of disfluent readers trying to decode words becomes something that fluent readers are no longer aware of.

The third aspect of the construct of fluency, according to Kuhn et al. (2010), is prosody. They defined prosody as reading with “appropriate expression or intonation coupled with phrasing that allows for the maintenance of meaning” (p. 233), and proposed that pitch (intonation), duration (of syllables), stress, and pausing as the primary features of prosody. They did not specifically mention the importance of connecting words to form meaningful phrases in this list of features; however, they introduced two of the most commonly used oral reading fluency rating scales, the NAEP scale (Pinnell et
al., 1995) and the Multidimensional Fluency Scale (Rasinski et al., 2009), both of which rely heavily on judgements of meaningful phrasing.

While the accuracy-automaticity-prosody conception of reading fluency is an appropriate and well-established starting point (Kuhn et al., 2010) for this study, there are a few questions to consider with regards to its application to L2 learners in EFL environments. The first is whether word recognition itself is a sufficient indicator of accuracy. A measure of comprehension of the text should also be included if a fuller picture of accuracy is to be obtained. The second is the question raised above about whether the inability to pronounce a word correctly necessarily indicates EFL learners’ inability to recognize and comprehend the word. If such learners can comprehend words they cannot pronounce while reading silently, it would be problematic to apply this model to EFL contexts as is. The third is similar to the second, but with regards to prosody rather than word recognition. That is to say, it is unknown whether meaningful phrasing and intonation in oral reading is a required condition of fluent silent reading for adult EFL readers. It is conceivable that fairly fluent silent readers might struggle with oral reading, due to the obstacles of pronunciation and prosody. Indeed, one purpose of this study was to shed light on this issue.

In examining the role of automaticity in reading fluency, I rely on Logan’s (1997) definition of automaticity, which is that automatic processing is fast, effortless, autonomous, and does not require conscious awareness. The mechanism of automaticity in reading is also well-explained in his Instance Theory of Automatization (1988), which states that every encounter with a word or task lays down a trace called an instance representation in the readers’ memory. When processing the meaning of each letter,
word, or phrase in a text, readers can either draw upon a cognitive algorithm (e.g., thinking about the pronunciation of a letter or the meaning of a word), or they can directly access the instance representation in memory. As the number of encounters with a given word increases, the direct recall of the trace representation from memory becomes faster, making it more efficient than the slower algorithm so that it becomes more likely that the meaning can be recalled instantaneously and automatically from memory. It is a complex and well-developed theory, but the bottom line is that increased practice and exposure lead to higher levels of automaticity.

Reading fluency and reading comprehension are central focuses of this study, so it is necessary to consider how successful reading comprehension occurs. Perfetti’s (1985) Verbal Efficiency Theory points out that the automatization of lower-level processes such as word recognition frees up limited attentional and cognitive resources for the higher-level processes involved in comprehension, which essentially involves integrating each new bit of information read into an overall picture of the content of the text, like adding pieces to a puzzle. The higher-level processes involved in this process include using existing background schema to understand the text, synthesizing information from multiple locations in the text, and making inferences when necessary.

The Audience for the Study

The primary audience for this study consists of L2 reading researchers, with a particular focus on researchers concerned with reading education in EFL contexts such as Japan. The results of this study contribute to the growing body of research on the effectiveness of ER as an approach to reading proficiency development. More
importantly, it bridges the gap between the large body of L1 reading fluency research and the smaller body of reading rate and reading fluency studies in EFL contexts. It introduces a model of reading fluency that is appropriate for adult EFL readers and develops instruments that measure reading fluency in such environments more accurately and thoroughly than current techniques. Finally, I provide researchers with detailed data on how reading fluency develops in EFL readers, as well as the role of prosody and oral reading rate and its relationship to silent reading rate and reading comprehension.

The secondary audience for this study consists of reading teachers and program administrators in EFL secondary and tertiary educational settings. The results of this study provide evidence to support placing more emphasis on and providing more resources for reading fluency development in an educational system that currently devotes nearly all of its reading instruction time and effort to form-focused instruction. My hope is that this research helps teachers and schools in EFL environments such as Japan determine whether, how, and to what extent they should implement ER programs or incorporate other types of fluency development activities, and that this will in turn positively affect students’ reading ability as well as their motivation toward and enjoyment of reading in English.

**Delimitations**

The external validity and generalizability of this study’s results are affected by a number of participant-related variables. The most obvious point of concern is the difference between EFL and ESL contexts. Because the focus of this study is on the L2 development of Japanese university students, the results can only be generalized to L2
learners in ESL environments with caution because the environment (in particular the extent to which the learner is exposed to the L2 in the course of daily life) and the motivations towards learning the L2 usually differ greatly between EFL and ESL learners. Even generalizing to non-Japanese EFL environments is problematic because specific cultural characteristics and educational background issues, for example, the emphasis on accuracy over fluency that is prevalent in the Japanese educational system as well as the high-stakes university entrance examination system and the content of those exams, are likely to affect the reading fluency development of the students in this study as well as their attitudes and approaches to the ER course.

Another area of concern regarding the generalizability of the results is the participants’ age and English proficiency. The students in this study were first-year university students with a fairly high level of general academic ability, but with an average level of English proficiency for university students in Japan. It would be wise not to generalize the results to younger children, nor to students of a much higher English proficiency level. However, because of the diminishing returns expected when learning any skill, students at lower proficiency levels should benefit even more from instructional approaches focusing on fluency development than the intermediate (within the context of the Japanese undergraduate population) students in this study.

A final delimitation concerns the participants’ field of study, which in this case is nursing. They are likely to have less motivation than, for example, English literature majors or international business majors. As with proficiency, however, this is likely a positive factor with regards to generalizing the results of this study. Because ER allows students to read at their own ability level and is designed to be motivating for students by
allowing them to self-select books that are interesting to them, the results are likely to be
generalizable to other fields in which students have only moderate to low motivation
towards learning English. The results might be less applicable, however, when it comes
to highly motivated students who have already achieved high levels of reading fluency.

**Organization of the Study**

In this chapter, I have provided a brief overview of the problems and purposes this study was designed to address, as well as its significance, its audience, and a discussion of its delimitations. In Chapter 2, I provide an extensive literature review covering the following areas: reading fluency, ER, reading rate gains through ER, oral reading versus silent reading, reading comprehension, and the relationship between reading rate and comprehension. In Chapter 3, I describe the overall methodology and treatment procedures used in the study, while in Chapter 4, I focus on instrument validation. In Chapter 5, I present the results of the study, and in Chapter 6, I highlight and explain the key findings while connecting them to previous research. Finally, in Chapter 7, I summarize the findings, point out key limitations, make suggestions for future research directions, and provide closing comments.
CHAPTER 2
REVIEW OF THE LITERATURE

In this chapter I review the literature in a number of areas relevant to this study. I begin with a brief but broad discussion of the definition, models, and subskills of reading fluency itself. I then introduce extensive reading, touching on proposed definitions as well as the type of research that has been conducted on this approach. Next, I provide an overview of studies that have investigated reading rate gains attributable to the ER approach, which is followed by a section covering the question of oral reading versus silent reading as a measurement of reading rate and fluency. Finally, there is a section on the construct of reading comprehension and measurement thereof, followed by overviews of studies concerning the relationship between reading rate and reading comprehension. Unresolved issues and gaps in the existing body of research are then identified and used as the basis for stating the purposes of this study and the specific hypotheses addressed in this study.

Reading Fluency

There is a broad consensus, from Sir Edmond Huey in 1908 to the seminal modern work by LaBerge and Samuels (1974) and nearly every expert on the subject since, for the notion that reading fluency consists of fast, accurate, and synchronized processing of component subskills, such as visual decoding, word recognition, phonological processing, and syntactic and semantic parsing or chunking, which are automatized and unitized in skilled readers so that their attentional resources can be
focused on the higher-level processes of comprehension, analysis, and interpretation (e.g., Grabe, 2009; Kuhn & Stahl, 2003; LaBerge & Samuels, 1974; Pikulski & Chard, 2005). LaBerge and Samuels argued:

> If each component process requires attention, performance of the complex skill will be impossible, because the capacity of attention will be exceeded. But if enough of the components and their coordinations can be processed automatically, then the load on attention will be within tolerable limits and the skill can be successfully performed (p. 293).

Nathan and Stanovich (1991) described the phenomenon in this way: “When processes of word recognition take little capacity (are fluent), most of the reader’s cognitive capacity can be focused on comprehending the text, criticizing it, elaborating on it, and reflecting on it—in short, doing all the things we know good readers do” (p. 176). Similarly, Adams (1994) explained that when the reader’s attention is focused on mechanics, it is not available to support general comprehension of the text as a whole. Word recognition and semantic access that is “rapid, effortless, and automatic,” however, frees up the requisite cognitive resources for comprehension (p. 5).

Concerning the specific components of fluent reading, Wolf and Katzir-Cohen (2001) listed sublexical decoding processes, phonological representation, lexical access and retrieval, prosodic knowledge, and connected-text comprehension. Kuhn and Stahl (2003) settled on three components—decoding, word recognition, and prosody—while Grabe (2009) listed four subprocesses essential to reading fluency—automaticity (defined as not requiring attentional resources, unconscious, and not subject to interference or suppression, as in Logan, 1988), accuracy, rapid overall rate, and recognition of prosodic
phrasing or chunking. Thus, there is general agreement on the conception of oral reading fluency used in this study, namely, that it consists of accuracy of decoding and word recognition, automaticity of decoding and word recognition, and prosody.

Breznitz (2012) suggested that fluency in reading is primarily based on the rate of word decoding, which itself is determined by processing speed. She elaborated that multiple brain systems are activated during word reading all process information at different speeds, and that synchronization of the information arriving from these different systems is essential for word decoding. Her research on dyslexic readers revealed that the slow decoding rate of dyslexics can be traced to slow processing speed, which itself is due to speed asynchrony between the components. Specifically, she argued that what she called the word regulator in the brain must synchronize these systems, and that in dysfluent readers this regulator often has difficulty synchronizing the visual and auditory modalities at lower levels of processing and also the phonological-auditory system at higher levels. She also stated that the phonological, orthographic, and semantic systems operate at their own speeds, so the regulator must synchronize these systems.

The picture that emerges is that reading fluency development is much more than simply increasing the speed or smoothness of reading. It is a complex, multifaceted phenomenon involving the automatization of lower-level processes and skills to free limited attentional resources so that these resources can be employed toward comprehending the global message of the text. It is also a matter of increasing and regulating the speed of the various systems and components involved. It therefore follows that improved reading rate and fluency should go hand in hand with improved comprehension. Reading fluency development should be the ultimate goal of reading
instruction and reading improvement, with the understanding that accuracy, comprehension, rate, and prosody are all components of fluency.

How, then, does reading fluency develop, and how can its development be encouraged through educational programs? Nathan and Stanovich (1991) provided the simple answer that large amounts of practice via exposure to print is what leads to automaticity and thereby fluency. In L1 settings, enormous gaps in amount of reading practice develop by early elementary age, creating a rich-get-richer phenomenon known as the Matthew effect (Stanovich, 1986). ER, repeated reading, speed reading, and other fluency training methods have been proposed to help students develop reading fluency.

Text Difficulty, Readability, and Simplification

Of particular concern in this study, due to its focus on reading fluency and the distinction between ER and IR, is the concept of text difficulty and readability. The difficulty that a given reader has in reading and comprehending a given text necessarily depends on the interaction between that reader’s abilities and knowledge on one hand, and specific features of the text on the other.

Text difficulty in L1 contexts has traditionally been measured by readability formulas such as Flesch reading ease and Flesch-Kincaid grade level that take into account the lexical features of the text as well as the sentence-level syntactic complexity (Chall & Dale, 1995). Such measurements are used because they are simple, easy to understand, and have demonstrated strong predictive power; however, they have been criticized because they do not account for the full scope of the multicomponent of reading (Crossley et al., 2008). Carrell (1987) argued that L2 readability formulas needed to
include reader characteristics as well as a broader range of text factors such as syntactic complexity, rhetorical organization, and propositional density.

In L2 contexts, it should come as no surprise that the degree to which a reader’s vocabulary knowledge matches the lexical difficulty of a text is the most important factor that determines how easily a reader can read and comprehend the text. This concept is known as “lexical coverage”, because it refers to the extent to which a reader’s existing vocabulary knowledge “covers” the vocabulary used in the text. Hu and Nation (2000) found that most readers need at least 98% coverage in order to gain adequate comprehension when reading a text, and that 95% coverage is the minimal level at which some readers can gain adequate comprehension. This means, according to Nation’s (2006) corpus study, that EFL readers need a vocabulary of 8,000–9,000 words to read a novel in English.

This brings us to the need for simplified readers, often called graded readers, in EFL reading contexts. Graded readers are complete fiction or non-fiction books that have been written or adjusted so that they use only a limited set of words and grammar structures (Nation, 2001). They can be original works or simplified versions of existing authentic (written for L1 readers with no explicit or systematic regard given to lexical/grammatical simplicity) literary works. Simplified texts are therefore easier for EFL readers to read and comprehend, and authentic texts are more difficult.

---

**Extensive Reading and Reading Fluency Training**

In their 1998 book on the subject, Day and Bamford laid out a set of characteristics of ER. In addition to reading a large volume of text, they proposed that
successful ER programs allow students to self-select from a wide variety of enjoyable reading material which is written well within their ability level, allow students to read individually and silently, orient students to the goals of ER including an emphasis on reading speed, avoid follow-up exercises as much as possible, and encourage students to read for pleasure or information rather than for vocabulary and grammar learning purposes. Waring and McLean (2015) suggested that in practice there are several distinct and equally valid types of ER, but propose the essential core attributes to be fluent comprehension, high reading rate, large amounts of text, and a meaning-focused approach. Yamashita (2015) proposed large amount of text, easy reading material, high reading rate, and reading for pleasure as the four features common to various definitions of ER. In their 2020 book, *Teaching Extensive Reading in Another Language*, Nation and Waring distill the essential features of extensive reading, in order of importance, to: reading at the right level (only 2% unknown words), reading with comprehension (small amount of dictionary use needed), reading a lot (upwards of 100,000 words per year, and more than 10 minutes per day), reading independently (each student reads a different book they have chosen based on their proficiency level and interests), and reading silently (because reading aloud limits reading speed).

Concerning the benefits of the ER approach, Day and Bamford (1998) focused their discussion on development of fluent reading ability as well as attitude and motivation toward L2 reading. However, when listing the components of fluent reading ability, they mentioned only sight vocabulary, general vocabulary knowledge, and cognitive schemata of real-world knowledge. Their summary of research results
supporting the effectiveness of ER covers the areas of general reading ability, affect, vocabulary, overall L2 competence/proficiency, writing, and even spelling.

A great deal of research on ER has been conducted in the past three decades, with tentatively positive results in the areas of incidental vocabulary acquisition (e.g., Day et al., 1991; Horst, 2005; Suk, 2017; Waring & Takaki, 2003), affect (e.g., Mori, 2004; Nishino, 2007; Robb & Susser, 1989; Yang, 2001), and general reading proficiency (Elley, 1991; Elley & Mangubhai, 1983; Mason & Krashen, 1997; Tudor & Hafiz, 1989). In the area of affect, Burrows (2012) conducted a study on the effect of ER and other reading strategies on self-efficacy development, finding that the combination of ER and reading strategies improved students’ self-efficacy, and that this was related to gains in reading comprehension. In the area of general reading proficiency, Jeon and Day (2016) conducted a meta-analysis which found that ER had a small to medium effect ($d = .57$ for experimental designs; $d = .79$ for pre-post designs) on reading proficiency (defined as reading comprehension, reading rate, and vocabulary). The question of whether reading comprehension is improved by extensive reading has also received attention. Some researchers have reported increased reading comprehension (Elley, 1991; Homan et al., 1993; Pichette, 2005; Suk, 2017), while others have found no difference in comprehension between ER and control groups (Nakanishi & Ueda, 2011). Another likely benefit of ER is in the area of improved reading rate and/or fluency. Fluency is often operationalized as the ability to read rapidly while accurately comprehending the text, although this is a somewhat simplistic measure of fluency. A number of researchers have measured and reported reading rate changes while confirming comprehension, as an indicator of fluency development (Beglar et al., 2012; Grabe, 2009; Huffman, 2014;
McLean & Rouault, 2017). Examples of this small but growing body of research are reviewed in the next section. In summary, investigations of the effectiveness of ER have traditionally focused largely on outcome variables relating to affect, vocabulary, and reading comprehension, with a more recent interest in reading rate also building.

Although ER has been suggested as a way to improve reading fluency, the definitions and research above reveal that most ER teachers and researchers place less emphasis on development of the reading rate aspect of fluency than on simply providing readers with the opportunity and encouragement to read large amounts of simplified text for pleasure. However, as I have previously noted (Huffman, 2016), reading rate and fluency enhancement activities should be included in extensive reading programs and courses, because they likely amplify the rate, fluency, and comprehension gains that are expected from extensive reading alone. It is therefore a wasted opportunity to not include such activities. Several methods have been developed, in both L1 and L2 environments, that are specifically designed to improve reading rate and fluency. Examples are speed reading and/or timed readings, repeated reading, and various integrated reading fluency development/training approaches. In the ER course that is the focus of the current study, elements of these fluency/rate development approaches are included to enhance the fluency improvement benefits of ER. I term this approach “extensive reading with fluency training (ERFT).” Shimono (2019) also conducted a longitudinal study (reviewed below) on a course that incorporated timed reading, repeated oral reading, and ER.

Repeated reading (Samuels, 1979) involves readers in reading a short text repeatedly, during which they are encouraged to notice their rate increasing and their errors decreasing. The students read orally; this might be preceded by a model reading
from the teacher or an audio tape. In L1 settings with children, this method has been shown to increase oral reading rates, accuracy (transferring also to novel passages), and comprehension (Dowhower, 1987). This method has recently been employed in L2 settings as well. For instance, Taguchi (1997) found that silent re-reading led to increased silent reading rates but the improvement did not transfer to novel texts. Taguchi and Gorsuch (2002) did not find significant transfer gains from a 10-week treatment period, while Gorsuch and Taguchi (2008) reported no transfer gains for reading rate, but they did find significant improvements in reading comprehension on unpracticed texts as a result of the 11-week repeated reading treatment.

Speed reading courses in which students measure and attempt to increase their silent reading rate by reading a series of simplified passages have been promoted and researched by researchers such as Nation (2009b). Chung and Nation (2006) reported reading rate gains, particularly from the first ten texts of a 23-text course lasting nine weeks. Macalister (2010) and Tran (2012) both conducted studies indicating that reading rate gains from speed reading likely transfers to other types of texts. It is likely that including timed readings as a component of an ER course not only gives students extra practice in reading fluently, but also motivates them to focus more intentionally on increasing their reading rate.

Rasinski et al. (1994) developed a reading fluency development lesson for L1 children that incorporates various principles of reading fluency instruction. The teacher begins by modelling fluent oral reading and then leads the class in a discussion about the rate, phrasing, expression, and intonation they noticed in the model reading. The class then reads the text chorally, followed by pair work. In pairs, the students read the text
several times to their partner, and the partner provides positive feedback. Finally, individuals are invited to perform their oral reading for the class. The researchers found significant treatment effects for this method, but this study involved a small number of participants (experimental \( n = 28 \) and control \( n = 26 \)). In the ER course that is the focus of the current study, I have incorporated a similar oral reading fluency development component to enhance the reading fluency benefits of ER. I adapted this training based on the Rasinski et al. (1994) because it is specifically designed for reading fluency development, as opposed to other methods (such as De Jong & Perfetti, 2011) that are focused on oral fluency development in general.

Shimono (2018) conducted a study in which both timed readings and oral reading fluency training were successfully used to improve students’ silent reading rates. In this study, 55 Japanese university students were divided into three treatment groups: timed reading with repeated oral reading (focusing on chunking and prosody), timed reading only, and oral communication training (the control group). Both treatment groups made significant reading rate gains, ranging from 13 to 27 swpm depending on the method of scoring used, and both outperformed the comparison group in both rate and comprehension gains. This study provides further evidence that timed readings and oral reading fluency training, both of which were used in the current study, are effective in increasing reading rate.

Shimono expanded on the above study in a subsequent study (2019), in which there were four treatment groups: 1) ER with timed reading and repeated oral reading, 2) ER with timed reading, 3) ER only, and 4) speaking and communication activities only (the comparison group). The results of this study showed that the ER with timed reading
and repeated oral reading group made the most reading rate gains, followed by the ER with timed reading group and then the ER group. However, the ER-only group did not significantly outperform the comparison group. The ER with timed reading and oral reading group also made the most gains in orthographic, semantic, and phonological processing. These results provide empirical support for adding timed reading and oral reading fluency development components, as I did in the current study, to the standard ER approach.

**Extensive Reading and Reading Rate Gains**

In this section I provide summaries of key quantitative investigations of reading rate gains associated with ER and other fluency development approaches in EFL contexts.

Robb and Susser’s (1989) investigation of ER vs. a reading skills approach (explicit instruction and training in reading skills such as identifying the main idea of a text, understanding important details, guessing vocabulary from context, and making inferences) in intact year-long university classes of freshman English majors in Japan included a pre- and post-treatment measurement of reading speed for both groups ($N = 125$). Although the two measurements were not equivalent, the ER students read significantly faster during the first minute of reading than the IR students post-treatment (86.55 vs. 76.75 wpm), while there was no significant difference between the two groups pre-treatment. However, the skills group read the total passage much more slowly than the ER group pre-treatment, while this gap was much lower post-treatment, muddying the results of this study. The ER group scored significantly higher than the IR group on
comprehension measures post-treatment, where there were no significant differences pre-treatment. The combination of greater increase in both reading rate and comprehension indicated that ER likely improved reading fluency to a greater extent than did IR.

Drawbacks in this study are typical of those found in ER studies using intact classes. Time-on-task was nearly double for the ER group, so the reading rate gains might have been due to increased time spent reading rather than the pedagogical approach itself. Another drawback related to the use of intact classes is that the students were taking a total of six English courses concurrently; therefore, their fluency gains might not have come from the ER course alone. A final drawback is that the ER students were reading from a library of authentic fiction written for L1 teenagers rather than simplified graded readers, so it is likely that they did not meet the conditions required for fluency development to occur, such as a low density of unknown words (Nation, 2009a).

Another study focused on speed and comprehension changes related to different reading approaches was conducted by Bell (2001). Bell investigated the effects of ER and IR methodology with young adult EFL learners in Yemen, reporting changes in reading speed and reading comprehension over one year. Time-on-task was equivalent for the two groups, and the reported results are impressive indeed. The IR group’s rate increased from 78.45 wpm to 92.54 wpm, while the ER group increased from 68.10 wpm to 127.53 wpm. Both of these gains and the post-treatment difference between the two groups were statistically significant. The ER group also made higher comprehension improvement than the IR group. The drawbacks of this study include the small number of participants \((N = 26)\) and the fact that the amount of text read by the participants was not reported. A much bigger drawback, however, is the fact that student comprehension of the text used
for the reading rate measure itself was apparently not checked; thus, the impressive reading rate gains seen by both groups in this experiment might have come at the expense of comprehension, a danger the author himself warned about multiple times in the paper.

Some studies designed to ascertain the effect of ER and other approaches on reading rate have reported negative results. For example, Taguchi et al. (2004) researched the effects of assisted repeated reading versus ER during a one-semester course on reading rate and comprehension of Japanese university students ($N = 20$). They reported a decrease in reading rate for both groups after the treatment, but it should be noted that the ER group read only 205 pages during the semester, the pre-treatment and post-treatment measures of reading rate were not equivalent, and that the books were likely well-above the level required for fluency development (comprehension scores were 1.90/16 on the pretest and 4.50/16 on the posttest). For these reasons, this study might have been limited in its ability to reveal reading fluency gains that result from semester- and year-long ER programs that involve a greater amount of reading and texts that are more closely matched to students’ reading ability levels.

A stronger study that avoided most of the pitfalls of previous studies on this topic was conducted by Beglar et al. (2012). This study provided robust evidence of the effect of reading extensively on fluency development. The participants were first-year Japanese university students ($N = 97$), consisting of an IR group and three treatment groups engaging in various amounts of pleasure reading (PR) during a one-year program. The IR group made negligible reading rate gains, while the PR groups ranged from gains of 8.02 standard words per minute (swpm) (89.71 swpm pre-treatment to 97.73 swpm post-treatment) to 16.85 swpm (103.9 swpm to 119.93 swpm). These results were in line with
the amount of reading done by the different PR groups, and the gains came without any accompanying decrease in comprehension of the passages. The reason this program was termed pleasure reading rather than ER is that students were provided with and permitted to read authentic, unsimplified novels in addition to simplified graded readers. However, the authors provide a clear analysis of the correlation between simplified and unsimplified reading and reading rate gains, finding that simplified reading resulted in greater rate gains. A minor limitation of this study, identified by the authors themselves, is that reading rate was measured in a slightly inaccurate way, by writing the elapsed time in 10-second increments on a whiteboard and having students look up and record their time when they finished reading the text. Overall, however, this study represents an improvement on previous studies and should be seen as a model and a springboard for further studies on fluency improvement through ER in a variety of EFL settings.

The Beglar et al. (2012) study is also notably the first in this line of research to use standard words per minute (swpm) to measure reading rate. Particularly when dealing with texts at differing difficulty levels and when comparing simplified and authentic texts, the length of words becomes an important variable. For this reason, standard words are a more accurate unit of measurement than words. Kramer and McLean (2019) conducted a study to determine the impact of word length on reading rate measurements, finding that the number of characters in a reading passage has a measurable effect on the time spent reading, and effect that goes beyond the effect of text length when counted using words. The use of standard words also allows for a more reliable comparison of results across studies (Beglar et al., 2012).
In a study that used only simplified readers for the ER group and added an explicit reading speed enhancement component, Huffman (2014) investigated the reading rate gains of Japanese college freshman in a one-semester ER course versus those in an intensive reading course ($N = 66$). The ER course included a speed reading component consisting of six timed readings, and the students were actively encouraged to work on increasing their reading rate. The ER group achieved significantly (by $t$-test) higher reading rate gains (20.73 swpm) than the IR group (-.62 swpm) with no accompanying reduction in comprehension. One limitation of this study was that the pre- and post-course reading rate measurements were conducted using different forms. Another limitation was that the ER group students spent more time on task (reading) during the semester, so the results could have come from this difference in time spent reading rather than the instructional approach (ER) under investigation. That study can be seen as a precursor to the current study, and indeed the current study attempts to improve on the Huffman (2014) study and other studies highlighted in this section.

A more recent study that avoided many of the limitations of previous studies was published by McLean and Rouault (2017). This study also involved Japanese first-year university EFL learners, and set up extensive reading as the experimental condition and grammar-translation as the control condition. The researchers effectively controlled for time on task by ensuring that both groups spent the same amount of time on their treatments, identical pre- and post-treatment test forms were used, and both treatment groups engaged in weekly timed reading exercises. The average reading rate of the extensive reading group increased significantly more (30.96 swpm) than that of the intensive reading group (5.15 swpm). While providing further evidence that extensive
reading yields increased reading rates with good comprehension levels, this study also represents a strong step forward toward the goal of balancing the need for experimental control with ecological validity in longitudinal studies in pedagogical settings.

In a notable study conducted by Suk (2017), intensive reading was compared with extensive reading, although the extensive reading group had a substantial in-class intensive reading component. The participants were Korean undergraduate university students. This researcher controlled for time on task well, and used a dual-purpose measurement for reading comprehension and reading rate. Although it seems logical that the focus on comprehension would put a damper on the reading rate results, but the extensive reading group was still able to demonstrate a greater reading rate increase than the intensive reading group under these conditions, and this provides greater confidence that the reading rate measurement can be used as an accurate measure of reading fluency, if we assume that rate with comprehension equals fluency. The two groups had different baseline reading rates (ER group: 133.29 wpm; IR group: 147.76 wpm), which is concerning, but the ER group clearly achieved a greater mean reading rate gain (ER: 35.13 wpm; IR: 15.53 wpm).

Kida (2016) conducted a notable study which measured the effect of ER on a different aspect of reading fluency, rather than reading rate: automatization and orthographic development. The 41 Japanese university students in this study completed a form-priming lexical decision task before and after 12 weeks of 30-minute sustained silent reading (SSR) activities (a core ER activity). The results indicated that participants increased their word recognition automaticity but not their orthographic representation. One important limitation of this study is that the students were engaging in what the
author describes only as “various text-based reading and vocabulary tasks” for the other 60 minutes of each 90 minute class. So the results seem more likely attributable to a combination of ER and IR activities, rather than solely to the SSR (ER) activity. Still, this study expands on the strand of research investigating the effect of ER on reading fluency in an important way, by focusing on automatization of word recognition and orthographic representation rather than reading rate gains.

Turning to the effect of timed readings (speed reading) on reading rate, Chung and Nation (2006) reported on a nine-week speed reading course with 40 Korean first-year university students. By calculating the average of the first three and last three texts, an average reading rate increase of 73 wpm (141 wpm to 214 wpm) was reported. The researchers also found that most students made greater rate gains during the first ten texts vis-à-vis the second ten texts. However, there was no comparison group, and no statistical analyses were conducted. Another weakness is that the question of whether these rate gains transfer to other types of texts, such as authentic texts, remained unanswered.

In an attempt to address some of the weaknesses in the Chung and Nation study, Macalister (2010) conducted a study of 36 ESL university preparation students in New Zealand, which involved 24 treatment students engaged in a 12-week speed reading course and 12 control students. Students were pre- and post-tested on both simplified speed reading texts and authentic texts to investigate transfer effects. All of the speed reading students showed an increase in reading speed, while just over half of the non-speed reading students increased their speed; the increases of the speed reading students were greater. However, no statistical analyses were reported. The treatment group also showed much bigger reading rate gains on authentic texts than the control group.
Keeping Macalister’s idea of investigating for transfer effects of ER rate gains to authentic texts, and also adding statistical analyses, Tran’s (2012) participants were 116 first-year university students in Vietnam, divided into experimental (speed reading in addition to the usual English program) and control (no speed reading) groups. The pretest and posttest were conducted using texts that were more difficult, longer, and on different topics compared to the texts in the speed reading course. The results showed an average increase of 50.9 to 57 wpm for the treatment groups on the speed reading texts (no statistical analyses) and a 48.26 wpm increase for the treatment group versus 15.3 wpm increase for the control group on the longer and more difficult texts on different topics. A one-way ANOVA on the effect of treatment on average speed increase revealed these differences to be significant ($p < .001$).

If reading rate and fluency improvement are to gain credence as an important educational goal in EFL contexts, it is important to consider what sort of targets are seen as ideal for readers. Carver (1982) found that the most efficient *rauding* rate for college students reading L1 college-level material was around 300 wpm; *rauding* refers to the rate at which a reader reads with maximum efficiency. Nuttall (2005) also reported that the average L1 reader reads at around 300 wpm, although there is a wide range around that figure. Higgins and Wallace (1989) noted that 180 wpm is “generally agreed to be close to the minimum at which reading becomes a pleasure” (p. 394) and that it represents “a threshold between immature and mature reading” (p. 392). Beglar et al. (2012) suggested that this minimum L1 rate could be used as a reasonable goal for L2 readers in many contexts. Nuttall (2005) noted that secondary school students in ESL countries—countries where English is not the L1 but is used as the primary or exclusive
language of education in all subjects, as opposed to countries where subjects other than English are taught in the L1—read at around 120 to 150 wpm before training. There is currently no empirically-based target for use in with EFL reading research.

The research on reading rate gains from ER and speed reading courses presented in this section represents growing body of research that preliminarily indicates a likelihood that these approaches are effective for reading rate improvement and other aspects of reading fluency improvement without sacrificing comprehension, and that such improvement might extend to other types of texts or even authentic texts.

**Intensive Reading**

Formal English reading instruction in EFL classroom settings at the secondary and tertiary level have typically consisted of explicit reading instruction and practice that can involve primarily the teaching of grammar rules and translation into the L1 (often termed the grammar-translation approach, or *yakudoku* in Japan), or alternatively avoid use of the L1 but still focus on explicit instruction and training in reading skills and strategies (often termed “intensive reading (IR)”). In Japan, yakudoku has traditionally been employed by most Japanese teachers of English (Hino, 1988; Tanaka & Stapleton, 2007), while IR (whether termed so or not) is more common in global EFL settings (Brown & Lee, 2015; Nation, 2009a).

In Brown and Lee’s (2015) chapter on teaching reading, the authors describe the teaching of reading strategies (identify the purpose, teach bottom-up decoding, skim for main ideas, scan for specific information, guess from context, analyze vocabulary, notice discourse markers, etc.) and the training of reading micro- and macroskills (discriminate
among graphemes and orthographic patterns; recognize grammatical word classes, patterns, and rules; recognize cohesive devices and rhetorical forms; infer non-explicit context using background knowledge; distinguish between literal and implied meanings; etc.) that can be used for explicit classroom reading instruction. They discuss all of this without explicitly mentioning IR, highlighting the fact that IR is often discussed and practiced under terminology such as “reading instruction” “reading skills/strategy instruction” without necessarily being termed “IR.” It only becomes necessary to refer to it as IR when it is compared, contrasted, or otherwise juxtaposed with ER, such as in the current study and the pedagogical approach described therein. Brown and Lee (2015) do go on to explicitly define IR as a classroom-oriented approach that focuses on the linguistic and semantic details of a text and referring to it as a “content-related reading strategy” (p. 409).

Nation (2009a) devotes an entire chapter to what he explicitly refers to as “intensive reading.” He describes it as focusing on studying the linguistic features of the text and improving readers’ reading strategies and comprehension skill. He refers to the grammar-translation approach as the “classic procedure for intensive reading” (p. 25), and regards it as a useful approach if it is part of a multi-faceted and well-balanced reading program. Typical focuses of IR activities described in this chapter are comprehension, vocabulary, grammar, information content, genre identification, and reading strategies.

Rather than being discrete approaches, ER and IR are properly situated at two ends of a rich spectrum of reading approaches and activities. There has been little empirical research establishing the benefits of the IR approach, but it is clear that
educators and researchers assume that it holds some benefit for the improvement of reading comprehension, vocabulary knowledge, and other reading skills and strategies.

**Reading Comprehension and Vocabulary Acquisition**

Perhaps the most influential theory of reading comprehension has been Perfetti’s (1985) Verbal Efficiency Theory. It accounts for the confluency of reading rate, automaticity, reading fluency, and reading comprehension that are the focus of this study. Moreover, it is similar to the description of reading fluency in the first section of this literature review. It starts with the fact that readers have limited attentional and cognitive resources (e.g., working memory) to apply to the complex process of reading a text, and it notably distinguishes between controlled, effortful processes that require the allocation of these resources and automatic processes that are essentially effortless. The more fluent readers become, the more processes become automatic, freeing up resources for higher-level skills such as processing the overall meaning of the text. This model explains how increased reading rate, and the accompanying increase in reading fluency, can result in improved comprehension.

Adding to the discussion of the higher-level processing skills is Kintsch’s (1988) construction-integration model. This model deals with the higher-level cognitive processes of constructing the meaning of the text as it is written (the textbase) and integrating this with the reader’s prior knowledge to create a coherent situation model of the underlying meaning. Therefore, multiple readers can potentially arrive at different situation models that represent individualized understandings from the same textbase.
This is due to the fact that texts are usually not fully explicit, and therefore “inferential thinking is the heart and soul of reading comprehension” (Farrall, 2012, p. 236).

Carver’s (1992) model of reading and four other reading “gears” is also useful for understanding the relationship between reading rate and reading comprehension. In his theory, which is based on Perfetti’s (1985) theory, readers can shift to different gears depending on their goal in reading a given text at a given time. Each gear represents a different reading rate and a different set of reading components. Gear 5 is scanning, where readers simply try to find specific information in a text. The only reading component involved is word recognition, so a typical native-speaker college student can scan at around 600 standard words per minute. Gear 4 is skimming, where readers skim over a text quickly to get a general overview. This type of reading involves word recognition and some semantic encoding, and therefore slows readers down to around 450 standard wpm. What people most commonly think of as reading is Gear 3, which Carter calls reading. Reading involves reading every word, semantically encoding it, and also integrating each word’s meaning into the overall meaning of the sentence. This type of reading can be achieved by typical, fluent, native-speaker college students at around 300 standard wpm. In Gear 2, which is called learning, readers want to understand and learn the information in the text at a high level of comprehension, such as would be required when they are tested on the content afterward. This gear involves the additional component of idea remembering and slows readers down to around 200 wpm. The slowest gear is Gear 1, memorizing. This approach is required when readers need to recite from memory all of the ideas expressed in the text. This introduces the component of fact rehearsal, which brings the reading rate down to 138 wpm. This model of reading
highlights the interrelation between reading comprehension and reading rate, which is the central focus of this study.

Vocabulary knowledge is known to be highly correlated with reading comprehension. Wagner and Meros (2010) suggested that vocabulary knowledge has a direct, indirect, and reciprocal influence on reading comprehension. In a meta-analysis of L2 reading comprehension studies, Jeon and Yamashita (2014) identified vocabulary as one of four measures that are highly correlated with L2 reading comprehension, and found reading comprehension to be more highly correlated with linguistic knowledge such as vocabulary than with general, cognitive, or metacognitive knowledge. Qian (1999) reported that vocabulary size and depth are related to reading comprehension and they are independent constructs that contribute to reading comprehension in different ways. It is also clear from Hu and Nation (2000) and Schmitt et al. (2011) that increasing vocabulary size yields increased reading comprehension. A vocabulary size measurement was therefore included in this study as both a predictor of reading comprehension and a likely benefit of IR and possibly ER approaches.

The question of whether and to what extent EFL reading approaches result incidental vocabulary acquisition has been the target of an active area of research. Much of this research has focused on the notion that the more frequently a word is encountered, the more likely it can be learned incidentally. Horst et al. (1998), for example, found that eight encounters were needed to learn a word, while Waring and Takaki (2003) found that more than 20 encounters were required. Webb (2014) noted that reasons for this wide range of results include learner variables such as L2 proficiency and methodological differences such as spaced (in a longitudinal ER study) versus massed (intensive and over
a short term) learning conditions. A recent meta-analysis by Uchihara et al. (2019) concluded that there is a medium effect ($r = .34$) of frequency of encounter on incidental vocabulary learning, but emphasized that it is just one of many variables affecting the degree of vocabulary learning that occurs through reading and other meaning-focused input approaches.

What is clear from this body of research is that even with the exposure to larger amounts of text in the ER approach, it is difficult to set up a situation where students encounter words often enough, and at short enough intervals, to make incidental vocabulary learning an effective way to increase vocabulary size. Extensive reading must be continued for many years and at a high volume for this to occur as it does with L1 children. Explicit vocabulary learning approaches likely need to be included for EFL adult learners to achieve substantial vocabulary size growth.

**Relationship Between Reading Rate and Comprehension: Silent Versus Oral Reading Fluency Measurement**

It is a paradox that mature, proficient readers spend most of their real-life reading time in the silent reading mode, but reading rate and fluency measures are almost always in the oral reading mode because it is impossible for researchers to know what readers are doing in their minds when they read silently. Furthermore, it has become well established that, at least with L1 children, oral reading rate/fluency is highly correlated with silent reading comprehension (see Fuchs et al., 2001 for a comprehensive discussion including empirical support). However, several researchers have questioned whether, particularly in EFL environments where the readers have an L1 orthography that differs greatly from the
orthography of English, oral reading is an accurate representation of how fluently and accurately readers read in the silent mode. In this section I summarize studies in which the association between reading rate or fluency and reading comprehension has been investigated in both L1 and L2 contexts, while also highlighting the results and researchers’ speculations relating to the appropriateness of silent versus oral measurements of reading rate and/or fluency.

Breznitz (1987) conducted a series of experiments designed to measure the effect of manipulating oral reading rates on reading accuracy and comprehension in first grade elementary students reading in their L1. Her first experiment was with 61 first graders in Israel. Reading rate, accuracy, and comprehension were measured in a self-paced and fast-paced condition. In the self-paced condition, the students read sentences at their own normal, comfortable reading rate. In the fast-paced condition, sentences were presented on a computer screen at a faster rate, which was equal to the fastest rate at which they had read and comprehended a sentence during the self-paced condition. In effect, the students were forced to read at their own fastest pace. The participants significantly increased their accuracy and comprehension in the fast-paced condition compared to the self-paced condition. The results also indicated that the poor readers benefitted more than the good readers. This experiment was repeated with American first graders, with similar results. In a separate experiment, slowing the participants’ reading rates increased their accuracy but reduced their comprehension. The author suggested that fast-paced reading frees up short-term memory resources that are normally tied up in slow decoding processes, allowing those resources to be directed to higher-level processing such as synthesizing and global comprehension.
In a follow-up study delving further into the reasons for this correlation between speed and comprehension, Breznitz (1988) found that “reading at a faster pace increased comprehension and reduced errors” (p. 47). Her results indicated that the improved comprehension was likely due to reduced distractibility because reading slowly allows too much opportunity for distracting stimuli to be introduced during the “empty time” between words or sentences, while a reading rate that pushes the limits of readers’ attentional resources leaves them with no spare processing capacity with which to register irrelevant stimuli.

Effect sizes were not reported in these studies (Breznitz, 1987, 1988), but evidence for a direct relationship between reading rate and comprehension was provided. They also suggest that incorporating fluency building activities, such as timed readings or re-reading activities, into reading courses is beneficial, as the students in the study were able to read much faster when forced to than when left alone, and faster reading resulted in increased accuracy and comprehension. However, these studies involved only oral reading, so it is not clear whether the findings would repeat themselves in a silent reading context with EFL students.

Another study focused on the correlation between reading rate and comprehension was conducted by Dhanapala and Yamada (2014) with Japanese students. The researchers set out to determine correlations between habitual and maximum rates, reading comprehension, and listening comprehension in Japanese EFL learners. The participants consisted of 117 undergraduate university students who took the listening comprehension and reading comprehension sections of the TOEIC® IP test as well as a
separate reading comprehension test and two oral reading rate measurements (habitual and maximum speed). The participants’ English proficiency level was not reported.

The resulting correlations between reading comprehension and oral reading rate as well as between listening comprehension and oral reading rate were significant but low ($r = .22-.29$). The correlation between reading and listening comprehension was significant and moderate ($r = .34$ for the researcher-made reading comprehension test and $r = .55$ for the TOEIC® IP reading comprehension section). The researchers identified eight “word callers” (fast readers with low comprehension) and nine slow readers with high comprehension. The mean habitual rate was 112 wpm, while the mean maximum rate was 124 wpm. Reading comprehension correlated significantly more highly with maximum rate than habitual rate. The researchers also hypothesized, based on the simple view of reading (Hoover & Gough, 1990), that reading comprehension would correlate more highly with the product of rate times listening comprehension than with rate alone or listening comprehension alone, but found this not to be the case.

Overall, the results of this study call into question the direct application of L1 reading fluency research results to L2 contexts. The authors postulated that the reason for the low correlation between rate and comprehension was that the Japanese EFL readers in this study were reading too slowly for the rate-comprehension connection to develop. They also offered an alternative explanation that because the reading comprehension measure did not include a time pressure component, slow readers were able to read and reread carefully, raising their comprehension to a level that would be unattainable under time pressure. This situation could result in the dissociation of rate from comprehension.
Providing further evidence that L1 reading research in general, and the correlation between oral reading fluency and comprehension in particular, might not be easily applied to L2 learners, Lems (2006, 2012) examined the relationship between oral reading fluency and silent reading comprehension in adult English language learners. The participants were 232 college students in an academic ESL program in the United States. The measure of oral reading fluency was a one-minute oral reading from which words correct per minute, total words read, total number of miscues, and total number of significant miscues were obtained. The measure of silent reading comprehension was the final exam for the course, which consisted of many sections, each of which was focused on grammar and vocabulary, and which also contained a reading passage with five comprehension questions. The results revealed a significant but low correlation \((r = .256, p < .001)\) between words correct per minute and the silent reading comprehension measure. Correlations were higher for more proficient students \((r = .41, p < .01)\) and lower for lower proficiency students \((r = .04, \text{non-significant})\). It should also be noted that over 60% of the students were Polish, and when analyzed by first language, significant correlations were only found for the Polish and Hispanic students, and not for Chinese students, who accounted for 10% of the participants. The researcher also looked at the correlation between a different measure of oral reading fluency, a subjectively judged fluency rubric called the Multidimensional Fluency Scale (Zutell & Rasinski, 1991), and silent reading comprehension, this time with 80 of the participants. The result was similar to that of the correlation between oral reading fluency and silent reading \((r = .29, p < .001)\), but the interrater reliability was low. A miscue analysis was also conducted with seven of the mid-level students, finding that both the oral reading fluency
measure and the fluency rubric failed to account for some of the unique characteristics of English language learners (ELLs), such as fossilized nonnative accents.

In the study described above, Lems concluded that oral reading fluency as a measure of adult ELL silent reading comprehension might only be valid after learners have achieved a minimal threshold level of listening comprehension and recoding skills (recoding here refers to phonological recoding, or pronouncing a word out loud). For students from L1 backgrounds where the orthography differs from that of English, however, recoding skills develop so slowly and are so far behind silent reading comprehension skills that it might be necessary to find an entirely different model and method for assessing reading fluency that does not rely on decoding and recoding.

Further insight on how well oral reading fluency predicts reading comprehension in L2 readers was provided by Jeon (2012), whose participants were 255 South Korean high school students. She used exploratory factor analysis to investigate the association of oral reading fluency with various reading predictors such as decoding, vocabulary knowledge, grammar knowledge, and metacognition. She then used multiple regression analyses to examine the predictive power of oral reading fluency on reading comprehension. The results of the factor analysis showed that oral passage reading fluency, but not word and pseudoword reading fluency, crossloaded with the other reading predictors as a single factor. The regression analysis showed that oral reading fluency as a whole explained 21.2% of variance in reading comprehension, with most of this due to passage-level rather than word-level reading fluency. These results, and particularly the weak association between oral reading fluency and reading comprehension, supported the conclusion of Lems (2006) that “comprehension without
recoding” (p. 240), meaning visually processing and comprehending the text without pronouncing words (even with the inner voice) likely occurs more often in L2 than in L1 reading. Jeon (2012) suggested that L2 readers are often be able to successfully recognize a word without recoding it phonologically, and that this might be due to “lack of oral rendition practice, failure to map proper phoneme onto the corresponding grapheme, and failure to properly sound out the phoneme of the corresponding grapheme” (p. 201). The low correlation between word knowledge and word reading in Jeon’s study also supports the notion that L2 readers can have orthographic recognition knowledge without phonological recoding ability.

Also contributing to this discussion is the work of Jiang et al. (2012), who investigated the relationship among word reading efficiency, text reading fluency, and reading comprehension in 185 adult Chinese EFL learners. Specifically, they questioned whether word reading efficiency and/or text reading fluency are reliable predictors of reading comprehension in Chinese EFL learners due to the fact that Chinese is a logographic writing system with low systematicity in character-to-phoneme mapping. The word reading efficiency instrument consisted of real English words as well as pronounceable nonwords. Number of words read correctly was recorded as the measure of efficiency, and total number of words read was recorded as the measure of rate. For silent text reading fluency, the participants read one sentence at a time on a computer monitor, clicking a Go button when finished to proceed to the next sentence. For oral reading, the same passage was presented again, this time paragraph by paragraph. For silent reading, the reading rate was recorded as words per minute. For oral reading, the number of words read correctly (accuracy), number of total words read per minute (rate),
and number of words read correctly per minute (fluency) were counted. TOEFL iBT Reading section scores were used as the measure of reading comprehension.

In addition to correlation coefficients, a sequential hierarchical multiple regression analysis was conducted, involving five regression models. The results showed moderate Pearson correlations between oral text reading and reading comprehension, ranging from .45 to .51. The results of the regression analysis showed that phonemic decoding accuracy, sight word efficiency, oral reading accuracy, oral reading efficiency, and silent reading rate all added significantly to the prediction of reading comprehension. Of these, oral reading accuracy and oral reading efficiency accounted for a combined 21% of variance in reading comprehension scores (all five variables together accounted for 32% of variance). The correlations in this study are somewhat higher than the results in Lems (2006, 2012) and Jeon (2012), but still much lower than the results for L1 children. The correlations between word reading efficiency measures and reading comprehension were even lower, ranging from .27 to .34, and silent text reading rate also had a low correlation of .27. It should be noted, however, that the silent text reading measure included no accompanying measure of comprehension, nor indeed any method at all of ensuring that the participants were actually reading. The sequential regression analysis began with phonemic decoding accuracy, which accounted for 7% of reading comprehension variance, and proceeded to add word reading efficiency (+2%), oral text reading accuracy (+15%), oral text reading efficiency (+6%), and silent text reading rate (+2%). Together the five independent variables accounted for 32% of the variance in reading comprehension, with oral text reading and particularly accuracy winning out over single word recognition and silent reading rate. The researchers concluded that the results
are consistent with L1 reading research inasmuch as text reading measures were more strongly correlated with reading comprehension than isolated word reading measures.

In the same study, the researchers compared the results of the Chinese learners with the data from a separate group of 172 Spanish EFL students, finding that while the Spanish students read more quickly but with more errors, the Chinese students read more slowly but more accurately. They concluded, similarly to Jeon (2012) and Lems (2006, 2012), that the influence of a logographic L1 might cause Chinese learners to learn orthography-to-phonology mapping on a more holistic and arbitrary basis. This results in slower reading even though they are more proficient readers.

In an L1 context with 255 fourth-graders in the Netherlands, van den Boer et al. (2014) conducted a direct comparison of oral and silent reading fluency. Using structural equation modelling, they found that phonological awareness, as measured by requiring participants to manipulate phonemes in oral readings of nonwords, significantly and uniquely predicted both oral reading \( r = .229, p = .031 \) and silent reading \( r = .191, p = .047 \). Rapid naming of numbers and letters (measured in terms of naming rate) correlated more strongly with oral reading \( r = .480, p < .001 \) than silent reading \( r = .219, p = .040 \). Visual attention span (the number of orthographic units that can be processed visually in a single glance) uniquely predicted only silent reading \( r = .365, p = .001 \). The results generally supported the notion that in L1 readers, oral reading and silent reading fluency measurements are for the most part tapping the same underlying cognitive processes. However, it also makes clear that there are differences between the two modes, and if these differences are detectable in L1 students, they would be even more apparent in L2 adult learners.
Although the connection between oral reading rate and comprehension has been well-established in L1 environments, the studies presented here reveal that the association is weaker, or has not yet been well-supported by empirical evidence, in ESL/EFL readers. It is also likely that the difference between silent reading fluency and oral reading fluency is greater for L2 readers than L1, and this gap is further widened when L2 learners have an L1 orthography that does not correspond as closely to phonological representations as English. However, options are still limited when it comes to measuring reading fluency in the silent mode. In the oral mode, there are more options available for getting a robust and multifaceted picture of reading fluency than in the silent mode. Specifically, in the oral mode accuracy, as measured by words correct per minute, and fluency, measured with a subjective rating scale, can be measured in addition to rate and comprehension, whereas in the silent mode only rate and comprehension can be measured.

**Prosody as a Component of Reading Fluency**

Kuhn et al. (2010) include prosody as one of the three central components (along with accuracy and automaticity) of reading fluency, and this view is adopted in the current study as well. In the L1 reading fluency research, reading fluency is generally studied (and taught) in the oral mode (and in the context of young readers), and in that context the construct of prosody refers to the ability to read (aloud) with expression, intonation, and phrasing that display an understanding of the meaning of the text (Miller & Schwanenflugel, 2008).

The features of prosody, according to Kuhn et al. (2010), are pitch (such as the rising pitch in yes-no questions), duration (such as the length of vowel sounds in stressed
and unstressed words), stress, and pausing. The researchers also note that prosody serves to break up, or parse, continuous speech into meaningful semantic and/or syntactic units for the listener. It can also be used to express emotion.

There is general consensus in the L1 research that prosody is an important construct in the development of reading fluency because it helps readers hold an auditory sequence of text in their working memory while they finish reading and analyzing a sentence or group of sentences (Koriat et al., 2002). It is also predictive of reading fluency development: as a reader’s fluency develops, their pauses will mark longer and more syntactically complex phrases and sentences (Schwanenflugel & Benjamin, 2017). Prosody has also been linked to reading comprehension. Schwanenflugel et al. (2004) showed that prosody predicted reading comprehension, but reading comprehension did not predict prosody. Klaudia and Guthrie (2008), however, found syntactic prosody and reading comprehension to be reciprocally related in fifth-grade students over the course of a year.

The question of whether prosody is also a component of silent reading fluency and comprehension has been taken up in the “implicit prosody” research (McCallum et al., 2004). This is the idea that as young readers progress from oral to silent reading, they take their prosody skill with them and apply it with their inner voice to the act of silent reading. This is still an unresolved area of research, even in the L1 context, but Rasinski et al. (2009) found that oral reading prosody ratings correlated well with silent reading comprehension scores.
Prosody is included as a component in one of the measures of oral reading fluency in the current study. See the previous section for a review of studies on the use of oral versus silent mode to measure reading rate and fluency in L2 contexts.

**Unresolved Issues and Gaps in the Literature**

The first unresolved issue addressed in this study is the need to establish whether ER yields greater reading fluency improvement than IR or other traditional approaches to teaching and learning EFL reading skills. A number of studies with various limitations and drawbacks (Bell, 2001; Huffman, 2014; Robb & Susser, 1989), as well as a few that have managed to avoid many of those limitations (Beglar et al., 2012; McLean & Rouault, 2017; Suk, 2017), have shown (while demonstrating comprehension) that ER yields greater reading rate improvement than IR. This “rate plus comprehension equals fluency” approach to operationalizing the measurement of reading fluency has served this line of research well, but it is not a complete measure of reading fluency. The construct of reading fluency goes beyond speed and accuracy, including prosodic chunking as well (Grabe, 2009; Kuhn & Stahl, 2003), so while there is a need to add to the already growing body of evidence on the effect of ER on reading rate, there is also need to push beyond the rate-plus-comprehension paradigm and measure reading fluency in a more multifaceted way.

The second gap addressed in this study is the fact that the ER and IR approaches have not been sufficiently compared to assess the different effects they have on students’ reading comprehension ability and vocabulary acquisition. The concept of IR as a reading approach has arisen primarily in EFL environments, so few researchers have focused on
the effects of IR. Since IR activities focus specifically on reading comprehension and vocabulary, it seems logical that this approach would yield greater improvement on these measures. A number of studies have researched the effect of ER on reading comprehension and vocabulary acquisition, and many of these compare ER with IR. Regarding the effect on reading comprehension, a number of these studies have found no difference between the effect of ER and IR on reading comprehension (Al-Homoud and Schmitt, 2009; Carney, 2016; Lai, 1993), while others have found a slightly greater effect for ER than IR on reading comprehension (Bell, 2001; Mason & Krashen, 1997; Suk, 2017). Regarding the effect on vocabulary knowledge acquisition, a clear line can be drawn between those that test students only on words they encountered during the texts they read versus those that use more general measures such as the Vocabulary Levels Test. In studies such as Suk (2017), where the vocabulary knowledge measure was drawn from the words encountered in the texts, ER was generally superior to IR in terms of vocabulary acquisition. However, when more general measures have been used, such as in Al-Homoud and Schmitt (2009), the vocabulary knowledge gains of the ER and IR groups tend to be similar. The current study is concerned with the differing effects of ER and IR on overall vocabulary knowledge, and thus falls into the latter category of studies. Overall, few studies have been focused on comparing the effects of ER and IR on reading comprehension and general vocabulary acquisition.

The third unresolved issue concerns the topic of oral versus silent reading rate and fluency measures, and specifically their correlation to each other and to reading comprehension, and whether this differs depending on whether students are reading in their L1 or L2. This issue is of particular importance in reading fluency research because
researchers are confronted with decisions about whether to use silent or oral reading when measuring reading rate and fluency. It is well established that in English L1 environments, oral reading rate highly correlates with reading comprehension (Fuchs et al., 2001). There is also a strong argument to be made that oral reading is a clearer and more accurate way to measure reading fluency because it is impossible to know what is going on inside the mind of a silent reader. In EFL contexts, however, studies on reading rate and fluency often use the silent reading mode. Researchers in these contexts have argued that as long as it can be confirmed that students are comprehending what they have read (e.g., through comprehension questions), reading rate can be considered a suitable proxy for reading fluency. Several researchers have also suggested that oral reading fluency does not correlate highly with EFL reading comprehension (and especially Asian EFL) readers (Jeon, 2012; Jiang et al., 2012; Lems, 2006 & 2012). Of particular interest is the fact that Jiang et al. (2012) compared Chinese and Spanish EFL students, finding that the Chinese students read more slowly but more accurately. Both Jeon (2012) and Jiang et al. argued that the influence of a logographic L1 might result in a phonological coding process that differs from that used by those alphabetic L1s. While this has only been hypothesized by a few researchers—contrasting with the vast volume of data supporting the high correlation of oral reading fluency with reading comprehension in English L1 environments—it does point to a need to investigate how oral reading fluency and silent reading rate are correlated with each other and with reading comprehension in Japanese readers and those in other EFL environments.

The fourth gap addressed in this study relates to the need to understand how reading fluency develops over the course of a semester or a year. This has implications
for curriculum design. For example, if students increase their reading speed rapidly during the first four weeks of a semester, but the improvement levels off after that, teachers could adjust their curriculum accordingly. Chung and Nation (2006), reported results that indicate just that sort of leveling. The participants’ reading rate gains were larger for the first ten texts of a 23-text course lasting nine weeks, and then the rate increases leveled off considerably during the second half of the course. However, approximately one-quarter of the students experienced greater reading rate gains in the second half of the course, and 70% of students increased their rate by more than 12 wpm during the second half, so the authors concluded that the full course was beneficial for most students. Tran (2012), on the other hand, found continuous improvement throughout a 20-text course over a three-month period. However, when looking at the pattern of reading rate change of individual readers, Chung and Nation (2006) discovered a variety of patterns, such as “gradual increase”, “erratic increase”, “plateau increase”, and “mixed increase” (pp. 194-197), and Tran (2012) replicated that finding. Few studies have provided this sort of fine-grained data, so few reliable conclusions can be drawn yet.

Overall, there is a steadily growing body of research on reading rate and fluency development in extensive reading programs, and some of the results are promising, but this is still an underdeveloped area of research, and further empirical evidence could help clarify whether claims regarding the effectiveness of these approaches are warranted. One reason for this gap in the research might be that the connection between ER and reading fluency seems more obvious than with other areas such as vocabulary or comprehension, perhaps because the connection between practice and automaticity is already well-established in L1 reading research (Kuhn & Stahl, 2003; LaBerge & Samuels, 1974;
Nathan & Stanovich, 1991) and in the larger body of skill acquisition research in general (Proctor & Vu, 2006). However, given the fact that teachers and institutions in EFL contexts have shown resistance toward implementing ER programs (Grabe, 2009), there is a need for further empirical studies in this area.

Additionally, what little research there is in EFL contexts on the development of reading fluency tends to focus on reading rate alone as the outcome variable. Because reading fluency is such a complex and multifaceted construct, there is a need for longitudinal studies that measure changes in oral reading rate, accuracy, and fluency, as well as silent reading rate and of course comprehension.

**Purposes of the Study**

The purpose of this study is to address these problems and gaps in the existing literature by investigating whether ER is more effective than IR at improving reading speed and fluency, generate evidence regarding the extent to which ER and IR can be linked to vocabulary acquisition and reading comprehension, examine whether silent or oral reading speed and fluency measurements are more highly correlated to reading comprehension, and provide a fine-grained description of how reading fluency develops during a one-semester ER (with fluency training) course.

The first purpose is to test the claim that ER is more effective in improving reading speed and fluency than IR. A gradually growing number of researchers have investigated this claim (Beglar et al., 2012; McLean & Roualt, 2017; Suk, 2017); however, even these studies have operationalized reading fluency by measure reading rate while including a reading comprehension measurement to confirm that the students
attained a reasonable level of comprehension while reading. In the current study I attempt to add to this body of evidence, but with a fuller, more multifaceted set of measurements of reading fluency. Such evidence tests the claim that ER should be added to existing IR-based EFL reading curricula.

The second purpose is to examine and compare the effect of both ER and IR on vocabulary knowledge and reading comprehension ability. IR approaches tend to include activities (e.g., comprehension questions, word cards, and vocabulary quizzes) designed explicitly to improve reading comprehension ability and vocabulary knowledge acquisition, so proponents of these approaches likely assume that they are better for improving these important reading skills. ER, on the other hand, is often promoted as a way to increase enjoyment and confidence in reading in English (Mori, 2004; Nishino, 2007) rather than as the most effective way to improve reading skills. The few studies that have compared the two, however, have found either a greater effect for ER or no difference. Regarding the effect on reading comprehension, for example, approximately the same number of studies have shown a slightly greater effect for ER (Bell, 2001; Mason & Krashen, 1997; Suk, 2017) or have found no difference (Al-Homoud & Schmitt, 2009; Carney, 2016; Lai, 1993). Turning to the effect on vocabulary knowledge, studies which have assessed general vocabulary knowledge gains (Al-Homoud & Schmitt, 2009) have found the effect of ER and IR to be similar; however, a study that tested only target vocabulary found in the texts read by the students (Suk, 2017) found a larger effect for ER. There is still little evidence in this area, so one purpose of the current study is to further shed light on the extent to which both ER and IR contribute to improved reading comprehension and vocabulary knowledge.
The third purpose of this study is to explore the measurements used to measure reading fluency, and particularly to determine which measurement is most appropriate for adult EFL readers. In L1 environments, oral reading rate has been established as a highly reliable proxy measurement for reading comprehension due to a consistently high correlation between the two (Fuchs et al., 2001). However, researchers in EFL contexts have often opted to use silent reading rate with confirmation of comprehension rather than oral reading rate as a proxy for reading fluency. This is likely due to concern that oral reading rate does not correlate highly with silent reading comprehension in EFL readers, especially where the L1 is a logographic language, such as Japanese. Both Jeon (2012) and Jiang et al. (2012), for example, have posited that speakers of these L1s might have phonological coding processes that are different from those used by L1 speakers of alphabetic languages. Although there is not yet much strong evidence for this proposition, it points to a need to investigate the correlation between oral reading and silent reading rate and fluency, and how they each correlate with reading comprehension, in EFL readers. The data gathered in this study allow for such a comparison.

The fourth purpose is to look more closely at the pattern of reading fluency development in a one-semester ER with fluency training (ERFT) course. Specifically, this study assesses how students’ initial reading rate correlates with the reading rate gains they achieve during the course. Little research has been conducted on this question, so this represents a relatively new area of exploration. It is hypothesized that students who start with a higher reading rate have less room to improve and therefore experience lower gains than those who start at a low reading rate. A final purpose is to provide a detailed picture of the reading rate gains during the ERFT course, regardless of initial reading
rate. A few previous studies have suggested a variety of possible rate change patterns for individual students. Chung and Nation (2006) reported a rapid rate increase during the first one-third of an ER program, with a subsequent plateau effect. However, Tran (2012) reported continuous improvement throughout the course. Both researchers reported a variety of rate change patterns for individual students, so it is possible that individual differences trump any patterns that can be identified. However, because of the paucity of research on this topic, which has potentially high practical relevance for curriculum design, it was included in the purposes of this study.

**Hypotheses**

The hypotheses for this study were as follows:

Hypothesis 1: The ERFT treatment will yield significantly greater gains for the silent reading rate and the oral reading rate and fluency measures, on both simplified and authentic texts, than the IR approach. This would be consistent with previous research on reading rate gains during ER programs (Bell, 2001; Beglar et al., 2012; Huffman, 2014) as well as the two studies that showed transfer effects of such rate gains to authentic texts (Macalister, 2010; Tran, 2012). Regarding reading rate, ER is compared with IR in this study for comparison purposes because IR is the default approach to reading instruction in Japan, rather than because of an expectation that IR should result in reading rate gains. English educators in Japan are generally concerned with whether to supplement IR with ER.
Hypothesis 2: The IR treatment will yield greater reading comprehension gains on the authentic text reading comprehension measure as well as greater written receptive vocabulary size gains, because these areas are where the students spent their time and effort during the semester. However, the ERFT approach was also hypothesized to yield gains here if the connection between reading rate and comprehension held true, and if it transferred from simplified to authentic texts. It was more difficult to hypothesize, however, which approach would yield greater gains on the simplified text reading comprehension measure. On one hand, oral reading fluency moderately correlates to reading comprehension in ESL/EFL contexts (Dhanapala & Yamada, 2014; Jeon, 2012; Lems, 2006 & 2012), so the ER approach would likely yield moderate gains in comprehension on a simplified text, which is the type of text the students read extensively during the ERFT semester. On the other hand, participants focused more explicitly on reading comprehension activities during the IR approach, and it is logical that if their comprehension ability on authentic texts increases, it would also increase on simplified texts.

Hypothesis 3: For the ERFT approach, changes in silent reading rate will correlate more highly with changes in reading comprehension (viewed as the ultimate desired outcome) than changes in oral reading rate and fluency do. The IR approach is not expected, based on the literature reviewed, to result in substantial gains in silent reading rate, so this approach is not included here. This hypothesis is based on the notion that EFL learners whose L1 is a non-alphabetic language are confronted with an undue burden when it comes to decoding and recoding (i.e., pronouncing) words during oral reading, meaning
that silent reading is a truer reflection of their reading ability. This hypothesis is in line with data and ideas found in Lems (2006, 2012), Jeon (2012), and Jiang et al. (2012). It would represent a departure from L1 studies such as van den Boer et al. (2014).

Hypothesis 4: For the ERFT treatment, the initial reading comprehension and silent reading rate of the students will correlate inversely with the magnitude of silent reading rate increase achieved during the semester. The IR approach is not expected, based on the literature reviewed, to result in substantial gains in silent reading rate, so this approach is not included here. This hypothesis was formed because the automatization of the lower-level reading processes involved in word recognition, and therefore reading rate itself, should follow the power law (Logan, 1997): “speed increases throughout practice, but the gains are largest early on and diminish with further practice” (p. 125). Therefore, students who had already progressed to a relatively higher level of reading fluency, as evidenced by a higher reading rate, before the start of the treatment might already have been at a relatively higher point on their individual learning curve, thus making it likely that the additional gains they achieved during this course would not be as high as students who started out at lower levels of fluency.

Hypothesis 5: For the ERFT treatment, a variety of patterns of silent reading rate change throughout the semester are predicted. Because the larger picture of an individual’s learning curve is not visible during a 15-week treatment period, the most common pattern will be a gradual, linear increase. This idea is consistent with both Chung and Nation (2006) and Tran (2012).
CHAPTER 3

METHODS

In this chapter, I describe the participants, context, and instrumentation as well as the design, procedure, and analyses employed in this study.

Participants

The participants were 77 first-year female students who were 19–20 years old attending a private, four-year nursing college in Japan. At the time of admission and the baseline point of this study, the participants had an average English proficiency of 581.33 (SD = 66.13; Median = 582.5) on the CASEC exam (described in the Instrumentation section below), which is equivalent to a TOEIC score of 523.97 (SD = 102.90; Median = 515) or TOEFL-PBT 456.83 (SD = 26.08; Median = 457). The TOEIC (Test of English for International Communication) is a standardized test of English language proficiency specifically designed for EFL/ESL business settings. It consists of a listening and a reading section with one hundred multiple-choice items per section, which combine for a total possible score of 990.

These 77 students were members of an 83-student incoming first-year class at this college, of which 6 students tested out of the required English I course, which is the focus of this study. The study took place during the spring (April through July) and fall (September through January) semesters of the 2016-2017 academic year. The 77 students were divided into four classes based on CASEC scores: a 17-student advanced class and three classes of 19-21 students each, assigned alphabetically by last name. Two classes
engaged in extensive reading with fluency training (ERFT) the first semester and intensive reading (IR) the second semester, while the remaining two classes did the opposite. As a result, there are data for 77 students who participated in the ERFT class and the same 77 students also participated in the IR class; thus, this study was a within-subjects design. The composition of participant classes and groups are displayed in Table 1.

Table 1. Composition of Participant Classes and Groups

<table>
<thead>
<tr>
<th>Class</th>
<th>No. of Students</th>
<th>Proficiency Category</th>
<th>Treatment Order</th>
<th>Average CASEC (TOEIC) Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>17</td>
<td>Advanced</td>
<td>IR-before-ERFT</td>
<td>670.88 (672.06)</td>
</tr>
<tr>
<td>B</td>
<td>21</td>
<td>Non-advanced</td>
<td>IR-before-ERFT</td>
<td>560.14 (486.43)</td>
</tr>
<tr>
<td>C</td>
<td>20</td>
<td>Non-advanced</td>
<td>ERFT-before-IR</td>
<td>554.90 (482.50)</td>
</tr>
<tr>
<td>D</td>
<td>19</td>
<td>Non-advanced</td>
<td>ERFT-before-IR</td>
<td>553.90 (479.00)</td>
</tr>
</tbody>
</table>

The participants had studied English approximately 5–6 hours per week during their six years of secondary education in Japan. They were enrolled in a beginning writing course while in the ERFT course, and an oral communication course focused on nursing situations while in the IR course. All courses met once a week for 90 minutes over the course of a 15-week semester, and I taught both the ERFT and IR courses. Lectures and other course activities were conducted in English. The data analyzed in this study were gathered as part of the educational and assessment activities of the course, and the students were informed that their data would be used in this study and given an opportunity to request that their data be removed from the study. None did so.

Three raters were recruited to conduct the prosody ratings as a measure of oral reading fluency. All three were full-time university EFL faculty in Japan with over ten years of experience teaching at the university level. One was a lecturer, one was an
associate professor, and one a professor. Two of the three had obtained Master’s degrees in TESOL. All three were English native speakers. The raters came to the researcher’s office for an initial one-hour training session, during which the researcher explained the rubric and asked each rater to do a series of ratings and then provided feedback which included the researcher’s own ratings.

**Context of the Study**

The participants were first-year nursing students at a small, private, highly-ranked (2016 Hensachi: 55.0; 2021 Hensachi: 60.0) four-year nursing college. Hensachi is a national, norm-referenced score used to assess the academic level of universities based on how selective they are. Scores range from 20 to 80, with 50 set as the mean and a standard deviation of 10. Approximately 80 new first-year students are accepted each year, with an additional 20 students who already have undergraduate degrees in other fields joining in the second-year. The total number of students in the undergraduate program, therefore, was approximately 380. The school has a long history of providing high-level nursing education in Japan, and it also has Master’s and doctoral programs in nursing. Relative to other nursing colleges in Japan, students have a large number of required and elective courses in the humanities, including eight required English courses, during their first two years, in addition to functional anatomy and nursing courses. From the end of their second year, they begin a rigorous schedule of practical training in nursing skills, both in the lab and on-site at local hospitals and clinics, which continues through the end of the third year. The fourth year sees a slight drop in the course load, with their attention turning to completion of a capstone research project and thesis.
Instrumentation

In order to investigate the five hypothesis of this study as explained at the end of Chapter 2, measures of silent and oral reading rate and fluency, reading comprehension, and vocabulary size were included, in addition to measures needed to ascertain participant characteristics and treatment conditions such as English proficiency, amount read, and time spent reading (time on task). Both silent and oral reading rate were included based on Hypothesis 3: one of the purposes of this study is to investigate whether silent or oral reading rate is more highly correlated with reading comprehension in Japanese EFL readers. The oral reading fluency (prosody) rating was included due to concern that rate measures alone might not measure all aspects of reading fluency, which includes prosody (Kuhn et al., 2010). These measures are listed in Table 2 and explained in detail in this section.

Table 2. List of Measurements

<table>
<thead>
<tr>
<th>Measurement</th>
<th>Units</th>
</tr>
</thead>
<tbody>
<tr>
<td>English proficiency</td>
<td>CASEC score (converted to TOEIC score)</td>
</tr>
<tr>
<td>Amount read (ERFT)</td>
<td>Standard words (exact)</td>
</tr>
<tr>
<td>Amount read (IR)</td>
<td>Standard words (estimated)</td>
</tr>
<tr>
<td>Time spent reading (ERFT)</td>
<td>Hours (time reported by Xreading online system)</td>
</tr>
<tr>
<td>Time spent reading (IR)</td>
<td>Hours (self-reported)</td>
</tr>
<tr>
<td>Silent reading rate</td>
<td>Standard words per minute (self-reported in-classroom)</td>
</tr>
<tr>
<td>Oral reading rate</td>
<td>Standard words correct per minute (timed and adjusted by researcher via audio recording)</td>
</tr>
<tr>
<td>Oral reading fluency</td>
<td>Prosody rating (by researcher and raters, via audio recording), converted to logits</td>
</tr>
<tr>
<td>Reading comprehension</td>
<td>Test score, converted to logits</td>
</tr>
<tr>
<td>Vocabulary size</td>
<td>Number of words known as calculated from test score, converted to logits</td>
</tr>
</tbody>
</table>
English Proficiency

All participants took a standardized general English proficiency test called Computerized Assessment System for English Communication (CASEC) at the beginning of their first year and again at the end of their second year. This test is used both for level-based placement into English classes and to evaluate the students’ English proficiency improvement during the first two years. The CASEC data are used in this study to determine and report the mean English proficiency level of the students.

CASEC (global.casec.com) is a computer adaptive test of English communication developed by the Society for Testing English Proficiency (STEP) in Japan. It uses item-response theory to select items of appropriate difficulty based on the test-taker’s response to previous items. It consists of four parts—vocabulary, conversational expression, listening, and dictation—with 250 points possible on each part, for a total of 1,000 points possible. Based on test-takers’ self-reported TOEIC and TOEFL scores, the test correlates moderately highly with TOEIC ($r = .84$) (based on 19,167 data sets) and moderately with TOEFL ($r = .63$) (based on 2,024 data sets). Results are provided to both test-takers and administration immediately upon completion of the test, and TOEIC and TOEFL reference scores are provided as well.

Amount Read

The amount of text read by each student was recorded in standard words (Carver, 1982), in which each 6-character unit is counted as a word; spaces and punctuation are also counted as characters. Take, for example, the following sentence (not including the quotation mark): “How many standard words is this?” The total number of characters,
including spaces and punctuation, can be automatically counted using online tools such as www.charactercountonline.com. The resulting total, in this case 32, is then divided by six to arrive at the number of standard words (5.33). The length of different books and the number of words on a page varies significantly, and this variation is of particular concern when books written at different levels are compared, and even more so when comparing simplified and authentic texts. For this reason, standard words are a more accurate unit of measurement than actual words. The use of standard words also allows for a more reliable comparison of results across studies (Beglar et al., 2012).

For the ERFT semester, students did all of their reading on the online extensive reading system, Xreading.com. Data on the exact number of standard words in each book were provided by Xreading.com.

For the intensive reading semester, the procedure for estimating the number of standard words in a given text was as follows. First, three full pages of text were randomly selected, and the standard words on those three pages counted and averaged. Then, the number of full pages of text (not including illustrations) in each book was counted at the quarter-page sensitivity level, and the number of full pages in each was then multiplied by the average number of standard words per page for that book (IR). Quarter-page sensitivity level means that if there appeared visually to be only a half-page or quarter page of text on a given page, they were counted as such rather than being counted as a full page.
Time Spent Reading

Time-on-task must be reported as accurately as possible in order to address concerns that any reading fluency development achieved by the ERFT group is simply due to increased time-on-task rather than the different approaches employed. The ERFT group did their reading on the Xreading.com system, so their reading times are reported exactly. When a reader is inactive for longer than two minutes, they are automatically logged out, so the times reported are reasonably accurate. For the IR group, the only way to report time-on-task accurately is to insist that students only read during class time. However, this was considered unreasonable due to the limited amount of time (90 minutes per week) available in class and the need to discuss the meaning of the text and engage in vocabulary study. The most accurate possible method for recording time-on-task, therefore, was to have the students self-report the time they spent reading by keeping a log throughout the semester. For each section of text they were assigned to read, there was a column for them to report the length of time they spent reading the text the first and second time (they are required to read each text twice), as well as additional columns for details regarding their word card use.

Silent Reading Rate

The definition of reading fluency used in this study includes accuracy, automaticity, and prosody (Kuhn et al., 2010). The silent reading rate measurements were intended as an indication of automaticity in silent reading, while also providing an indication of accuracy. This measurement does not, however, provide a picture of readers’ prosody, which can only be ascertained via oral reading.
Each silent reading rate measurement instrument included six texts of approximately 200 standard words each, with accompanying comprehension questions. The first three texts were general-topic narrative fiction from low-level graded readers, written at Flesch-Kincaid grade levels 5.0, 5.4, and 5.8. The second three texts were nursing-related narrative non-fiction from authentic texts, written at Flesch-Kincaid grade levels 9.4, 10.1, and 11.0. All details including lexical profiles generated using the vocabulary profiler at www.lextutor.ca, based on the BNC-COCA-25 list, are presented in Table 3. This table provides information on the difficulty and readability, as well as the length, of the texts used in this measurement. Four unique sets of six texts were used, for the pretest and posttest of the first and second semesters. Each of the six 200-word texts had three accompanying multiple-choice comprehension questions placed on the back side of the paper. These questions focused on details in the text, but they were not intended to be overly challenging. Their primary purpose was to encourage students to attempt to comprehend the text while reading it, rather than just skimming, and also to provide evidence that the students were actually reading the texts. In this case, a large percentage of students answering fewer than half of the questions correctly would throw suspicion on whether the recorded rate accurately reflects their reading rate. The texts and questions are provided in Appendix A.

For the during-semester silent reading rate measurements, administered only during the ERFT semester, one excerpt of approximately 450 standard words from a simplified, low-level graded reader with a Flesch-Kincaid grade level of approximately 5.0 was used for each of three administrations. Six comprehension questions were on the back of the paper. Details regarding these texts are also included in Table 3, and the texts...
and questions are included in Appendix A. Using the rate from only the simplified texts of the pre- and post- administrations, along with the three during-semester measurements, allows for five data points (from the pre-test, the three during-semester rate measurements, and the post-test) to be utilized in examining the pattern of silent reading rate change during the ER course.

Students were also given a practice version of this test one week prior to the pre-course administration in the first semester, in order to become familiarized with the procedure. For each administration, when instructed to begin, students read the first text, and wrote their time when they finished reading by referring to a large stopwatch projected at the front of the classroom. They then turned the paper over and answered the comprehension questions at their own pace. They then waited for the other students to finish, at which time they were instructed to begin reading the second text, which followed the same procedure. The students were encouraged to read as fast as possible while understanding what they were reading. They were also told not to refer back to the text once they had begun answering the comprehension questions, and were monitored carefully on this point.

Results for this measure are reported in standard words per minute. On the pre-course and post-course measurements, the rates on the three simplified texts are combined into a single rate, as are those on the three authentic texts.
Table 3. *Silent Reading Rate Texts*

<table>
<thead>
<tr>
<th>Time 1</th>
<th>Source</th>
<th>Standard words</th>
<th>Grade level</th>
<th>Ease</th>
<th>1,000</th>
<th>2,000</th>
<th>3,000</th>
<th>4,000</th>
<th>5,000-6,000</th>
<th>7,000+</th>
</tr>
</thead>
<tbody>
<tr>
<td>Simplified-1</td>
<td><em>Kidnapped</em></td>
<td>201.83</td>
<td>5.4</td>
<td>82.6</td>
<td>112</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>Simplified-2</td>
<td><em>The Call of the Wild</em></td>
<td>201.17</td>
<td>5.0</td>
<td>84.5</td>
<td>106</td>
<td>6</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Simplified-3</td>
<td><em>Moondial</em></td>
<td>201.67</td>
<td>5.8</td>
<td>79.5</td>
<td>111</td>
<td>8</td>
<td>2</td>
<td>0</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>Authentic-1</td>
<td><em>The Heart’s Truth</em></td>
<td>201.67</td>
<td>11.0</td>
<td>63.4</td>
<td>98</td>
<td>25</td>
<td>7</td>
<td>0</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>Authentic-2</td>
<td><em>The Heart’s Truth</em></td>
<td>200.00</td>
<td>10.1</td>
<td>59.9</td>
<td>98</td>
<td>12</td>
<td>15</td>
<td>7</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>Authentic-3</td>
<td><em>The Heart’s Truth</em></td>
<td>200.5</td>
<td>9.4</td>
<td>67.4</td>
<td>105</td>
<td>13</td>
<td>5</td>
<td>3</td>
<td>8</td>
<td>7</td>
</tr>
<tr>
<td>During-ER</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Simplified-A</td>
<td><em>The Mystery of Allegra</em> †</td>
<td>444.83</td>
<td>5.0</td>
<td>80.5</td>
<td>191</td>
<td>15</td>
<td>7</td>
<td>4</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Simplified-B</td>
<td><em>The Mystery of Allegra</em> †</td>
<td>485.83</td>
<td>4.8</td>
<td>79.2</td>
<td>187</td>
<td>8</td>
<td>4</td>
<td>2</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Simplified-C</td>
<td><em>The Mystery of Allegra</em> †</td>
<td>485.53</td>
<td>5.0</td>
<td>81.5</td>
<td>181</td>
<td>5</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Time 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Simplified-1</td>
<td><em>Rabbit-Proof Fence</em></td>
<td>201.00</td>
<td>5.4</td>
<td>78.4</td>
<td>119</td>
<td>10</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Simplified-2</td>
<td><em>Tales of Mystery and Imagination</em></td>
<td>200.5</td>
<td>5.0</td>
<td>85.0</td>
<td>114</td>
<td>11</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Simplified-3</td>
<td><em>Wyatt’s Hurricane</em></td>
<td>199.5</td>
<td>5.8</td>
<td>76.3</td>
<td>110</td>
<td>12</td>
<td>3</td>
<td>3</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Table 3 continues.

---

70
Table 3 (continued)

<table>
<thead>
<tr>
<th>Time</th>
<th>Book Title</th>
<th>Mean</th>
<th>SD</th>
<th>Median</th>
<th>Median-SD</th>
<th>Corrected Median-SD</th>
<th>Corrected Median-SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>Authentic-1 <strong>The Heart's Truth</strong></td>
<td>199.17</td>
<td>11.0</td>
<td>56.5</td>
<td>100 (76%)</td>
<td>16 (12%)</td>
<td>9 (7%)</td>
</tr>
<tr>
<td></td>
<td>Corrected Median-SD</td>
<td>199.17</td>
<td>11.0</td>
<td>56.5</td>
<td>100 (76%)</td>
<td>16 (12%)</td>
<td>9 (7%)</td>
</tr>
<tr>
<td></td>
<td>Authentic-2 <strong>The Heart's Truth</strong></td>
<td>199.67</td>
<td>10.1</td>
<td>57.5</td>
<td>96 (76%)</td>
<td>12 (9%)</td>
<td>7 (6%)</td>
</tr>
<tr>
<td></td>
<td>Corrected Median-SD</td>
<td>199.67</td>
<td>10.1</td>
<td>57.5</td>
<td>96 (76%)</td>
<td>12 (9%)</td>
<td>7 (6%)</td>
</tr>
<tr>
<td></td>
<td>Authentic-3 <strong>The Heart's Truth</strong></td>
<td>200.67</td>
<td>9.4</td>
<td>57.5</td>
<td>111 (81%)</td>
<td>19 (14%)</td>
<td>3 (2%)</td>
</tr>
<tr>
<td></td>
<td>Corrected Median-SD</td>
<td>200.67</td>
<td>9.4</td>
<td>57.5</td>
<td>111 (81%)</td>
<td>19 (14%)</td>
<td>3 (2%)</td>
</tr>
<tr>
<td></td>
<td>Time 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Simplified-1 <strong>The Mysterious Death of Charles Bravo</strong></td>
<td>200.5</td>
<td>5.4</td>
<td>78.5</td>
<td>123 (90%)</td>
<td>5 (4%)</td>
<td>1 (1%)</td>
</tr>
<tr>
<td></td>
<td>Corrected Median-SD</td>
<td>200.5</td>
<td>5.4</td>
<td>78.5</td>
<td>123 (90%)</td>
<td>5 (4%)</td>
<td>1 (1%)</td>
</tr>
<tr>
<td></td>
<td>Simplified-2 <strong>Moondial</strong></td>
<td>200.00</td>
<td>5.0</td>
<td>79.5</td>
<td>112 (86%)</td>
<td>9 (7%)</td>
<td>1 (1%)</td>
</tr>
<tr>
<td></td>
<td>Corrected Median-SD</td>
<td>200.00</td>
<td>5.0</td>
<td>79.5</td>
<td>112 (86%)</td>
<td>9 (7%)</td>
<td>1 (1%)</td>
</tr>
<tr>
<td></td>
<td>Simplified-3 <strong>Wyatt's Hurricane</strong></td>
<td>200.83</td>
<td>5.8</td>
<td>79.0</td>
<td>104 (83%)</td>
<td>14 (11%)</td>
<td>0 (0%)</td>
</tr>
<tr>
<td></td>
<td>Corrected Median-SD</td>
<td>200.83</td>
<td>5.8</td>
<td>79.0</td>
<td>104 (83%)</td>
<td>14 (11%)</td>
<td>0 (0%)</td>
</tr>
<tr>
<td></td>
<td>Authentic-1 <strong>The Heart's Truth</strong></td>
<td>202.17</td>
<td>11.0</td>
<td>55.3</td>
<td>87 (75%)</td>
<td>16 (14%)</td>
<td>6 (5%)</td>
</tr>
<tr>
<td></td>
<td>Corrected Median-SD</td>
<td>202.17</td>
<td>11.0</td>
<td>55.3</td>
<td>87 (75%)</td>
<td>16 (14%)</td>
<td>6 (5%)</td>
</tr>
<tr>
<td></td>
<td>Authentic-2 <strong>The Heart's Truth</strong></td>
<td>200.5</td>
<td>10.1</td>
<td>56.1</td>
<td>93 (73%)</td>
<td>15 (12%)</td>
<td>10 (8%)</td>
</tr>
<tr>
<td></td>
<td>Corrected Median-SD</td>
<td>200.5</td>
<td>10.1</td>
<td>56.1</td>
<td>93 (73%)</td>
<td>15 (12%)</td>
<td>10 (8%)</td>
</tr>
<tr>
<td></td>
<td>Authentic-3 <strong>The Heart's Truth</strong></td>
<td>199.0</td>
<td>9.4</td>
<td>72.7</td>
<td>117 (84%)</td>
<td>8 (6%)</td>
<td>5 (4%)</td>
</tr>
<tr>
<td></td>
<td>Corrected Median-SD</td>
<td>199.0</td>
<td>9.4</td>
<td>72.7</td>
<td>117 (84%)</td>
<td>8 (6%)</td>
<td>5 (4%)</td>
</tr>
<tr>
<td></td>
<td>Time 4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Simplified-1 <strong>Rabbit-Proof Fence</strong></td>
<td>200.00</td>
<td>5.4</td>
<td>79.8</td>
<td>112 (87%)</td>
<td>10 (8%)</td>
<td>1 (1%)</td>
</tr>
<tr>
<td></td>
<td>Corrected Median-SD</td>
<td>200.00</td>
<td>5.4</td>
<td>79.8</td>
<td>112 (87%)</td>
<td>10 (8%)</td>
<td>1 (1%)</td>
</tr>
<tr>
<td></td>
<td>Simplified-2 <strong>The Last Sherlock Holmes Story</strong></td>
<td>199.33</td>
<td>5.0</td>
<td>80.3</td>
<td>109 (90%)</td>
<td>5 (4%)</td>
<td>3 (2%)</td>
</tr>
<tr>
<td></td>
<td>Corrected Median-SD</td>
<td>199.33</td>
<td>5.0</td>
<td>80.3</td>
<td>109 (90%)</td>
<td>5 (4%)</td>
<td>3 (2%)</td>
</tr>
<tr>
<td></td>
<td>Simplified-3 <strong>On the Edge</strong></td>
<td>200.83</td>
<td>5.8</td>
<td>80.0</td>
<td>126 (93%)</td>
<td>7 (5%)</td>
<td>0 (0%)</td>
</tr>
<tr>
<td></td>
<td>Corrected Median-SD</td>
<td>200.83</td>
<td>5.8</td>
<td>80.0</td>
<td>126 (93%)</td>
<td>7 (5%)</td>
<td>0 (0%)</td>
</tr>
<tr>
<td></td>
<td>Authentic-1 <strong>The Heart's Truth</strong></td>
<td>201.33</td>
<td>11.0</td>
<td>51.4</td>
<td>100 (74%)</td>
<td>18 (13%)</td>
<td>9 (7%)</td>
</tr>
<tr>
<td></td>
<td>Corrected Median-SD</td>
<td>201.33</td>
<td>11.0</td>
<td>51.4</td>
<td>100 (74%)</td>
<td>18 (13%)</td>
<td>9 (7%)</td>
</tr>
<tr>
<td></td>
<td>Authentic-2 <strong>The Heart's Truth</strong></td>
<td>202.17</td>
<td>10.1</td>
<td>54.8</td>
<td>94 (72%)</td>
<td>13 (10%)</td>
<td>16 (12%)</td>
</tr>
<tr>
<td></td>
<td>Corrected Median-SD</td>
<td>202.17</td>
<td>10.1</td>
<td>54.8</td>
<td>94 (72%)</td>
<td>13 (10%)</td>
<td>16 (12%)</td>
</tr>
<tr>
<td></td>
<td>Authentic-3 <strong>The Heart's Truth</strong></td>
<td>201.00</td>
<td>9.4</td>
<td>57.4</td>
<td>106 (76%)</td>
<td>17 (12%)</td>
<td>6 (4%)</td>
</tr>
<tr>
<td></td>
<td>Corrected Median-SD</td>
<td>201.00</td>
<td>9.4</td>
<td>57.4</td>
<td>106 (76%)</td>
<td>17 (12%)</td>
<td>6 (4%)</td>
</tr>
</tbody>
</table>

*Note. † = Oxford Bookworms, Level 3; * = Oxford Bookworms, Level 3.
Oral Reading Rate and Fluency

These measurements reflect all three aspects of reading fluency, according to the definition used in this study, which includes accuracy, automaticity, and prosody (Kuhn et al., 2010). The oral reading rate itself is recorded in words per minute, and the number of words uttered incorrectly are then subtracted to arrive at a rate-plus-accuracy figure, called words correct per minute. Finally, the recordings of the oral text reading were rated on an oral reading fluency scale for the three indicators of prosody: phrasing, intonation, and stress.

Each oral reading rate and fluency measurement instrument included six texts of approximately 40 standard words each, with accompanying comprehension questions. The first three texts were general-topic narrative fiction from low-level graded readers, written at Flesch-Kincaid grade levels of 5.0, 5.3, and 5.9. The second three texts were nursing-related, narrative non-fiction extracts from authentic texts, written at Flesch-Kincaid grade levels of approximately 9.1, 9.8, and 11.4. In order to obtain three texts of different difficulty level, the texts were each taken from a different book. All details, including the lexical profiles generated using the vocabulary profiler at www.lextutor.ca, which is based on the BNC-COCA-25 list, are presented in Table 4. This table provides information on the difficulty and readability, as well as the length, of the texts used in this measurement. Four unique sets of six texts were used for the pretest and posttest in both semesters. For each 40-word text, a single multiple-choice comprehension question was placed on the back of the paper. The three simplified texts were on one sheet of paper, with the three corresponding questions on the back, and the
authentic texts and comprehension questions were on a second sheet of paper. As with the silent reading rate measurement, the comprehension questions focused on details in the text, but were not intended to be overly challenging. Their primary purpose was to motivate students to attempt to comprehend the text while reading it, and also to provide a way of assessing whether the students were actually reading the texts. The texts and questions are provided in Appendix B.

Each student read the three 40-word simplified texts aloud while being audio-recorded using an Olympus Voice-Trek (V-803) voice recorder, then turned over the paper to answer the three comprehension questions regarding those texts. The student then repeated the procedure with the authentic texts and comprehension questions. The procedure was explained to the students one week in advance, and students were encouraged to read as quickly as possible while also understanding the text, and to try to use correct pronunciation while also trying to read smoothly and fluently. They were encouraged to avoid making long pauses if they found a word they were unfamiliar with, but rather to try their best and continue reading.

Results for this measure are reported in standard words per minute, standard words correct per minute (wcpm), and a prosody rating rubric that evaluates phrasing, stress, and intonation. To calculate wcpm, the number of standard words pronounced incorrectly by the student (not including L1-influenced pronunciations that likely indicate accurate comprehension of the word) were subtracted from the total number of words in the passage, and this was be divided by the time taken to read the passage. The prosody rating incorporates elements (phrasing and smoothness) of Zutell and
Rasinski’s (1991) Multidimensional Fluency Scale, but also reflects the features of prosody (stress and intonation) discussed by Kuhn et al. (2010).

The rubric consisted of four ability levels for each of the three features of prosody: phrasing, stress (word- and sentence-level), and intonation. The full rubric with descriptors is shown in Table 5. The results of the three prosody measurements were added together, and the aggregate scores reported in terms of variation in person measures using the Rasch model, as suggested in Beglar (2010).

**Reading Comprehension**

Reading comprehension was measured using six 200-word texts, each followed by four comprehension questions. As with the reading rate and fluency measures, the first three texts were general-topic narrative fiction from low-level graded readers, written at Flesch-Kincaid grade levels of approximately 4.8, 5.3, and 5.8. The last three texts were nursing-related narrative non-fiction from authentic texts, written at Flesch-Kincaid grade levels of approximately 9.1, 9.8, and 11.5. All details including lexical profiles generated using the vocabulary profiler at www.lextutor.ca, based on the BNC-COCA-25 list, are presented in Table 6. This table provides information on the difficulty and readability, as well as the length, of the texts used in this measurement. Each text was followed by a set of 4 multiple choice comprehension questions including two repetitive questions for which the answer can be found with slightly different wording at a single location in the text, one synthesis question for which the answer can be found by synthesizing information from two locations in the text, and one inferencing question.
for which the answer can be found by synthesizing information from one or more
locations in the text with information from the reader’s existing schemata of knowledge
or experience that is not found in the text. These question types are based on those
described in Yano et al. (1994).

Four unique sets of six tests were used for the pretests and posttests of both
semesters. However, for the second, third, and fourth administrations, one of the three
simplified texts and accompanying comprehension questions and one of the three
authentic texts and accompanying comprehension questions were repeated from a
previous administration, so the responses could be anchored in order to equate the four
test forms. For the purpose of reporting reading comprehension results, the three
simplified texts/items were combined, and the three authentic texts/items were
combined, so that a perfect comprehension score on either the simple texts or the
authentic texts is 12 out of 12. The results of the reading comprehension measurement
are reported in terms of variation in person measures using the Rasch model, as
suggested in Beglar (2010). The texts and items are provided in Appendix C.

This test was administered in the classroom setting together with the Vocabulary
Size test at the beginning and end of each semester. Students were given 60 minutes to
complete the combined test. All students were able to complete the test during the
allotted time.
<table>
<thead>
<tr>
<th>Source</th>
<th>Standard words</th>
<th>Flesch-Kincaid Source Standard</th>
<th>Grade level</th>
<th>Ease 1,000</th>
<th>2,000</th>
<th>3,000</th>
<th>4,000</th>
<th>5,000-6,000</th>
<th>7,000+</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Time 1</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Simplified-1</td>
<td>The Mysterious Death of Charles Bravo*</td>
<td>40.5</td>
<td>5.9</td>
<td>78.5</td>
<td>32 (86%)</td>
<td>3 (8%)</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
</tr>
<tr>
<td>Simplified-2</td>
<td>Wyatt’s Hurricane*</td>
<td>40.5</td>
<td>5.3</td>
<td>75.9</td>
<td>26 (72%)</td>
<td>4 (11%)</td>
<td>1 (3%)</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
</tr>
<tr>
<td>Simplified-3</td>
<td>On the Edge*</td>
<td>39.5</td>
<td>5.0</td>
<td>84.0</td>
<td>31 (89%)</td>
<td>1 (3%)</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
</tr>
<tr>
<td>Authentic-1</td>
<td>The Heart’s Truth</td>
<td>39.83</td>
<td>11.4</td>
<td>50.5</td>
<td>25 (74%)</td>
<td>3 (9%)</td>
<td>4 (12%)</td>
<td>0 (0%)</td>
<td>1 (3%)</td>
</tr>
<tr>
<td>Authentic-2</td>
<td>The Heart’s Truth</td>
<td>39.5</td>
<td>9.1</td>
<td>69.3</td>
<td>28 (76%)</td>
<td>4 (11%)</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
<td>1 (3%)</td>
</tr>
<tr>
<td>Authentic-3</td>
<td>The Heart’s Truth</td>
<td>40.33</td>
<td>9.8</td>
<td>59.0</td>
<td>31 (86%)</td>
<td>2 (6%)</td>
<td>0 (0%)</td>
<td>1 (3%)</td>
<td>1 (3%)</td>
</tr>
<tr>
<td><strong>Time 2</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Simplified-1</td>
<td>On the Edge*</td>
<td>40.00</td>
<td>5.3</td>
<td>74.6</td>
<td>33 (87%)</td>
<td>2 (5%)</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
</tr>
<tr>
<td>Simplified-2</td>
<td>Justice*</td>
<td>40.33</td>
<td>5.0</td>
<td>77.33</td>
<td>32 (84%)</td>
<td>1 (3%)</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
</tr>
<tr>
<td>Simplified-3</td>
<td>Moondial*</td>
<td>39.5</td>
<td>5.9</td>
<td>78.1</td>
<td>32 (91%)</td>
<td>1 (3%)</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
</tr>
<tr>
<td>Authentic-1</td>
<td>The Heart’s Truth</td>
<td>41.17</td>
<td>9.1</td>
<td>69.3</td>
<td>29 (74%)</td>
<td>6 (15%)</td>
<td>1 (3%)</td>
<td>2 (5%)</td>
<td>0 (0%)</td>
</tr>
<tr>
<td>Authentic-2</td>
<td>The Heart’s Truth</td>
<td>39.17</td>
<td>9.9</td>
<td>59.6</td>
<td>23 (62%)</td>
<td>5 (14%)</td>
<td>3 (8%)</td>
<td>2 (5%)</td>
<td>1 (3%)</td>
</tr>
<tr>
<td>Authentic-3</td>
<td>The Heart’s Truth</td>
<td>40.5</td>
<td>11.4</td>
<td>50.5</td>
<td>24 (69%)</td>
<td>9 (26%)</td>
<td>1 (3%)</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
</tr>
<tr>
<td><strong>Time 3</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Simplified-1</td>
<td>Rabbit-Proof Fence*</td>
<td>41.0</td>
<td>5.3</td>
<td>84.7</td>
<td>40 (95%)</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
</tr>
<tr>
<td>Simplified-2</td>
<td>Tales of Mystery and Imagination*</td>
<td>40.33</td>
<td>5.9</td>
<td>81.4</td>
<td>40 (98%)</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
</tr>
</tbody>
</table>

Table 4 continues.
<table>
<thead>
<tr>
<th>Time 4</th>
<th>Simplified-3</th>
<th>The Crown of Violet*</th>
<th>40.5</th>
<th>5.0</th>
<th>78.8</th>
<th>25 (74%)</th>
<th>3 (9%)</th>
<th>2 (6%)</th>
<th>1 (3%)</th>
<th>0 (0%)</th>
<th>0 (0%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Authentic-1</td>
<td>The Heart’s Truth</td>
<td>41.33</td>
<td>9.1</td>
<td>56.2</td>
<td>27 (71%)</td>
<td>5 (13%)</td>
<td>2 (5%)</td>
<td>0 (0%)</td>
<td>2 (5%)</td>
<td>1 (3%)</td>
</tr>
<tr>
<td></td>
<td>Authentic-2</td>
<td>The Heart’s Truth</td>
<td>41.5</td>
<td>9.8</td>
<td>48.5</td>
<td>26 (70%)</td>
<td>7 (19%)</td>
<td>3 (8%)</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
</tr>
<tr>
<td></td>
<td>Authentic-3</td>
<td>The Heart’s Truth</td>
<td>40.67</td>
<td>11.4</td>
<td>49.8</td>
<td>27 (77%)</td>
<td>2 (6%)</td>
<td>2 (6%)</td>
<td>2 (6%)</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
</tr>
<tr>
<td></td>
<td>Time 4</td>
<td>Simplified-1</td>
<td>Wyatt’s Hurricane*</td>
<td>40.67</td>
<td>5.9</td>
<td>78.8</td>
<td>31 (82%)</td>
<td>4 (11%)</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
</tr>
<tr>
<td></td>
<td>Simplified-2</td>
<td>Justice*</td>
<td>41.17</td>
<td>5.3</td>
<td>75.9</td>
<td>34 (89%)</td>
<td>2 (5%)</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
</tr>
<tr>
<td></td>
<td>Simplified-3</td>
<td>Moondial*</td>
<td>40.00</td>
<td>5.0</td>
<td>88.3</td>
<td>40 (91%)</td>
<td>2 (5%)</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
</tr>
<tr>
<td></td>
<td>Authentic-1</td>
<td>The Heart’s Truth</td>
<td>41.00</td>
<td>9.1</td>
<td>52.6</td>
<td>31 (86%)</td>
<td>3 (8%)</td>
<td>1 (3%)</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
</tr>
<tr>
<td></td>
<td>Authentic-2</td>
<td>The Heart’s Truth</td>
<td>40.67</td>
<td>9.8</td>
<td>63.3</td>
<td>28 (78%)</td>
<td>1 (3%)</td>
<td>0 (0%)</td>
<td>2 (6%)</td>
<td>2 (6%)</td>
<td>0 (0%)</td>
</tr>
<tr>
<td></td>
<td>Authentic-3</td>
<td>The Heart’s Truth</td>
<td>40.17</td>
<td>11.3</td>
<td>47.2</td>
<td>25 (69%)</td>
<td>6 (17%)</td>
<td>2 (6%)</td>
<td>1 (3%)</td>
<td>1 (3%)</td>
<td>0 (0%)</td>
</tr>
</tbody>
</table>

*Note.* Level 3; * = Oxford Bookworms, Level 3
Table 5. Oral Reading Prosody Rubric

<table>
<thead>
<tr>
<th>Phrasing</th>
<th>Stress</th>
<th>Intonation</th>
</tr>
</thead>
<tbody>
<tr>
<td>4 = Native-like</td>
<td>Student reads in clause and sentence units; phrase boundaries are meaningful and sound smooth and natural, or nearly so, to a native speaker.</td>
<td>Student shows understanding of the text by using appropriate stress at or near a native-speaker level.</td>
</tr>
<tr>
<td>3 = Good</td>
<td>Student reads in short clauses and attempts meaningful phrasing, but with some choppiness and unnaturalness.</td>
<td>Student makes use of stress, but makes occasional errors.</td>
</tr>
<tr>
<td>2 = Fair</td>
<td>Student uses some two- or three-word phrasing, but the overall effect is rather choppy.</td>
<td>Student’s use of stress is quite infrequent or highly inaccurate.</td>
</tr>
<tr>
<td>1 = Poor</td>
<td>Student reads word by word, with little or no phrasing.</td>
<td>Student does not make use of stress when reading.</td>
</tr>
</tbody>
</table>

Vocabulary Size

In order to test the hypothesis that the IR approach would result in greater gains in vocabulary size, which is highly correlated with reading comprehension, written receptive vocabulary size was measured using an adapted version of the Vocabulary Size Test (Beglar, 2010; Nation & Beglar, 2007). The original version consists of 140 items, ten representative items each from the 1st to the 14th 1,000-word frequency lists. However, due to administration time constraints, it was necessary to use a shorter version that was more specifically targeted to the students in this study. Recent previous research with students at the same college (Huffman, 2014) indicated that they had a mean vocabulary size of over 3,300 words ($SD = 430$), so 99.7% of students were
## Table 6. Reading Comprehension Texts

<table>
<thead>
<tr>
<th>Source</th>
<th>Standard words</th>
<th>Grade level</th>
<th>Ease</th>
<th>1,000</th>
<th>2,000</th>
<th>3,000</th>
<th>4,000</th>
<th>5,000-6,000</th>
<th>7,000+</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Time 1</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Simplified-1</td>
<td><em>Rabbit-Proof Fence</em></td>
<td>198.67</td>
<td>5.8</td>
<td>76.4</td>
<td>107 (84%)</td>
<td>6 (5%)</td>
<td>2 (2%)</td>
<td>0 (0%)</td>
<td>3 (2%)</td>
</tr>
<tr>
<td>Simplified-2</td>
<td><em>The Crown of Violet</em></td>
<td>201.33</td>
<td>5.3</td>
<td>74.4</td>
<td>108 (84%)</td>
<td>5 (4%)</td>
<td>1 (1%)</td>
<td>2 (2%)</td>
<td>0 (0%)</td>
</tr>
<tr>
<td>Simplified-3</td>
<td><em>Justice</em></td>
<td>202.83</td>
<td>4.6</td>
<td>78.7</td>
<td>117 (84%)</td>
<td>8 (6%)</td>
<td>2 (1%)</td>
<td>1 (1%)</td>
<td>0 (0%)</td>
</tr>
<tr>
<td>Authentic-1</td>
<td><em>The Heart’s Truth</em></td>
<td>198.33</td>
<td>11.7</td>
<td>57.2</td>
<td>97 (68%)</td>
<td>13 (9%)</td>
<td>9 (6%)</td>
<td>6 (4%)</td>
<td>8 (6%)</td>
</tr>
<tr>
<td>Authentic-2</td>
<td><em>The Heart’s Truth</em></td>
<td>201.33</td>
<td>9.1</td>
<td>64.2</td>
<td>96 (74%)</td>
<td>16 (12%)</td>
<td>8 (6%)</td>
<td>1 (1%)</td>
<td>3 (2%)</td>
</tr>
<tr>
<td>Authentic-3</td>
<td><em>The Heart’s Truth</em></td>
<td>201.00</td>
<td>9.5</td>
<td>59.8</td>
<td>99 (73%)</td>
<td>9 (7%)</td>
<td>15 (11%)</td>
<td>0 (0%)</td>
<td>1 (1%)</td>
</tr>
<tr>
<td><strong>Time 2</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Simplified-1</td>
<td><em>The Mysterious Death of Charles Bravo</em></td>
<td>201.33</td>
<td>4.9</td>
<td>76.9</td>
<td>109 (83%)</td>
<td>12 (9%)</td>
<td>4 (3%)</td>
<td>2 (2%)</td>
<td>0 (0%)</td>
</tr>
<tr>
<td>Simplified-2</td>
<td><em>The Crown of Violet</em></td>
<td>201.33</td>
<td>5.3</td>
<td>74.4</td>
<td>108 (84%)</td>
<td>5 (4%)</td>
<td>1 (1%)</td>
<td>2 (2%)</td>
<td>0 (0%)</td>
</tr>
<tr>
<td>Simplified-3</td>
<td><em>The Last Sherlock Holmes Story</em></td>
<td>202.33</td>
<td>5.8</td>
<td>70.8</td>
<td>106 (79%)</td>
<td>18 (13%)</td>
<td>2 (1%)</td>
<td>1 (1%)</td>
<td>0 (0%)</td>
</tr>
<tr>
<td>Authentic-1</td>
<td><em>The Heart’s Truth</em></td>
<td>198.33</td>
<td>11.7</td>
<td>57.2</td>
<td>97 (68%)</td>
<td>13 (9%)</td>
<td>9 (6%)</td>
<td>6 (4%)</td>
<td>8 (6%)</td>
</tr>
<tr>
<td>Authentic-2</td>
<td><em>The Heart’s Truth</em></td>
<td>198.83</td>
<td>10.0</td>
<td>54.8</td>
<td>91 (71%)</td>
<td>15 (12%)</td>
<td>14 (11%)</td>
<td>0 (0%)</td>
<td>2 (2%)</td>
</tr>
<tr>
<td>Authentic-3</td>
<td><em>The Heart’s Truth</em></td>
<td>200.17</td>
<td>9.3</td>
<td>67.8</td>
<td>111 (84%)</td>
<td>10 (8%)</td>
<td>2 (2%)</td>
<td>4 (3%)</td>
<td>2 (2%)</td>
</tr>
<tr>
<td><strong>Time 3</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Simplified-1</td>
<td><em>Kidnapped</em></td>
<td>202.00</td>
<td>4.7</td>
<td>85.6</td>
<td>116 (90%)</td>
<td>5 (4%)</td>
<td>1 (1%)</td>
<td>0 (0%)</td>
<td>3 (2%)</td>
</tr>
<tr>
<td>Simplified-2</td>
<td><em>The Call of the Wild</em></td>
<td>201.00</td>
<td>5.3</td>
<td>86.0</td>
<td>101 (84%)</td>
<td>10 (8%)</td>
<td>2 (2%)</td>
<td>0 (0%)</td>
<td>2 (2%)</td>
</tr>
<tr>
<td>Simplified-3</td>
<td><em>The Last Sherlock Holmes Story</em></td>
<td>202.33</td>
<td>5.8</td>
<td>70.8</td>
<td>106 (79%)</td>
<td>18 (13%)</td>
<td>2 (1%)</td>
<td>1 (1%)</td>
<td>0 (0%)</td>
</tr>
</tbody>
</table>

Table 6 continues.
<p>| | | | | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Table 6 (continued)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Authentic-1</strong></td>
<td><em>The Heart's Truth</em></td>
<td>200.67</td>
<td>11.3</td>
<td>55.3</td>
<td>105 (78%)</td>
<td>18 (13%)</td>
<td>3 (2%)</td>
<td>4 (3%)</td>
<td>2 (1%)</td>
</tr>
<tr>
<td><strong>Authentic-2</strong></td>
<td><em>The Heart's Truth</em></td>
<td>198.83</td>
<td>10.0</td>
<td>54.8</td>
<td>91 (71%)</td>
<td>15 (12%)</td>
<td>14 (11%)</td>
<td>0 (0%)</td>
<td>2 (2%)</td>
</tr>
<tr>
<td><strong>Authentic-3</strong></td>
<td><em>The Heart's Truth</em></td>
<td>200.33</td>
<td>9.1</td>
<td>63.4</td>
<td>99 (79%)</td>
<td>13 (10%)</td>
<td>5 (4%)</td>
<td>1 (1%)</td>
<td>2 (2%)</td>
</tr>
<tr>
<td><strong>Time 4</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Simplified-1</strong></td>
<td><em>The Crown of Violet</em></td>
<td>202.33</td>
<td>4.6</td>
<td>81.4</td>
<td>115 (86%)</td>
<td>8 (6%)</td>
<td>0 (0%)</td>
<td>2 (1%)</td>
<td>1 (1%)</td>
</tr>
<tr>
<td><strong>Simplified-2</strong></td>
<td><em>The Call of the Wild</em></td>
<td>201.00</td>
<td>5.3</td>
<td>86.0</td>
<td>101 (84%)</td>
<td>10 (8%)</td>
<td>2 (2%)</td>
<td>0 (0%)</td>
<td>2 (2%)</td>
</tr>
<tr>
<td><strong>Simplified-3</strong></td>
<td><em>Tales of Mystery and Imagination</em></td>
<td>201.00</td>
<td>5.7</td>
<td>77.8</td>
<td>126 (89%)</td>
<td>8 (6%)</td>
<td>1 (1%)</td>
<td>3 (2%)</td>
<td>0 (0%)</td>
</tr>
<tr>
<td><strong>Authentic-1</strong></td>
<td><em>The Heart's Truth</em></td>
<td>201.17</td>
<td>11.6</td>
<td>52.4</td>
<td>94 (73%)</td>
<td>17 (13%)</td>
<td>11 (9%)</td>
<td>2 (2%)</td>
<td>0 (0%)</td>
</tr>
<tr>
<td><strong>Authentic-2</strong></td>
<td><em>The Heart's Truth</em></td>
<td>201.5</td>
<td>9.8</td>
<td>55.3</td>
<td>99 (73%)</td>
<td>19 (14%)</td>
<td>10 (7%)</td>
<td>2 (1%)</td>
<td>1 (1%)</td>
</tr>
<tr>
<td><strong>Authentic-3</strong></td>
<td><em>The Heart's Truth</em></td>
<td>200.33</td>
<td>9.1</td>
<td>63.4</td>
<td>99 (79%)</td>
<td>13 (10%)</td>
<td>5 (4%)</td>
<td>1 (1%)</td>
<td>2 (2%)</td>
</tr>
</tbody>
</table>

*Note. * = Oxford Bookworms, level 3
expected to fall between 2,010 and 4,590 words. Due to the entrance exam system in Japan, students at a given university tend to have proficiency levels that vary little from year to year. Most students aim to enter a school whose examination they are reasonably sure they can pass, the examination system usually prevents students from entering universities above their academic ability level, and few students aim to enter universities ranked below their own estimated ability level. Even given some yearly variability, it was therefore considered unlikely that students would know a substantial number of words above the 7th 1,000-word level. The shortened version of the test used in this study, therefore, consisted of 5 representative items from each of the first 1,000 words through the seventh 1000-word lists, for a total of 35 items.

Four unique forms of this adjusted VST were used, for the pretest and posttest in both semesters. However, for the second, third, and fourth administrations, 10 of the 35 items were repeated from a previous administration, so the responses could be anchored in order to equate the four test forms. The items are provided in Appendix D.

This adapted VST was administered along with the reading comprehension test at the beginning and end (weeks 1 and 13 or 14) of each semester. Students had 60 minutes to complete the combined test. The amount of time was determined based on students’ performance the previous year, so all students had enough time to finish the test.

The results of the VST measurement are reported in terms of variation in person measures using the Rasch model, as suggested in Beglar (2010).
Research Design

This study was a quasi-experimental within-subjects (repeated-measures) design using intact classes. All students participated in the ERFT treatment and the IR treatment for one semester each, so when ERFT and IR results are compared in this study, it is the results of the same students’ performance during each treatment that are being compared. Field (2009) noted that a repeated-measures design introduces less unsystematic variation than a between-subjects design, thus reducing error variation and making the effect of the experimental manipulation more apparent. The reason for this is that in both designs, there is unsystematic variation due to unknown factors affecting how a given participant behaves or performs at a given time, but only in between-subjects design is there additional variation due to differences in the characteristics of the participants themselves. That variation is eliminated in a within-subjects design.

The duration of the data collection period was one academic year, but because the groups switched treatments after the first semester, the treatment period for each approach (ERFT and IR) was one 15-week semester (students started reading from the second week of the semester). Measures relating to reading rate, fluency, comprehension, and vocabulary size were administered at the beginning and end of each semester, while the measure of general English proficiency was administered once at the beginning of the school year. All measures were administered during class time as a part of the regular educational activities of the course. For the measurement of oral reading rate and fluency, students recorded their oral readings individually in my office.
Most research on reading comprehension and reading fluency consists of one-time correlational studies, so it is not possible to see how fluency develops over time, nor how different instructional approaches impact fluency development in different ways. This design represents a much-needed addition to the body of literature in that it allows for the comparison of two instructional approaches and also provides a picture of the development of reading fluency within individual learners over the course of one academic semester. However, the longitudinal nature of this study presents problems of its own, particularly in that it introduces a wide variety of uncontrollable and unknowable outside influences. Of particular concern is the fact that half of the students began the ERFT semester in the second semester, after having already gone through the IR course and in some cases studying abroad during the summer holiday. Of the 77 students in this study, 16 participated in a three-week study abroad program in Canada, so this might have increased the baseline English proficiency of those students for the second semester. The design and timeline of the measurements and treatments is summarized in Table 7.

**Procedures**

**Extensive Reading with Fluency Training**

During the ERFT semester, students read from a large selection of graded readers of various genres (mostly fiction) from the Cengage Foundations, Cengage Page Turners, ELI Readers, Oracle Readers, MPI Building Blocks Library, Oxford Bookworms, Cambridge English Readers, Cambridge Experience Readers, Young...
Learners Classic Readers, and Macmillan Readers series. The students chose and read their own books online using the Xreading.com system. They were restricted to books at Xreading.com’s Level 4 (301–400 headwords, corresponding to Cambridge Level 1) and above. Students were required to start at Levels 4, 5, or 6 because levels lower than this were considered to be too easy for even the lowest-proficiency students at this university, and they were required to read at least one book and not more than six books at each level before moving up to the next level. While generally moving to more difficult texts, they were also allowed to take occasional breaks at their own discretion, reading a lower-level book after they had already moved up to a higher level. The books that counted as breaks accounted for no more than 20% of all the books they read during the semester. The result of this system was that most students started at Level 4 or 5, and few students got much higher than Level 6 or 7 by the end of the semester. For reference, the Flesch-Kincaid readability for the Oxford Bookworms series ranges from Grade 1.1 through 3.9 (Beglar et al., 2012). The students were strongly encouraged to read at least one book per week, and they were required to read at least one book during each two-week period. This approach ensured that the participants continued reading throughout the course of the semester. They were also required to read at least 80,000 tokens during the semester in order to pass the course. This was set as the pass-fail line, so it was intended to be near the minimum of what could qualify as ER. They knew that their grade for the course was heavily weighted on amount read, and that their grade would increase from the minimum for every additional 10,000 tokens between 80,000 and 120,000 tokens. The number of standard words per book varied widely, but as a
point of reference, Cambridge’s *Bad Love* is a Cambridge Level 1 book (which is Level 4 in Xreading, the lowest level allowed in the ERFT class in this study) has a standard word count of 3,586.67. The number of tokens is 4,031, so students would have to have read 20 books of this length to pass the course (80,000 tokens), or 30 books to get the

Table 7. *Timeline of Measurements and Treatments*

<table>
<thead>
<tr>
<th>Week</th>
<th>ERFT (Classes A &amp; B)</th>
<th>IR (Classes C &amp; D)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>RC &amp; V (60 min.), practice SRR (15 min.)</td>
<td>RC &amp; V (60 min., practice SRR (15 min.)</td>
</tr>
<tr>
<td>2</td>
<td>SRR (30 min.), course explanation (60 min.)</td>
<td>SRR (30 min.), course explanation (60 min.)</td>
</tr>
<tr>
<td>3</td>
<td>ORF (90 min.)</td>
<td>ORF (90 min.)</td>
</tr>
<tr>
<td>4</td>
<td>ER/IR lecture (60 min.), Explain book reports (30 min.)</td>
<td>Discussion and checking answers to comprehension questions on assigned readings (70 min.); intensive vocabulary study using word cards (15 min. for in-class Q&amp;A); nursing/medical vocabulary quizzes (5 min.)</td>
</tr>
<tr>
<td>5</td>
<td>ORF training session 1 (90 min.), book reports due</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>SRR 1 (15 min.), book familiarization activity (45 min.), free silent reading time (30 min.)</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>Activity (60 min.), free silent reading time (30 min.)</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>ORF 2 (90 min.), book reports due</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>SRR 2 (15 min.), Review ER/IR (45 min.), free silent reading time (30 min.)</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>Activity (60 min.), free silent reading time (30 min.)</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>ORF 3 (90 min.), book reports due</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>SRR 3 (15 min.), activity (45 min.), free silent reading time (30 min.)</td>
<td>Presentation (90 min.)</td>
</tr>
<tr>
<td>13</td>
<td>RC &amp; V (60 min.), free silent reading time (30 min.)</td>
<td>Final Exam (90 min.)</td>
</tr>
<tr>
<td>14</td>
<td>SRR (30 min.), Final Exam (30 min.)</td>
<td>RC&amp;V (60 min.), SRR (30 min.)</td>
</tr>
<tr>
<td>15</td>
<td>ORF (90 min.)</td>
<td>ORF (90 min.)</td>
</tr>
</tbody>
</table>

8-week summer holiday

*Note.* RC&V = reading comprehension and vocabulary size test; SRR = silent reading rate measurement; ORF = oral reading rate and fluency measurement or training session; USSR = uninterrupted, sustained, silent reading. For the second semester, the classes switched between ERFT and IR, and the same timeline was followed again. The practice SRR, however, was not conducted again at the beginning of the second semester.
highest possible grade in the course (120,000 tokens). Xreading.com provides a five-item multiple-choice quiz at the end of each book, and students were only given credit for having read a book if they correctly answered four or five of the questions. Students were able to do their reading and quizzes on Xreading.com either on their own time or during free silent reading sessions in class, so they were not supervised when answering the quiz questions.

Xreading.com also provides data on how much time a student spends reading a book and, based on that figure, how fast they read. As an additional confirmation that the students were reading the books, their reading times and reading speeds were monitored periodically. When irregular reading speeds (e.g., above 400 wpm) were detected, I talked to them directly and reminded them that they need to read all of the text and understand it while reading. When continuing irregularities were noticed, books exceeding 400 wpm were no longer counted as read (this only occurred on three occasions).

As a third and final way of ensuring that the students were reading the books, each student had to write a book report including a summary of the story for three books during the semester. These book reports were assigned by me on an impromptu basis within 10 days after the student had finished the book, so that students would not know in advance which book they would have to write a report on. If the summary portion of a book report did not adequately display that the student had read and understood the book, the student was given a chance to rewrite the report, and if the problem was still
not rectified, the book was not counted. This happened a total of 22 out of 234 times (9.4%).

Students in the ERFT class spent class time engaged in the following activities: (a) instruction and discussion of the ER approach, (b) book familiarization activities, (c) silent timed readings (speed reading), (d) oral reading fluency training sessions, and (e) silent free reading time. Lectures and discussions regarding the ER approach were conducted during the first two class sessions, and the students were evaluated on their knowledge of this information in the final course exam. The content focused on the difference between ER and IR as well as the purpose, method, and expected benefits of each approach. All lectures and other activities in the ERFT class were conducted in English, except to occasionally ensure that students understood complex but important rules or assignments.

The book familiarization activities, which included simple, enjoyable communicative activities as well as a poster contest, were designed to get students talking to each other in English about the books they have read, as well as to create a positive and communal atmosphere while getting to know what books were available.

The series of three silent timed readings were spread evenly throughout the semester, and the procedure for these were the same as for the pre-semester and post-semester silent reading rate measurements. These were intended to serve both an instructional purpose (raising students’ awareness of the importance reading rate and encouraging them to monitor their reading rate) and a research purpose (measuring their
reading rate). The texts used for these were described in detail in the Silent Reading Rate subsection of the Instrumentation section above.

The three oral reading fluency training sessions were also spread across the semester, and these were modeled after Rasinski et al.’s (1994) fluency development lesson. Students were informed that the ultimate goal of the exercise was to improve their silent reading fluency and thereby improve their reading comprehension as well. The texts used for these sessions were excerpts from Oxford Bookworm’s *The Mystery of Allegra* (Level 2), each text being around 300 standard words in length. Students were given a copy of the text and first listened to my oral reading of the text, three times. While listening, they were encouraged to notice and mark any observations or questions related to oral reading fluency, with the goal being to become a more fluent reader. They then shared their observations with a partner, followed by a whole-class discussion of these issues. The first oral reading fluency training session was done without telling the students what the important features of oral reading fluency are, but rather letting them notice and discuss this organically. (They noticed features such as rising/falling intonation, pauses, word stress and the dropping of certain consonants or syllables, how I slowed down or speeded up at times, expressiveness, and volume changes.) In the second and third sessions they were instructed to focus specifically on pronunciation, stress, intonation, phrasing, and speed.) Next, I read each sentence and the students repeated. This often involved breaking the sentence into meaningful phrases and pointing out important issues for oral reading fluency improvement. Next, the whole class did a final choral reading of the text, and then students got into pairs and
read the text out loud, sentence by sentence, while coaching and giving feedback to each other. Finally, a number of students performed a solo oral reading of a few sentences for the class, with brief teacher feedback and coaching following. Approximately one-third of students did a performance for each oral reading fluency training session, so that every student had a chance to do it once during the semester. Each of the three sessions took approximately 90 minutes to complete.

The final component of the ERFT course was silent reading. The students often had 20–30 minutes of silent reading time at the end of each 90-minute course session. The reading was conducted online using Xreading.com. They were also allowed to use this time to write book reports.

**Intensive Reading**

During the IR semester, students read 20,258.32 standard words of authentic text from three books written for adult native-speaker readers. The books—*The Heart’s Truth: Essays on the Art of Nursing*, *Chicken Soup for the Nurse’s Soul*, and *Chicken Soup for the Nurse’s Soul, Second Dose*—consisted of essays and short stories related to nursing experiences. The students read authentic nursing-related texts for two reasons. The first was that the end goal of their college-level education in English at a nursing college is to improve their reading skill with these types of texts, and ultimately even to be able to read international research articles in their field in English. The second reason was that texts at this difficulty level and in this genre are rich in unknown vocabulary, medical and nursing-related vocabulary, and difficult grammar, so it provided rich
opportunities to expand their interlanguage, which does not occur as much if they are reading a text that matches their current level. All students read the same texts.

During class, the students engaged in group work as well as whole-class teacher-led discussions as they worked through a set of content-oriented comprehension questions related to each assigned reading. The content of the texts, as clarified in these comprehension questions and discussions, was the focus of a large part of their final exam, giving them motivation to understand the content.

In the IR course, the students were also required to make word cards and engage in intensive study of the vocabulary encountered in the text. Specifically, they were given a list of vocabulary (average = 165 words) from each of five reading assignments. It should be noted that four of these words (weep, scrub, unravel, decent) also appeared on one or more of the Vocabulary Size Test forms used in this study. Students were instructed to find each word as they read through the text and determine whether or not they already knew that word as used in that context. If not, they were instructed to make a word card for it. All students were required to make at least 25 word cards per assignment, for a total minimum of 125 cards during the semester. The students were informed that any word from the lists could appear on the vocabulary section of the final exam, so they were motivated to learn all of the words they did not already know to the best of their ability. They were instructed to write an English definition and example sentence using the word on the back of each card (they were allowed to include an optional Japanese translation as well), and they were instructed to go through each set of cards every day until they could recall the definitions and usages for at least
90% of them. At that point, they were advised to use space repetition by waiting for 3-4 days and then going through them again, and then waiting a week before going through them again. There was also a Vocabulary Question Time each class so that the students could ask for an explanation of words they had difficulty understanding on their own.

Students made and studied their word cards outside of class, and their short-term motivation was assumed to be mainly the final exam, so although there was a productive aspect to this component of the course (students were encouraged to create their own original example sentences on the back of the cards), it focused primarily on receptive vocabulary knowledge.

The course grade was based on in-class participation, word card checks, a presentation, and a final exam. The presentation and the final examination served as high-stakes incentives for the students to read and understand the content of the texts, and the word card checks and the final examination fulfilled the same purpose with regards to vocabulary study. There were also a series of vocabulary quizzes covering medical and nursing vocabulary; these words did not come from the reading texts. The final exam items were approximately 50% vocabulary knowledge, 25% content knowledge of the texts they had read during the course, and 25% reading comprehension ability using the texts they had read during the course.

The focus on content, accuracy, comprehension, and explicit vocabulary study in this was in line with Nation’s (2009a) description of IR, but it did not include explicit instruction of reading strategies and skills, which is often included as a component of IR courses (see Chapter 2).
Data Analysis

Hypothesis 1: ER with fluency training (ERFT) will yield greater gains than IR in silent and oral reading rate and fluency, for both simplified and authentic texts.

Because all participants underwent both treatments, six repeated-measures ANOVA were conducted, with group (ERFT and IR) as the independent variable and silent reading rate, oral reading rate, and oral reading fluency as the dependent variables.

Hypothesis 2: IR will yield greater gains than ERFT in reading comprehension on the authentic text and vocabulary size, but the gains in reading comprehension on the simplified text will be similar for ERFT and IR.

A repeated-measures ANOVA was conducted with group (ERFT and IR) as the independent variable and reading comprehension and vocabulary size as the dependent variables.

Hypothesis 3: For the ERFT treatment, changes in silent reading rate will correlate more highly with changes in reading comprehension than changes in oral reading rate and fluency.

Pearson product-moment correlation coefficient tables showing the correlations between and among silent reading rate change, reading comprehension change, oral reading rate change, and fluency change were generated.
Hypothesis 4: For the ERFT treatment, initial reading comprehension ability and initial reading rate will correlate negatively with the silent reading rate change during the semester.

Pearson product-moment correlation coefficient tables showing the correlations between and among pre-course silent reading rate, pre-course reading comprehension, and silent reading rate change during the semester were generated.

Hypothesis 5: For the ERFT treatment, a variety of patterns of silent reading rate change over the course of the semester will be identified, with a gradual increase being the most common pattern.

Including the pre- and post-course measurements, there were a total of five silent reading rate measurements (simplified text) over the course of the 15-week treatment period. To analyze the pattern of change over the treatment period, rate change graphs for all participants were examined. Data-driven categories of change patterns were identified and defined, and the number of participants in each pattern category were tallied and reported.

Rasch analysis was conducted in order investigate gather validity evidence and assess the reliability of the tests and rating rubrics used in this study. This analysis entails calculating the degree to which the observed data fit an a priori model. These validity and reliability analyses are described and presented in the next chapter.
In addition to instrument validation, Rasch analysis was also used to create interval-scale measures of student performance from the raw scores obtained from each instrument described in the previous section. These scaled measures, rather than the raw scores, were used when conducting the statistical analyses presented in Chapter 5.
In this chapter I examine the validity and reliability of the rating scale used for the oral reading prosody measurement and the two dichotomous tests used to measure reading comprehension and vocabulary size. These analyses also provide the interval measures to be used in place of raw scores in the statistical analyses in the next chapter.

In order to gather validity evidence and estimate the reliability of the oral reading prosody ratings, and also to generate Rasch ability estimates to be used in later analyses, I conducted multifaceted Rasch analyses using FACETS (Linacre, 2011a). An example of the command file used is presented in Appendix E. For each analysis, a Wright map is first produced in order to visualize the students’ performances (ability), the strictness of the raters, and the difficulty of each prosody criterion (phrasing, stress, and intonation), all on a single scale using a common unit of measurement (Rasch logits). Next, the infit and outfit statistics of the raters are reported, which indicate how well their ratings fit to the Rasch model. As an acceptable criterion range for item fit, I used 0.40–1.20, a range proposed by Wright et al. (1994) for judged ratings when agreement is encouraged. Estimates of person and item separation and reliability are also reported. For evaluation of separation and reliability of the rating scale instrument, the criteria suggested by Fisher (2007) were used. Fisher’s criteria for person and item measurement reliability are as follows: Poor: <.67; Fair: .67–.80; Good: .81–.90; Very Good: .91–.94; and Excellent: >.94. Fisher’s criteria for person and item separation are
as follows: Poor: 2 or less; Fair: 2–3; Good: 3–4; Very Good: 4–5; and Excellent: >5.
Similar statistics for the rating scale criteria are also reported, followed by an analysis of how well the rating scale levels functioned. Because separate analyses were conducted for each of simplified and authentic texts as well as each of four time points, the rating scales in the second, third, and fourth times were anchored to that of the first time by inputting the Rasch-Andrich thresholds measures from the first time a priori.

Rasch-Andrich thresholds are also reported for the prosody ratings as an indication of whether the four levels of the rating scale are clearly distinguished. Wolfe and Smith (2007) reported that the minimum separation for four-point rating scale thresholds is 1.10, so this criteria is adopted here.

In order to explore the reliability and validity of the dichotomous tests, I used Winsteps version 3.73 (Linacre, 2011b) software to assess the degree to which the data fit the Rasch model. An example of the command file used is presented in Appendix F. Lack of fit to the model can indicate that students are not answering items seriously, or alternatively that the items themselves are flawed or biased in some way. These analyses can also indicate whether the test measures a single, unidimensional construct.

The first statistics reported for each set of items are measures of misfit: infit and outfit. These fit statistics indicate the degree of unexpected responses. For example, if a large number of high-ability students who answered most items correctly suddenly answered an easy question incorrectly, the fit statistic alerts researchers to look closely at the question and correct response. Rasch analysis sets 1.00 as the expected value, so increasing misfit is shown by values that are farther above or below 1.00. As an
acceptable item fit criterion, I use 0.70–1.30, a range suggested by Wright et al. (1994) for low-stakes tests. Outfit is the more sensitive measure, sometimes falling outside the suggested range for very easy or very difficult items due to one or two students who “accidentally” get the correct or incorrect answer. The more robust infit MNSQ statistic was therefore treated in this study as the more reliable indicator of misfit.

Next, a principal components analysis (PCA) of Rasch residuals is reported as a measure of unidimensionality. A Rasch PCA of item residuals examines systematic variance caused by factors other than the Rasch item difficulty measures. Such variance should ideally consist only of random noise, but a substantial amount of systematic variance caused by other factors can indicate that more than a single construct is being measured by the test. In this study, the following criteria proposed by Linacre (2010) are applied when assessing the PCA results:

1. If the percent of variance explained by the items is more than quadruple the percent of variance explained by the first contrast, this indicates good unidimensionality.

2. If the variance explained by the measures if over 50%, this indicates good unidimensionality.

3. If the variance explained by the first contrast is less than 3.0 eigenvalue units, this indicates good unidimensionality.

4. If the variance explained by the first contrast is less than 1.5 eigenvalue units, this indicates excellent unidimensionality.
5. If less than 5% of the variance is explained by the first contrast, this indicates excellent unidimensionality.

The first two of these criteria are difficult to meet, and the second criteria is nearly impossible to meet if the person separation statistic is low, as it is in this study. Therefore, the variance explained by the first contrast serves as the most important indicator of dimensionality in this study because if a second construct exists in the data, this is where it would appear.

Next, an item-person map, which presents student ability and item difficulty on a single interval scale, is used to assess how well the test items matched the students’ ability. The item difficulty estimates are also used to examine in a detailed way the factors contributing to item difficulty, and identify possible problematic items.

Finally, basic statistics of item and person reliability are reported. Item reliability refers to how well the results of the test are reproducible with a different set of students, while person reliability indicates whether the same students would likely reproduce the results on a different set of items measuring the same construct. Item and person separation statistics are also reported, in order to examine how many distinct levels of difficulty (or ability) can be identified.

Oral Reading Prosody Ratings

The oral reading prosody assessment was administered four times. Each administration consisted of six short passages, which the students read aloud into a recording device. Three passages were simplified, general topic, narrative fictional texts
taken from graded readers, and three were authentic, nursing-related, narrative non-
fiction texts. The audio recordings of the students’ readings were rated using a four-
level scale on each of three reading prosody criteria: phrasing, stress, and intonation
(see Table 5 in Chapter 3 for the descriptors). The four raters first participated in a
training session. I rated all performances myself, and the three additional raters each
rated approximately one-third of the performances, with some overlap, so that most
performances for each time and text type were rated by two raters, but a few
performances were rated by three raters. A separate FACETS analysis was conducted
for each of two text types—simplified-general and authentic-nursing—and each of four
performance times. The second, third, and fourth times were anchored to the first time
by inputting the Rasch-Andrich thresholds measurements from the first time a priori.

**Oral Reading Prosody Ratings (Time 1, Simplified)**

The Wright map for Time 1, simplified is shown in Figure 1. The students are
spread out over the entirety of the four-point scale, indicating that the raters used the
full extent of the scale. A student ability estimate of 0.00 logits corresponds with 2.5 on
the four-point scale, and 61.5% of students are at or below this level. The raters were all
between 1.00 and 2.50 logits, indicating strong agreement in terms of their application
of the scale. Intonation was slightly more challenging than phrasing and stress, but all
three criteria have difficulty estimates between -0.50 and 1.00 logit.

The rater statistics in Table 8 indicate that the mean-square fit statistics fell
within the desired range of 0.40–1.20, providing evidence of intra-rater reliability
Figure 1. Wright Map for Oral Prosody Ratings (Time 1, Simplified)

<table>
<thead>
<tr>
<th>Measure</th>
<th>+Students</th>
<th>-Rater</th>
<th>Criteria</th>
<th>Scale</th>
</tr>
</thead>
<tbody>
<tr>
<td>9</td>
<td>+ .</td>
<td>+</td>
<td></td>
<td>(4)</td>
</tr>
<tr>
<td></td>
<td>*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>+</td>
<td>+</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>+</td>
<td>+</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>+</td>
<td>+</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>+</td>
<td>+</td>
<td></td>
<td>---</td>
</tr>
<tr>
<td></td>
<td>.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>* .</td>
<td>+</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>* .</td>
<td>+</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>* .</td>
<td>+ 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>*</td>
<td></td>
<td></td>
<td>3</td>
</tr>
<tr>
<td>1</td>
<td>.</td>
<td>+ 4</td>
<td>intonation</td>
<td></td>
</tr>
<tr>
<td></td>
<td>**</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>* 0</td>
<td>* *</td>
<td>*</td>
<td></td>
<td>--- *</td>
</tr>
<tr>
<td></td>
<td>*</td>
<td></td>
<td>phrasing stress</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-1</td>
<td>.</td>
<td>+</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-2</td>
<td>.</td>
<td>+</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-3</td>
<td>* **</td>
<td>+</td>
<td></td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-4</td>
<td>.</td>
<td>+</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-5</td>
<td>.</td>
<td>+</td>
<td></td>
<td>---</td>
</tr>
<tr>
<td></td>
<td>.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-6</td>
<td>+ *******</td>
<td>+</td>
<td></td>
<td>(1)</td>
</tr>
</tbody>
</table>

Note. Each "." represents 1 student, and each "***" is two students.
Table 8. Rater Measurement Report for Oral Prosody Ratings (Time 1, Simplified)

<table>
<thead>
<tr>
<th>Rater</th>
<th>Measure</th>
<th>SE</th>
<th>Infit MNSQ</th>
<th>Infit ZSTD</th>
<th>Outfit MNSQ</th>
<th>Outfit ZSTD</th>
<th>Pt-measure correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>0.90</td>
<td>0.25</td>
<td>1.15</td>
<td>0.9</td>
<td>1.14</td>
<td>0.7</td>
<td>.86</td>
</tr>
<tr>
<td>3</td>
<td>1.53</td>
<td>0.26</td>
<td>0.89</td>
<td>-0.6</td>
<td>0.94</td>
<td>-0.2</td>
<td>.85</td>
</tr>
<tr>
<td>1</td>
<td>1.88</td>
<td>0.17</td>
<td>0.99</td>
<td>0.0</td>
<td>1.00</td>
<td>0.0</td>
<td>.90</td>
</tr>
<tr>
<td>2</td>
<td>2.17</td>
<td>0.24</td>
<td>0.94</td>
<td>-0.4</td>
<td>0.94</td>
<td>-0.3</td>
<td>.92</td>
</tr>
</tbody>
</table>

Note. MNSQ = mean square; ZSTD = Z-score standardized; Pt-measure = point-measure. All statistics are based on Rasch logits.

(consistency). For raters, the Rasch separation statistic was 2.13 (Fair), and the reliability estimate was .82 (Good), indicating that differing levels of severity among the raters can be distinguished. The measurement statistics for the three measurement criteria are reported in Table 9. The Rasch separation statistic was 3.56 (Good), and the Rasch reliability estimate was .93 (Very Good). The Rasch-Andrich thresholds indicated that the four levels of the rating scale were clearly distinguished (1: low; 2: -4.89; 3: -.03; 4: 4.92). Wolfe and Smith (2007) reported that the minimum separation for four-point rating scale thresholds is 1.10, so these thresholds met the criteria.

Table 9. Criteria Measurement Report for Oral Prosody Ratings (Time 1, simplified)

<table>
<thead>
<tr>
<th>Criterion</th>
<th>Measure</th>
<th>SE</th>
<th>Infit MNSQ</th>
<th>Infit ZSTD</th>
<th>Outfit MNSQ</th>
<th>Outfit ZSTD</th>
<th>Pt-measure correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stress</td>
<td>-0.44</td>
<td>0.19</td>
<td>0.95</td>
<td>-0.4</td>
<td>0.98</td>
<td>-0.1</td>
<td>.90</td>
</tr>
<tr>
<td>Phrasing</td>
<td>-0.37</td>
<td>0.19</td>
<td>1.09</td>
<td>0.8</td>
<td>1.15</td>
<td>1.0</td>
<td>.89</td>
</tr>
<tr>
<td>Intonation</td>
<td>0.80</td>
<td>0.19</td>
<td>0.93</td>
<td>-0.5</td>
<td>0.88</td>
<td>-0.7</td>
<td>.89</td>
</tr>
</tbody>
</table>

Note. MNSQ = mean square; ZSTD = Z-score standardized; Pt-measure = point-measure. All statistics are based on Rasch logits.
Oral Reading Prosody Ratings (Time 2, Simplified)

The Wright map for Time 2, simplified is shown in Figure 2. The students are spread out over the full four-point scale. A student ability estimate of 0.00 logits corresponds with 2.5 on the four-point scale, and 65.4% of students were at or below this level. The raters were all between 1.50 and 3.50 logits, which was a slightly wider range than at Time 1. Intonation was still the most challenging of the three criteria, and stress was more challenging than phrasing; the three criteria are between -1.50 and 1.50 logits, a slightly wider range than in Time 1.

The rater statistics are shown in Table 10. The Rasch separation statistic was 2.63 (Fair), and the reliability estimate was .87 (Good), indicating differing levels of severity among the raters. The mean-square fit statistics fall within the desired range of 0.40–1.20, providing evidence of intra-rater reliability (consistency). The criteria measurement statistics are reported in Table 11. The Rasch separation statistic was 7.11 (Excellent), and the Rasch reliability estimate was .98 (Excellent).

The Rasch-Andrich thresholds were anchored by inputting those from Time 1 (1: low; 2: -4.89; 3: -.03; 4: 4.92). Wolfe and Smith (2007) reported that the minimum separation for four-point rating scale thresholds is 1.10, so these thresholds met the criteria.
Figure 2. *Wright Map for Oral Prosody Ratings (Time 2, Simplified)*

<table>
<thead>
<tr>
<th>Measure</th>
<th>Students</th>
<th>Rater</th>
<th>Criteria</th>
<th>ANC</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>*</td>
<td></td>
<td></td>
<td>(4)</td>
</tr>
<tr>
<td>9</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>**</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>**</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>**</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>**</td>
<td>3</td>
<td></td>
<td>3</td>
</tr>
<tr>
<td>1</td>
<td>****</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>*</td>
<td>1</td>
<td>intonation</td>
<td></td>
</tr>
<tr>
<td>-1</td>
<td>***</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-2</td>
<td>**</td>
<td></td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>-3</td>
<td>****</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-4</td>
<td>**</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-5</td>
<td>**</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-6</td>
<td>*********</td>
<td></td>
<td></td>
<td>(1)</td>
</tr>
</tbody>
</table>

Note. Each "**" represents 1 student.
Table 10. *Rater Measurement Report for Oral Prosody Ratings (Time 2, Simplified)*

<table>
<thead>
<tr>
<th>Rater</th>
<th>Measure</th>
<th>SE</th>
<th>Infit MNSQ</th>
<th>Infit ZSTD</th>
<th>Outfit MNSQ</th>
<th>Outfit ZSTD</th>
<th>Pt-measure correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.66</td>
<td>0.16</td>
<td>0.68</td>
<td>-3.5</td>
<td>0.65</td>
<td>-3.0</td>
<td>.89</td>
</tr>
<tr>
<td>4</td>
<td>2.13</td>
<td>0.23</td>
<td>0.96</td>
<td>-0.2</td>
<td>0.95</td>
<td>-0.2</td>
<td>.88</td>
</tr>
<tr>
<td>3</td>
<td>2.30</td>
<td>0.27</td>
<td>0.67</td>
<td>-2.1</td>
<td>0.59</td>
<td>-2.0</td>
<td>.89</td>
</tr>
<tr>
<td>2</td>
<td>3.28</td>
<td>0.29</td>
<td>1.06</td>
<td>0.3</td>
<td>0.94</td>
<td>0.0</td>
<td>.86</td>
</tr>
</tbody>
</table>

*Note. MNSQ = mean square; ZSTD = Z-score standardized; Pt-measure = point-measure. All statistics are based on Rasch logits.*

Table 11. *Criteria Measurement Report for Oral Prosody Ratings (Time 2, Simplified)*

<table>
<thead>
<tr>
<th>Criterion</th>
<th>Measure</th>
<th>SE</th>
<th>Infit MNSQ</th>
<th>Infit ZSTD</th>
<th>Outfit MNSQ</th>
<th>Outfit ZSTD</th>
<th>Pt-measure correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phrasing</td>
<td>-1.37</td>
<td>0.18</td>
<td>0.86</td>
<td>-1.2</td>
<td>0.86</td>
<td>-0.9</td>
<td>.87</td>
</tr>
<tr>
<td>Stress</td>
<td>0.04</td>
<td>0.19</td>
<td>0.81</td>
<td>-1.7</td>
<td>0.75</td>
<td>-1.8</td>
<td>.88</td>
</tr>
<tr>
<td>Intonation</td>
<td>1.33</td>
<td>0.19</td>
<td>0.71</td>
<td>-2.6</td>
<td>0.65</td>
<td>-1.9</td>
<td>.88</td>
</tr>
</tbody>
</table>

*Note. MNSQ = mean square; ZSTD = Z-score standardized; Pt-measure = point-measure. All statistics are based on Rasch logits.*

**Oral Reading Prosody Ratings (Time 3, Simplified)**

The Wright map for Time 3, simplified is shown in Figure 3. The students spread out over the full four-point scale, with more students at the upper end of the scale than in Time 1 and Time 2. A student ability estimate of 0.00 logits corresponds with 2.5 on the four-point scale, and 60.3% of students were still at or below this level. The raters were all between 1.50 and 3.00 logits, indicating good agreement; a slightly tighter range than in Time 2 but slightly wider than in Time 1. The range and order of difficulty of the three criteria was unchanged from Time 2; the three criteria were between -1.50 and 1.50 logits.

The rater statistics are shown in Table 12. For raters, the Rasch separation statistic was 2.99 (Fair), and the reliability estimate was .90 (Good), indicating that
**Figure 3. Wright Map for Oral Prosody Ratings (Time 3, Simplified)**

<table>
<thead>
<tr>
<th>Measure</th>
<th>+Students+</th>
<th>+Rater*</th>
<th>+Criteria</th>
<th>+ANC</th>
</tr>
</thead>
<tbody>
<tr>
<td>7</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>**.</td>
<td>+</td>
<td>+</td>
<td>---</td>
</tr>
<tr>
<td>4</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>+ 4</td>
<td>+</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>* 1</td>
<td>2</td>
<td></td>
<td>3</td>
</tr>
<tr>
<td>1</td>
<td>*</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>* 0</td>
<td>* **.</td>
<td>*</td>
<td>*</td>
<td>---</td>
</tr>
<tr>
<td>-1</td>
<td>.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-2</td>
<td>**.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-3</td>
<td>**</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-4</td>
<td>**.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-5</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-6</td>
<td>* *****</td>
<td></td>
<td></td>
<td>(1)</td>
</tr>
</tbody>
</table>

Note. Each "**" represents 2 students, and each "." is 1 student.

Differing levels of severity among the raters can be distinguished. The mean-square fit statistics generally fell within the desired range of .40-1.20, providing evidence of intra-
rater reliability (consistency). The infit MNSQ for Rater 2 was outside the range, indicating some minor inconsistency.

Table 12. *Rater Measurement Report for Oral Prosody Ratings (Time 3, Simplified)*

<table>
<thead>
<tr>
<th>Rater</th>
<th>Measure</th>
<th>SE</th>
<th>Infit MNSQ</th>
<th>Infit ZSTD</th>
<th>Outfit MNSQ</th>
<th>Outfit ZSTD</th>
<th>Pt-measure correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>1.19</td>
<td>0.27</td>
<td>0.73</td>
<td>-1.6</td>
<td>0.69</td>
<td>-1.3</td>
<td>.92</td>
</tr>
<tr>
<td>1</td>
<td>2.23</td>
<td>0.16</td>
<td>0.61</td>
<td>-4.4</td>
<td>0.50</td>
<td>-3.6</td>
<td>.91</td>
</tr>
<tr>
<td>2</td>
<td>2.66</td>
<td>0.26</td>
<td>1.28</td>
<td>1.5</td>
<td>1.09</td>
<td>0.3</td>
<td>.84</td>
</tr>
<tr>
<td>4</td>
<td>2.90</td>
<td>0.24</td>
<td>1.00</td>
<td>0.0</td>
<td>0.89</td>
<td>-0.2</td>
<td>.84</td>
</tr>
</tbody>
</table>

*Note.* MNSQ = mean square; ZSTD = Z-score standardized; Pt-measure = point-measure. All statistics are based on Rasch logits.

The criteria measurement statistics are reported in Table 13. The Rasch separation statistic was 8.30 (Excellent), and the Rasch reliability estimate was .99 (Excellent).

The Rasch-Andrich thresholds were anchored by inputting those from Time 1 (1: low; 2: -4.89; 3: -.03; 4: 4.92). The minimum separation for four-point rating scale thresholds is 1.10, so these thresholds met the criteria (Wolfe & Smith, 2007).

Table 13. *Criteria Measurement Report for Oral Prosody Ratings (Time 3, Simplified)*

<table>
<thead>
<tr>
<th>Criterion</th>
<th>Measure</th>
<th>SE</th>
<th>Infit MNSQ</th>
<th>Infit ZSTD</th>
<th>Outfit MNSQ</th>
<th>Outfit ZSTD</th>
<th>Pt-measure correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phrasing</td>
<td>-1.52</td>
<td>0.18</td>
<td>0.78</td>
<td>-2.1</td>
<td>0.72</td>
<td>-2.3</td>
<td>.89</td>
</tr>
<tr>
<td>Stress</td>
<td>-0.18</td>
<td>0.19</td>
<td>0.93</td>
<td>-0.5</td>
<td>0.88</td>
<td>-0.7</td>
<td>.89</td>
</tr>
<tr>
<td>Intonation</td>
<td>1.70</td>
<td>0.21</td>
<td>0.76</td>
<td>-1.9</td>
<td>0.57</td>
<td>-1.5</td>
<td>.88</td>
</tr>
</tbody>
</table>

*Note.* MNSQ = mean square; ZSTD = Z-score standardized; Pt-measure = point-measure. All statistics are based on Rasch logits.
Oral Reading Prosody Ratings (Time 4, Simplified)

The Wright map for Time 4, simplified is shown in Figure 4. The students were spread out over the full four-point scale. A student ability estimate of 0.00 logits corresponds to 2.5 on the four-point scale, and 64.1% of students were still at or below this level. The raters were all between 1.00 and 2.50 logits, a tight range indicating strong agreement. The range and order of difficulty of the three criteria were little changed from Times 1 and 2; the three criteria were between -1.50 and 2.00 logits. The rater statistics are shown in Table 14. For raters, the Rasch separation statistic was 1.58 (Poor), and the reliability estimate was .71 (Fair), indicating that differing levels of severity among the raters could be distinguished rather poorly. The mean-square fit statistics fell within the desired range of 0.40–1.20, providing evidence of intra-rater reliability (consistency). The criteria measurement statistics are reported in Table 15. The Rasch separation statistic was 8.04 (Excellent), and the Rasch reliability estimate was .98 (Excellent). The Rasch-Andrich thresholds were anchored by inputting those from Time 1 (1: low; 2: -4.89; 3: -.03; 4: 4.92). Wolfe and Smith (2007) reported that the minimum separation for four-point rating scale thresholds is 1.10, so these thresholds met the criteria.

Oral Reading Prosody Ratings (Time 1, Authentic)

The Wright map for Time 1, authentic is shown in Figure 5. The students spread out over the entirety of the four-point scale, indicating that the raters used the full extent of the scale. A student ability estimate of 0.00 logits corresponds with 2.5 on the four-
Figure 4. *Wright Map for Oral Prosody Ratings (Time 4, Simplified)*

<table>
<thead>
<tr>
<th>Measure</th>
<th>Students</th>
<th>-Rater-</th>
<th>-Criteria</th>
<th>ANC</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>**</td>
<td>+</td>
<td>+</td>
<td>(4)</td>
</tr>
<tr>
<td>7</td>
<td>*</td>
<td>+</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>**</td>
<td>+</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>---</td>
</tr>
<tr>
<td>4</td>
<td>****</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>**</td>
<td></td>
<td>2 4</td>
<td>3</td>
</tr>
<tr>
<td>2</td>
<td>**</td>
<td>+</td>
<td>3</td>
<td>+</td>
</tr>
<tr>
<td>1</td>
<td>**</td>
<td>+</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>****</td>
<td>*</td>
<td>*</td>
<td>---</td>
</tr>
<tr>
<td>-1</td>
<td>**</td>
<td>+</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>-2</td>
<td>****</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-3</td>
<td>**</td>
<td></td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>-4</td>
<td>**</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-5</td>
<td>**</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-6</td>
<td>**</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. Each "**" represents 1 student.
Table 14. **Rater Measurement Report for Oral Prosody Ratings (Time 4, Simplified)**

<table>
<thead>
<tr>
<th>Rater</th>
<th>Measure</th>
<th>SE</th>
<th>Infit MNSQ</th>
<th>Infit ZSTD</th>
<th>Outfit MNSQ</th>
<th>Outfit ZSTD</th>
<th>Pt-measure correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.31</td>
<td>0.16</td>
<td>0.76</td>
<td>-2.7</td>
<td>0.66</td>
<td>-3.1</td>
<td>.90</td>
</tr>
<tr>
<td>3</td>
<td>1.90</td>
<td>0.27</td>
<td>0.56</td>
<td>-3.1</td>
<td>0.50</td>
<td>-3.1</td>
<td>.92</td>
</tr>
<tr>
<td>2</td>
<td>2.21</td>
<td>0.21</td>
<td>1.00</td>
<td>0.0</td>
<td>0.94</td>
<td>-0.2</td>
<td>.87</td>
</tr>
<tr>
<td>4</td>
<td>2.21</td>
<td>0.26</td>
<td>1.03</td>
<td>0.2</td>
<td>1.05</td>
<td>0.2</td>
<td>.90</td>
</tr>
</tbody>
</table>

*Note. MNSQ = mean square; ZSTD = Z-score standardized; Pt-measure = point-measure. All statistics are based on Rasch logits.*

Table 15. **Criteria Measurement Report for Oral Prosody Ratings (Time 4, Simplified)**

<table>
<thead>
<tr>
<th>Criteria</th>
<th>Measure</th>
<th>SE</th>
<th>Infit MNSQ</th>
<th>Infit ZSTD</th>
<th>Outfit MNSQ</th>
<th>Outfit ZSTD</th>
<th>Pt-measure correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phrasing</td>
<td>-1.31</td>
<td>0.18</td>
<td>0.85</td>
<td>-1.3</td>
<td>0.77</td>
<td>-1.8</td>
<td>.90</td>
</tr>
<tr>
<td>Stress</td>
<td>-0.28</td>
<td>0.18</td>
<td>0.77</td>
<td>-2.3</td>
<td>0.73</td>
<td>-2.3</td>
<td>.90</td>
</tr>
<tr>
<td>Intonation</td>
<td>1.58</td>
<td>0.19</td>
<td>0.87</td>
<td>-1.1</td>
<td>0.81</td>
<td>-1.0</td>
<td>.88</td>
</tr>
</tbody>
</table>

*Note. MNSQ = mean square; ZSTD = Z-score standardized; Pt-measure = point-measure. All statistics are based on Rasch logits.*

point scale, and 59.0% of students were at or below this level. The raters were in a tight range compared to the simplified texts—all between 0.00 and 1.00 logits, indicating strong agreement in terms of their application of the scale. The range and order of difficulty of the three criteria was relatively unchanged, with intonation being the most challenging, stress in the middle, and phrasing easiest; all three criteria were between -1.00 and 1.00 logit.

The rater statistics are shown in Table 16. For raters, the Rasch separation statistic was 1.37 (Poor), and the reliability estimate was .65 (Poor), indicating that differing levels of severity among the raters can be distinguished rather poorly. The mean-square fit statistics fell for the most part within the desired range of .40-1.20,
providing evidence of intra-rater reliability (consistency). The outfit MNSQ for Rater 2 was outside the range, indicating evidence of some minor inconsistencies. The criteria
measurement statistics are reported in Table 17. The Rasch separation statistic was 5.37 (Excellent), and the Rasch reliability estimate was .97 (Excellent).

Table 16. Rater Measurement Report for Oral Prosody Ratings (Time 1, Authentic)

<table>
<thead>
<tr>
<th>Rater</th>
<th>Measure</th>
<th>SE</th>
<th>Infit MNSQ</th>
<th>Infit ZSTD</th>
<th>Outfit MNSQ</th>
<th>Outfit ZSTD</th>
<th>Pt-measure correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.07</td>
<td>0.26</td>
<td>1.06</td>
<td>0.4</td>
<td>1.25</td>
<td>1.1</td>
<td>.93</td>
</tr>
<tr>
<td>3</td>
<td>0.47</td>
<td>0.35</td>
<td>1.12</td>
<td>0.6</td>
<td>1.14</td>
<td>0.5</td>
<td>.91</td>
</tr>
<tr>
<td>4</td>
<td>0.61</td>
<td>0.27</td>
<td>0.75</td>
<td>-1.8</td>
<td>0.75</td>
<td>-1.1</td>
<td>.91</td>
</tr>
<tr>
<td>1</td>
<td>1.19</td>
<td>0.18</td>
<td>0.95</td>
<td>-0.4</td>
<td>0.91</td>
<td>-0.5</td>
<td>.91</td>
</tr>
</tbody>
</table>

Note. MNSQ = mean square; ZSTD = Z-score standardized; Pt-measure = point-measure. All statistics are based on Rasch logits.

Table 17. Criteria Measurement Report for Oral Prosody Ratings (Time 1, Authentic)

<table>
<thead>
<tr>
<th>Criteria</th>
<th>Measure</th>
<th>SE</th>
<th>Infit MNSQ</th>
<th>Infit ZSTD</th>
<th>Outfit MNSQ</th>
<th>Outfit ZSTD</th>
<th>Pt-measure correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phrasing</td>
<td>-1.23</td>
<td>0.21</td>
<td>1.12</td>
<td>1.0</td>
<td>1.30</td>
<td>1.5</td>
<td>.91</td>
</tr>
<tr>
<td>Stress</td>
<td>0.21</td>
<td>0.20</td>
<td>0.85</td>
<td>-1.4</td>
<td>0.86</td>
<td>-0.8</td>
<td>.93</td>
</tr>
<tr>
<td>Intonation</td>
<td>1.02</td>
<td>0.21</td>
<td>0.90</td>
<td>-0.8</td>
<td>0.79</td>
<td>-1.1</td>
<td>.92</td>
</tr>
</tbody>
</table>

Note. MNSQ = mean square; ZSTD = Z-score standardized; Pt-measure = point-measure. All statistics are based on Rasch logits.

The Rasch-Andrich thresholds indicated that the four levels of the rating scale were well-distinguished from each other (1: low; 2: -6.49; 3: .28; 4: 6.21). Wolfe and Smith (2007) reported that the minimum separation for four-point rating scale thresholds is 1.10, so these thresholds met the criteria.
Oral Prosody Ratings (Time 2, Authentic)

The Wright map for Time 2, authentic is shown in Figure 6. The students spread out over the entirety of the four-point scale. A student ability estimate of 0.00 logits corresponds with 2.5 on the four-point scale, and 52.3% of students were at or below this level, indicating that overall, the students performed somewhat better than on the simplified texts and on Time 1 of the authentic text. This result was also seen in the higher ratings given by the raters, which were between 2.00 and 3.00 logits. This is also a relatively tight range, indicating strong rater agreement. The order of difficulty of the three criteria was relatively unchanged, but the range was somewhat wider, with the three criteria between -2.00 and 2.00 logits.

The rater statistics are shown in Table 18. For raters, the Rasch separation statistic was 1.93 (Poor), and the reliability estimate was .79 (Fair), indicating that differing levels of severity among the raters can be distinguished, but rather poorly. The mean-square fit statistics fell for the most part within the desired range of 0.40–1.20, providing evidence of intra-rater reliability (consistency). However, the infit and outfit MNSQ for Rater 4 was outside the range, indicating some inconsistency. The criteria measurement statistics are reported in Table 19. The Rasch separation statistic was 8.31 (Excellent), and the Rasch reliability estimate was .99 (Excellent). The Rasch-Andrich thresholds were anchored by inputting those from Time 1 (1: low; 2: -6.49; 3: .28; 4: 6.21). Wolfe and Smith (2007) reported that the minimum separation for four-point rating scale thresholds is 1.10, so these thresholds met the criteria.
Oral Prosody Ratings (Time 3, Authentic)

The Wright map for Time 3, authentic is shown in Figure 7. The students spread out over the four-point scale. A student ability estimate of 0.00 logits corresponds
Table 18. Rater Measurement Report for Oral Prosody Ratings (Time 2, Authentic)

<table>
<thead>
<tr>
<th>Rater</th>
<th>Measure</th>
<th>SE</th>
<th>Infit MNSQ</th>
<th>Infit ZSTD</th>
<th>Outfit MNSQ</th>
<th>Outfit ZSTD</th>
<th>Pt-measure correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>2.14</td>
<td>0.25</td>
<td>1.16</td>
<td>1.1</td>
<td>1.13</td>
<td>0.6</td>
<td>.92</td>
</tr>
<tr>
<td>1</td>
<td>2.17</td>
<td>0.17</td>
<td>0.84</td>
<td>-1.7</td>
<td>0.68</td>
<td>-2.1</td>
<td>.91</td>
</tr>
<tr>
<td>3</td>
<td>3.09</td>
<td>0.33</td>
<td>0.83</td>
<td>-0.9</td>
<td>0.67</td>
<td>-0.9</td>
<td>.89</td>
</tr>
<tr>
<td>4</td>
<td>3.21</td>
<td>0.29</td>
<td>1.29</td>
<td>1.7</td>
<td>1.25</td>
<td>0.7</td>
<td>.86</td>
</tr>
</tbody>
</table>

Note. MNSQ = mean square; ZSTD = Z-score standardized; Pt-measure = point-measure. All statistics are based on Rasch logits.

Table 19. Criteria Measurement Report for Oral Prosody Ratings (Time 2, Authentic)

<table>
<thead>
<tr>
<th>Criterion</th>
<th>Measure</th>
<th>SE</th>
<th>Infit MNSQ</th>
<th>Infit ZSTD</th>
<th>Outfit MNSQ</th>
<th>Outfit ZSTD</th>
<th>Pt-measure correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phrasing</td>
<td>-1.66</td>
<td>0.20</td>
<td>1.13</td>
<td>1.2</td>
<td>1.03</td>
<td>0.2</td>
<td>.89</td>
</tr>
<tr>
<td>Stress</td>
<td>-0.15</td>
<td>0.19</td>
<td>0.93</td>
<td>-0.6</td>
<td>0.80</td>
<td>-1.4</td>
<td>.89</td>
</tr>
<tr>
<td>Intonation</td>
<td>1.81</td>
<td>0.23</td>
<td>0.85</td>
<td>-1.0</td>
<td>0.81</td>
<td>-0.6</td>
<td>.92</td>
</tr>
</tbody>
</table>

Note. MNSQ = mean square; ZSTD = Z-score standardized; Pt-measure = point-measure. All statistics are based on Rasch logits.

with 2.5 on the four-point scale, and 56.4% of students were at or below this level. The four raters fell within a relatively wide range between 0.50 and 2.50 logits. The order of difficulty of the three criteria was unchanged, and the range was relatively wide, between -2.00 and 2.00 logits.

The rater statistics are shown in Table 20. For raters, the Rasch separation statistic was 2.39 (Fair), and the reliability estimate was .85 (Good), indicating that differing levels of severity among the raters could be distinguished. Raters 1 and 3 had mean-square fit statistics that fell within the desired range of 0.40–1.20, providing evidence of intra-rater reliability (consistency). However, the infit and outfit MNSQ of Raters 2 and 4 were outside the range, indicating some inconsistency. The criteria
measurement statistics are reported in Table 21. The Rasch separation statistic was 9.42 (Excellent), and the Rasch reliability estimate was .99 (Excellent).
Table 20. Rater Measurement Report for the Oral Prosody Ratings (Time 3, Authentic)

<table>
<thead>
<tr>
<th>Rater</th>
<th>Measure</th>
<th>SE</th>
<th>Infit MNSQ</th>
<th>Infit ZSTD</th>
<th>Outfit MNSQ</th>
<th>Outfit ZSTD</th>
<th>Pt-measure correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>0.61</td>
<td>0.29</td>
<td>0.95</td>
<td>-0.2</td>
<td>0.87</td>
<td>-0.3</td>
<td>.89</td>
</tr>
<tr>
<td>2</td>
<td>1.21</td>
<td>0.32</td>
<td>1.68</td>
<td>3.0</td>
<td>1.76</td>
<td>1.6</td>
<td>.93</td>
</tr>
<tr>
<td>1</td>
<td>1.60</td>
<td>0.18</td>
<td>0.81</td>
<td>-1.8</td>
<td>0.72</td>
<td>-1.3</td>
<td>.93</td>
</tr>
<tr>
<td>4</td>
<td>2.30</td>
<td>0.28</td>
<td>1.30</td>
<td>1.6</td>
<td>1.34</td>
<td>1.0</td>
<td>.87</td>
</tr>
</tbody>
</table>

Note. MNSQ = mean square; ZSTD = Z-score standardized; Pt-measure = point-measure. All statistics are based on Rasch logits.


<table>
<thead>
<tr>
<th>Criteria</th>
<th>Measure</th>
<th>SE</th>
<th>Infit MNSQ</th>
<th>Infit ZSTD</th>
<th>Outfit MNSQ</th>
<th>Outfit ZSTD</th>
<th>Pt-measure correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phrasing</td>
<td>-1.97</td>
<td>0.22</td>
<td>1.37</td>
<td>2.6</td>
<td>1.49</td>
<td>1.9</td>
<td>.89</td>
</tr>
<tr>
<td>Stress</td>
<td>-0.15</td>
<td>0.22</td>
<td>0.82</td>
<td>-1.3</td>
<td>0.74</td>
<td>-1.3</td>
<td>.94</td>
</tr>
<tr>
<td>Intonation</td>
<td>2.12</td>
<td>0.21</td>
<td>0.98</td>
<td>-0.1</td>
<td>0.86</td>
<td>-0.3</td>
<td>.90</td>
</tr>
</tbody>
</table>

Note. MNSQ = mean square; ZSTD = Z-score standardized; Pt-measure = point-measure. All statistics are based on Rasch logits.

The Rasch-Andrich thresholds were anchored by inputting those from Time 1 (1: low; 2: -6.49; 3: .28; 4: 6.21). These thresholds met the 1.10 criterion for the minimum separation for four-point rating scale thresholds (Wolfe & Smith, 2007).

Oral Prosody Ratings (Time 4, Authentic)

The Wright map for Time 4, authentic is shown in Figure 8. The students spread out over the entirety of the four-point scale. A student ability estimate of 0.00 logits corresponds with 2.5 on the four-point scale, and 52.3% of students were at or below this level, indicating somewhat higher performance compared with Time 3. The four raters fell between 0.00 and 1.50 logits. The order of difficulty of the three criteria is unchanged, and the range is again wide, between -2.00 and 2.00 logits.
The rater statistics are shown in Table 22. For raters, the Rasch separation statistic was 2.46 (Fair), and the reliability estimate was .86 (Good), indicating that
differing levels of severity among the raters can be distinguished. The raters’ mean-square fit statistics fell mostly within the desired range of 0.40–1.20, providing evidence of intra-rater reliability (consistency). Rater 2 showed minor inconsistency. The criteria measurement statistics are reported in Table 23. The Rasch separation statistic was 10.07 (Excellent), and the Rasch reliability estimate was .99 (Excellent). The Rasch-Andrich thresholds were anchored by inputting those from Time 1 (1: low; 2: -6.49; 3: .28; 4: 6.21). Wolfe and Smith (2007) reported that the minimum separation for four-point rating scale thresholds is 1.10, so these thresholds met the criteria.

Table 22. Rater Measurement Report for the Oral Prosody Ratings (Time 4, Authentic)

<table>
<thead>
<tr>
<th>Rater</th>
<th>Measure</th>
<th>SE</th>
<th>Infit</th>
<th>MNSQ</th>
<th>Infit</th>
<th>ZSTD</th>
<th>Outfit</th>
<th>MNSQ</th>
<th>Outfit</th>
<th>ZSTD</th>
<th>Pt-measure correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>-0.17</td>
<td>0.31</td>
<td>1.12</td>
<td>0.7</td>
<td>1.25</td>
<td>0.7</td>
<td>.92</td>
<td></td>
<td></td>
<td></td>
<td>.92</td>
</tr>
<tr>
<td>1</td>
<td>0.37</td>
<td>0.17</td>
<td>0.77</td>
<td>-2.4</td>
<td>0.66</td>
<td>-1.8</td>
<td>.92</td>
<td></td>
<td></td>
<td></td>
<td>.92</td>
</tr>
<tr>
<td>4</td>
<td>1.00</td>
<td>0.29</td>
<td>0.86</td>
<td>-0.8</td>
<td>0.94</td>
<td>0.0</td>
<td>.91</td>
<td></td>
<td></td>
<td></td>
<td>.90</td>
</tr>
<tr>
<td>3</td>
<td>1.43</td>
<td>0.26</td>
<td>0.78</td>
<td>-1.4</td>
<td>0.62</td>
<td>-1.3</td>
<td>.90</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. MNSQ = mean square; ZSTD = Z-score standardized; Pt-measure = point-measure. All statistics are based on Rasch logits.

Table 23. Criteria Measurement Report for the Oral Prosody Ratings (Time 4, Authentic)

<table>
<thead>
<tr>
<th>Criteria</th>
<th>Measure</th>
<th>SE</th>
<th>Infit</th>
<th>MNSQ</th>
<th>Infit</th>
<th>ZSTD</th>
<th>Outfit</th>
<th>MNSQ</th>
<th>Outfit</th>
<th>ZSTD</th>
<th>Pt-measure correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phrasing</td>
<td>-2.16</td>
<td>0.20</td>
<td>0.87</td>
<td>-1.0</td>
<td>0.83</td>
<td>-0.6</td>
<td>.91</td>
<td></td>
<td></td>
<td></td>
<td>.91</td>
</tr>
<tr>
<td>Stress</td>
<td>0.10</td>
<td>0.20</td>
<td>0.89</td>
<td>-0.9</td>
<td>0.81</td>
<td>-1.0</td>
<td>.91</td>
<td></td>
<td></td>
<td></td>
<td>.91</td>
</tr>
<tr>
<td>Intonation</td>
<td>2.06</td>
<td>0.22</td>
<td>0.74</td>
<td>-2.0</td>
<td>0.73</td>
<td>-0.8</td>
<td>.92</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. MNSQ = mean square; ZSTD = Z-score standardized; Pt-measure = point-measure. All statistics are based on Rasch logits.
Reading Comprehension Tests

The reading comprehension tests were administered four times. Each test consisted of six passages with four questions each, for a total of 24 questions. Three of the passages were simplified, general topic, narrative fiction tests from graded readers, and three were authentic, nursing-related topic, narrative non-fiction texts. Within the simplified texts, there were three difficulty levels: easy (Flesch-Kincaid grade level 4.6-4.9), medium (F-K 5.3), and difficult (F-K 5.7-5.8). Within the authentic texts, text difficulty was categorized as follows: easy (F-K 9.1-9.3), medium (F-K 9.5-10.0), and difficult (F-K 11.3-11.7). Regarding item type, the first two questions for each text were repetitive questions, meaning that the answer could be found with slightly different wording at a single location in the text. The third question for each text was a synthesis question requiring the readers to synthesize information found in two locations in the text. The fourth question for each text was an inferencing question requiring the readers to synthesize information from one or more locations in the text with their own existing schemata of knowledge and experience.

In order to avoid the learning effect that would likely occur with four administrations of the same test, while also allowing for the analysis of validity and reliability across test forms, on the second, third, and fourth administrations I used two texts—one simplified and one authentic—from previous administrations, allowing for anchoring of the accompanying items (8 of the 24 items) to previous administrations. The determination of which items (and therefore texts) to use as anchors was made after analyzing the item difficulty estimates and fit statistics of the previous administration.
Anchored items were carefully selected based on their difficulty estimates in order to ensure a good range of difficulty among the items while avoiding using misfitting items as anchors and avoiding using items that were too easy or too difficult. The resulting anchoring plan is presented in Table 24.

Table 24. Anchoring Plan for the Reading Comprehension Tests

<table>
<thead>
<tr>
<th>Time 1</th>
<th>Time 2</th>
<th>Time 3</th>
<th>Time 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rabbit-Proof Fence</td>
<td>Mysterious Death of Charles Bravo (new)</td>
<td>Kidnapped (new)</td>
<td>Crown of Violet (Good seats) (new)</td>
</tr>
<tr>
<td><strong>Simplified</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Crown of Violet (Going to the theater) (anchored)</td>
<td>Crown of Violet (Going to the theater) (anchored)</td>
<td>Call of the Wild (new)</td>
<td>Call of the Wild (anchored)</td>
</tr>
<tr>
<td>Justice</td>
<td>Last Sherlock Holmes Story (new)</td>
<td>Last Sherlock Holmes Story (anchored)</td>
<td>Tales of Mystery and Imagination (new)</td>
</tr>
<tr>
<td>HT: Washing feet</td>
<td>HT: Washing feet (anchored)</td>
<td>HT: Comforting grieving patients (new)</td>
<td>HT: Writing about patients (new)</td>
</tr>
<tr>
<td><strong>Authentic</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HT: Women’s health</td>
<td>HT: Patients’ stories (new)</td>
<td>HT: Patients’ stories (anchored)</td>
<td>HT: When a nurse becomes ill (new)</td>
</tr>
<tr>
<td>HT: Susan’s pain</td>
<td>HT: A troubled childhood (new)</td>
<td>HT: The line between wellness and sickness (new)</td>
<td>HT: The line between wellness and sickness (anchored)</td>
</tr>
</tbody>
</table>

*Note. Anchored texts are in bold.*

**Reading Comprehension Test (Time 1)**

The Rasch descriptive statistics for the 24 items administered at Time 1 are reported in Table 25. These statistics indicate a wide range of item difficulty, nearly 5
logits (-2.11 to 2.71). Using 0.70–1.30 as the acceptable range of item fit for low-stakes tests (Wright et al., 1994), all of the items demonstrated acceptable fit to the Rasch model. The infit MNSQ statistic ranged from 0.82–1.19, and the point-measure correlations ranged from .14 to .57.

Table 25. Rasch Descriptive Statistics for the Reading Comprehension Test Items (Time 1)

<table>
<thead>
<tr>
<th>Item</th>
<th>Measure</th>
<th>SE</th>
<th>Infit MNSQ</th>
<th>Infit ZSTD</th>
<th>Outfit MNSQ</th>
<th>Outfit ZSTD</th>
<th>Pt-measure correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>23</td>
<td>2.71</td>
<td>0.34</td>
<td>1.03</td>
<td>0.2</td>
<td>0.96</td>
<td>0.0</td>
<td>.25</td>
</tr>
<tr>
<td>8</td>
<td>1.53</td>
<td>0.26</td>
<td>0.99</td>
<td>0.0</td>
<td>0.98</td>
<td>-0.1</td>
<td>.36</td>
</tr>
<tr>
<td>14</td>
<td>1.46</td>
<td>0.26</td>
<td>1.04</td>
<td>0.4</td>
<td>1.06</td>
<td>0.4</td>
<td>.31</td>
</tr>
<tr>
<td>24</td>
<td>1.39</td>
<td>0.26</td>
<td>1.19</td>
<td>1.8</td>
<td>1.24</td>
<td>1.6</td>
<td>.14</td>
</tr>
<tr>
<td>16</td>
<td>1.02</td>
<td>0.25</td>
<td>0.82</td>
<td>-2.1</td>
<td>0.80</td>
<td>-1.8</td>
<td>.57</td>
</tr>
<tr>
<td>21</td>
<td>0.89</td>
<td>0.25</td>
<td>1.13</td>
<td>1.5</td>
<td>1.17</td>
<td>1.6</td>
<td>.22</td>
</tr>
<tr>
<td>13</td>
<td>0.77</td>
<td>0.24</td>
<td>0.92</td>
<td>-0.9</td>
<td>0.94</td>
<td>-0.6</td>
<td>.45</td>
</tr>
<tr>
<td>20</td>
<td>0.77</td>
<td>0.24</td>
<td>1.16</td>
<td>1.8</td>
<td>1.19</td>
<td>1.8</td>
<td>.19</td>
</tr>
<tr>
<td>17</td>
<td>0.66</td>
<td>0.24</td>
<td>0.97</td>
<td>-0.3</td>
<td>1.00</td>
<td>0.0</td>
<td>.40</td>
</tr>
<tr>
<td>1</td>
<td>0.37</td>
<td>0.25</td>
<td>1.06</td>
<td>0.7</td>
<td>1.12</td>
<td>1.1</td>
<td>.29</td>
</tr>
<tr>
<td>18</td>
<td>0.05</td>
<td>0.25</td>
<td>0.97</td>
<td>-0.3</td>
<td>1.02</td>
<td>0.2</td>
<td>.38</td>
</tr>
<tr>
<td>19</td>
<td>0.05</td>
<td>0.25</td>
<td>1.06</td>
<td>0.7</td>
<td>1.10</td>
<td>0.8</td>
<td>.28</td>
</tr>
<tr>
<td>22</td>
<td>-0.08</td>
<td>0.25</td>
<td>0.86</td>
<td>-1.5</td>
<td>0.79</td>
<td>-1.5</td>
<td>.53</td>
</tr>
<tr>
<td>9</td>
<td>-0.32</td>
<td>0.27</td>
<td>1.13</td>
<td>1.1</td>
<td>1.16</td>
<td>0.9</td>
<td>.19</td>
</tr>
<tr>
<td>12</td>
<td>-0.42</td>
<td>0.27</td>
<td>0.90</td>
<td>-0.8</td>
<td>0.84</td>
<td>-0.8</td>
<td>.45</td>
</tr>
<tr>
<td>7</td>
<td>-0.49</td>
<td>0.27</td>
<td>0.96</td>
<td>-0.3</td>
<td>0.84</td>
<td>-0.8</td>
<td>.41</td>
</tr>
<tr>
<td>4</td>
<td>-0.57</td>
<td>0.27</td>
<td>0.92</td>
<td>-0.6</td>
<td>0.86</td>
<td>-0.7</td>
<td>.42</td>
</tr>
<tr>
<td>6</td>
<td>-0.81</td>
<td>0.29</td>
<td>0.97</td>
<td>-0.2</td>
<td>1.03</td>
<td>0.2</td>
<td>.33</td>
</tr>
<tr>
<td>3</td>
<td>-1.07</td>
<td>0.31</td>
<td>0.96</td>
<td>-0.2</td>
<td>0.96</td>
<td>0.0</td>
<td>.32</td>
</tr>
<tr>
<td>15</td>
<td>-1.28</td>
<td>0.33</td>
<td>1.01</td>
<td>0.1</td>
<td>0.93</td>
<td>-0.1</td>
<td>.27</td>
</tr>
<tr>
<td>10</td>
<td>-1.39</td>
<td>0.34</td>
<td>0.91</td>
<td>-0.3</td>
<td>0.88</td>
<td>-0.3</td>
<td>.36</td>
</tr>
<tr>
<td>5</td>
<td>-1.50</td>
<td>0.35</td>
<td>0.93</td>
<td>-0.2</td>
<td>0.79</td>
<td>-0.5</td>
<td>.34</td>
</tr>
<tr>
<td>11</td>
<td>-1.63</td>
<td>0.37</td>
<td>1.02</td>
<td>0.2</td>
<td>1.09</td>
<td>0.4</td>
<td>.20</td>
</tr>
<tr>
<td>2</td>
<td>-2.11</td>
<td>0.43</td>
<td>1.01</td>
<td>0.1</td>
<td>1.02</td>
<td>0.2</td>
<td>.18</td>
</tr>
</tbody>
</table>

*Note.* MNSQ = mean square; ZSTD = Z-score standardized; Pt-measure = point-measure. All statistics are based on Rasch logits.

The dimensionality of the reading comprehension items was investigated using the Linacre (2010) criteria described above. The variance explained by the items was
18.7%, while the variance explained by measures was 28.3%. The unexplained variance explained by the first contrast was 2.2 eigenvalue units (6.6%), indicating a good likelihood of unidimensionality.

The item-person map (Wright map) for Time 1 (Figure 9) shows how well the difficulty of the items matches the students’ ability. The students’ ability estimates are displayed on the left side, with each student shown represented as an X, and the items are on the right side. Students with higher ability are shown at the top, and those with lower ability are at the bottom. Similarly, more difficult items are at the top and easier items are at the bottom. Overall, there is some mismatch between item difficulty and student ability, with 13 of the students (16.5%) higher than 23 of the 24 items, and 6 of the items (25%) easier than all of the students’ ability levels. The mean of the person ability estimates was 0.66 logits ($SD = 0.84$), and the mean of the item difficulty estimates was 0.00 logits ($SD = 1.17$).

Table 26 displays the effect of text type, text difficulty, and item type on item difficulty estimates. With few exceptions, it is clear that the authentic text items were more difficult than the simplified text items. Within the authentic texts, the items for the most difficult text and the medium-level text were more difficult than those for the easier text. However, within the simplified texts, text difficulty had no apparent effect on item difficulty.

Item type also affected item difficulty. Inferencing questions were the most difficult to answer overall; however, there was no clear difference between synthesis and repetitive questions in terms of their item difficulty estimates. This result prompted
Figure 9. Abridged Item-Person Map for the Reading Comprehension Test (Time 1)

<table>
<thead>
<tr>
<th>More able persons</th>
<th>More difficult items</th>
</tr>
</thead>
<tbody>
<tr>
<td>XXX</td>
<td>T</td>
</tr>
<tr>
<td>T</td>
<td>T</td>
</tr>
<tr>
<td>2</td>
<td>XXX</td>
</tr>
<tr>
<td>XXXXXX</td>
<td>S</td>
</tr>
<tr>
<td>XXX</td>
<td>A3-4</td>
</tr>
<tr>
<td>X</td>
<td>S</td>
</tr>
<tr>
<td>XXXXXXX</td>
<td>1</td>
</tr>
<tr>
<td>XXXXXXXX</td>
<td>A3-1</td>
</tr>
<tr>
<td>M</td>
<td>A2-1</td>
</tr>
<tr>
<td>XXXX</td>
<td></td>
</tr>
<tr>
<td>XXXXX</td>
<td>S1-1</td>
</tr>
<tr>
<td>XXXXXX</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>XXXXXXXX</td>
</tr>
<tr>
<td>A3-2</td>
<td></td>
</tr>
<tr>
<td>XXXX</td>
<td>S</td>
</tr>
<tr>
<td>S</td>
<td>A1-3</td>
</tr>
<tr>
<td>X</td>
<td>S3-2</td>
</tr>
<tr>
<td>XXXXX</td>
<td>S3-4</td>
</tr>
<tr>
<td>XXX</td>
<td>S2-2</td>
</tr>
<tr>
<td>XXX</td>
<td>S1-2</td>
</tr>
<tr>
<td>-1</td>
<td>T+</td>
</tr>
<tr>
<td></td>
<td>S</td>
</tr>
<tr>
<td></td>
<td>S3-2</td>
</tr>
<tr>
<td></td>
<td>S2-1</td>
</tr>
<tr>
<td></td>
<td>S3-3</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>-2</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td>T</td>
</tr>
</tbody>
</table>

Less able persons | Less difficult items
-------------------|----------------------

Note. Each “X” equals 1 student. M = Mean; S = One standard deviation from the mean; T = Two standard deviations from the mean; A = authentic text; S = simplified text.
Table 26. Item Difficulty by Text Type, Text Difficulty, and Item Type for the Reading Comprehension Test (Time 1)

<table>
<thead>
<tr>
<th>Item</th>
<th>Measure</th>
<th>Simp-easy (S3)</th>
<th>Simp-mid (S2)</th>
<th>Simp-difficult (S1)</th>
<th>Auth-easy (A2)</th>
<th>Auth-mid (A3)</th>
<th>Auth-difficult (A1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>23</td>
<td>(A3-3)</td>
<td>2.71</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>synth</td>
</tr>
<tr>
<td>8</td>
<td>(S2-4)</td>
<td>1.53</td>
<td>-</td>
<td>infer</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>14</td>
<td>(A1-2)</td>
<td>1.46</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>rep</td>
</tr>
<tr>
<td>24</td>
<td>(A3-4)</td>
<td>1.39</td>
<td>-</td>
<td>-</td>
<td>infer</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>16</td>
<td>(A1-4)</td>
<td>1.02</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>infer</td>
<td>-</td>
</tr>
<tr>
<td>21</td>
<td>(A3-1)</td>
<td>0.89</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>rep</td>
</tr>
<tr>
<td>13</td>
<td>(A1-1)</td>
<td>0.77</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>rep</td>
</tr>
<tr>
<td>20</td>
<td>(A2-4)</td>
<td>0.77</td>
<td>-</td>
<td>infer</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>17</td>
<td>(A2-1)</td>
<td>0.66</td>
<td>-</td>
<td>-</td>
<td>rep</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>1</td>
<td>(S1-1)</td>
<td>0.37</td>
<td>-</td>
<td>rep</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>18</td>
<td>(A2-2)</td>
<td>0.05</td>
<td>-</td>
<td>rep</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>19</td>
<td>(A2-3)</td>
<td>0.05</td>
<td>-</td>
<td>synth</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>22</td>
<td>(A3-2)</td>
<td>-0.08</td>
<td>-</td>
<td>-</td>
<td>rep</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>9</td>
<td>(S3-1)</td>
<td>-0.32</td>
<td>rep</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>12</td>
<td>(S3-4)</td>
<td>-0.42</td>
<td>infer</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>7</td>
<td>(S2-3)</td>
<td>-0.49</td>
<td>synth</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>4</td>
<td>(S1-4)</td>
<td>-0.57</td>
<td>-</td>
<td>infer</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>6</td>
<td>(S2-2)</td>
<td>-0.81</td>
<td>rep</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>3</td>
<td>(S1-3)</td>
<td>-1.07</td>
<td>synth</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>15</td>
<td>(A1-3)</td>
<td>-1.28</td>
<td>-</td>
<td>synth</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>10</td>
<td>(S3-2)</td>
<td>-1.39</td>
<td>rep</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>5</td>
<td>(S2-1)</td>
<td>-1.50</td>
<td>rep</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>11</td>
<td>(S3-3)</td>
<td>-1.63</td>
<td>synth</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>2</td>
<td>(S1-2)</td>
<td>-2.11</td>
<td>rep</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

Note. A = authentic text; S = simplified text; Simp = simplified text; Auth = authentic text; mid = medium level of difficulty; rep = repetitive (item type); synth = synthesis (item type); infer = inferencing (item type). All statistics are based on Rasch logits.

A closer look at some of the easiest synthesis questions in this test, namely item 11 and item 15. Regarding item 11, in addition to this text being the easiest of the simplified texts, the locations of the pieces of information required to answer the question were relatively close to each other, all occurring within a span of three lines of text. The question asks, *When did someone put a bomb under Parliament?* and the word *bomb*...
only occurs once in the passage, with the phrase under your Parliament coming immediately after, and the time expressed as a numeral, 1605, only two lines later. For these reasons, it seems that the cognitive load required to synthesize this information was light. Regarding item 15, upon reflection, this item might not fit the definition of a synthesis question, but could rather be described as requiring light inferencing. The question was, What did the nurses do before meals? and the correct answer was, They opened the curtains around the beds. Although nurses were not clearly identified as the agent, and before meals was not explicitly mentioned, a single sentence in the text (When all the curtains were opened, as they were at night and for meals, the two rows of patients faced each other like chess pieces.) apparently supplied enough information for most students to answer correctly. Incidentally, the original reasoning for using this as a synthesis question was that the fact that the location and nature of these curtains (around the patients’ beds) was introduced in a previous sentence.

The Rasch item reliability estimate for this instrument was .94 (Very Good), and the Rasch item separation statistic was 3.87 (Good), indicating that the items were spread over four distinct levels of difficulty. The Rasch person reliability estimate was .61 (Poor), and the Rasch person separation statistic was 1.26 (Poor). These are rather low, indicating that many participants performed so similarly that they could not be distinguished statistically.
Reading Comprehension Test (Time 2)

Table 27 shows the descriptive statistics of the Rasch analysis of the 24 items administered at Time 2. Items 5-8 and 13-16 were initially anchored by inputting their difficulty estimates from the Time 1 administration. By anchoring the difficulty estimates for those items, the students’ performance on the two tests can be compared as if it were the same test (Bond & Fox, 2007). However, after comparing the difficulty estimates of the anchored items in Time 1 to what they would be in Time 2 if not

Table 27. Rasch Descriptive Statistics for the Reading Comprehension Test Items (Time 2)

<table>
<thead>
<tr>
<th>Item</th>
<th>Measure</th>
<th>SE</th>
<th>Infit MNSQ</th>
<th>Infit ZSTD</th>
<th>Outfit MNSQ</th>
<th>Outfit ZSTD</th>
<th>Pt-measure correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>19</td>
<td>2.47</td>
<td>0.29</td>
<td>1.06</td>
<td>0.5</td>
<td>1.62</td>
<td>2.1</td>
<td>.29</td>
</tr>
<tr>
<td>18</td>
<td>1.54</td>
<td>0.25</td>
<td>1.07</td>
<td>0.8</td>
<td>1.06</td>
<td>0.5</td>
<td>.35</td>
</tr>
<tr>
<td>8</td>
<td>1.53A</td>
<td>0.26</td>
<td>1.10</td>
<td>1.0</td>
<td>1.10</td>
<td>0.7</td>
<td>.41</td>
</tr>
<tr>
<td>14</td>
<td>1.46A</td>
<td>0.25</td>
<td>0.88</td>
<td>-1.3</td>
<td>0.82</td>
<td>-1.3</td>
<td>.51</td>
</tr>
<tr>
<td>4</td>
<td>1.10</td>
<td>0.25</td>
<td>1.18</td>
<td>1.9</td>
<td>1.28</td>
<td>1.9</td>
<td>.25</td>
</tr>
<tr>
<td>16</td>
<td>1.02A</td>
<td>0.25</td>
<td>0.92</td>
<td>-0.9</td>
<td>0.85</td>
<td>-1.1</td>
<td>.49</td>
</tr>
<tr>
<td>13</td>
<td>0.77A</td>
<td>0.25</td>
<td>1.08</td>
<td>0.9</td>
<td>1.17</td>
<td>1.2</td>
<td>.37</td>
</tr>
<tr>
<td>11</td>
<td>0.40</td>
<td>0.26</td>
<td>0.94</td>
<td>-0.6</td>
<td>1.09</td>
<td>0.6</td>
<td>.44</td>
</tr>
<tr>
<td>23</td>
<td>0.27</td>
<td>0.26</td>
<td>1.02</td>
<td>0.2</td>
<td>1.05</td>
<td>0.3</td>
<td>.38</td>
</tr>
<tr>
<td>20</td>
<td>0.20</td>
<td>0.26</td>
<td>1.01</td>
<td>0.2</td>
<td>0.90</td>
<td>-0.5</td>
<td>.40</td>
</tr>
<tr>
<td>22</td>
<td>-0.09</td>
<td>0.28</td>
<td>0.96</td>
<td>-0.3</td>
<td>0.93</td>
<td>-0.2</td>
<td>.41</td>
</tr>
<tr>
<td>3</td>
<td>-0.17</td>
<td>0.28</td>
<td>1.05</td>
<td>0.4</td>
<td>0.98</td>
<td>0.0</td>
<td>.35</td>
</tr>
<tr>
<td>10</td>
<td>-0.34</td>
<td>0.29</td>
<td>0.96</td>
<td>-0.2</td>
<td>0.84</td>
<td>-0.5</td>
<td>.41</td>
</tr>
<tr>
<td>24</td>
<td>-0.34</td>
<td>0.29</td>
<td>1.04</td>
<td>0.3</td>
<td>0.93</td>
<td>-0.2</td>
<td>.35</td>
</tr>
<tr>
<td>12</td>
<td>-0.42</td>
<td>0.30</td>
<td>0.94</td>
<td>-0.3</td>
<td>1.09</td>
<td>0.4</td>
<td>.39</td>
</tr>
<tr>
<td>17</td>
<td>-0.42</td>
<td>0.30</td>
<td>0.98</td>
<td>-0.1</td>
<td>0.89</td>
<td>-0.3</td>
<td>.39</td>
</tr>
<tr>
<td>7</td>
<td>-0.49A</td>
<td>0.30</td>
<td>1.12</td>
<td>0.8</td>
<td>1.25</td>
<td>0.9</td>
<td>.25</td>
</tr>
<tr>
<td>6</td>
<td>-0.81A</td>
<td>0.33</td>
<td>1.32</td>
<td>1.5</td>
<td>1.87</td>
<td>2.0</td>
<td>.21</td>
</tr>
<tr>
<td>1</td>
<td>-1.03</td>
<td>0.35</td>
<td>1.03</td>
<td>0.2</td>
<td>0.83</td>
<td>-0.3</td>
<td>.32</td>
</tr>
<tr>
<td>9</td>
<td>-1.44</td>
<td>0.40</td>
<td>0.88</td>
<td>-0.3</td>
<td>0.72</td>
<td>-0.4</td>
<td>.38</td>
</tr>
<tr>
<td>5</td>
<td>-1.50A</td>
<td>0.40</td>
<td>0.83</td>
<td>-0.5</td>
<td>0.65</td>
<td>-0.6</td>
<td>.27</td>
</tr>
<tr>
<td>21</td>
<td>-1.80</td>
<td>0.45</td>
<td>0.99</td>
<td>0.1</td>
<td>0.88</td>
<td>0.0</td>
<td>.27</td>
</tr>
<tr>
<td>2</td>
<td>-2.60</td>
<td>0.61</td>
<td>0.68</td>
<td>-0.5</td>
<td>0.17</td>
<td>-1.2</td>
<td>.49</td>
</tr>
<tr>
<td>15</td>
<td>-2.60</td>
<td>0.61</td>
<td>0.85</td>
<td>-0.1</td>
<td>0.27</td>
<td>-0.9</td>
<td>.39</td>
</tr>
</tbody>
</table>

Note. MNSQ = mean square; ZSTD = Z-score standardized; Pt-measure = point-measure. All statistics are based on Rasch logits.
anchored, it was found that Item 15 did not meet the < 0.50 logit displacement criterion proposed by Wright and Douglas (1976); this item was therefore left unanchored.

The fit statistics (Table 27) reveal two items with infit MNSQ statistics that fell just outside the 0.70–1.30 criterion for model fit; all of the other items fell within the suggested range. Item 6 had an infit MNSQ of 1.32. This item was an anchor item, and its infit MNSQ on the first administration was 0.97, so it was not a consistently misbehaving item. Item 2, which had an infit MNSQ of 0.68, was one of the easiest items on the test, which likely explains the slight misfit. The point-measure correlations ranged from .21 to .51.

Analysis of the dimensionality of this test revealed that the variance explained by the items was 17.3%, and the variance explained by measures was 29.3%. The unexplained variance accounted for by the first contrast was 2.6 eigenvalue units (7.5%), indicating good unidimensionality.

The Wright map in Figure 10 shows that, like the first test, there was a ceiling effect, with not enough difficult items to differentiate between the high-ability students. Item difficulty estimates ranged from -2.60 to 2.47, but the mean was -0.14 logits ($SD = 1.28$), compared with a mean person ability of 1.11 ($SD = 1.04$), indicating that overall, the test was too easy for the students.

Although not as clear as in Time 1, an inspection of the effect of text difficulty and item type on item difficulty (Table 28) revealed that the items for the authentic texts were more difficult than those for the simplified texts, and that within the authentic texts, the more difficult items were usually associated with the more difficult texts. The
Figure 10. Abridged Item-Person Map for the Reading Comprehension Test (Time 2)

More able persons | More difficult items

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>XX</td>
<td>T</td>
</tr>
<tr>
<td>3</td>
<td>XXX</td>
</tr>
<tr>
<td></td>
<td>A2-3</td>
</tr>
<tr>
<td>XXXXXX</td>
<td>T</td>
</tr>
<tr>
<td>S</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>XXXXX</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>XXXXX</td>
</tr>
<tr>
<td></td>
<td>A1-2  A2-2  S2-4</td>
</tr>
<tr>
<td>XXXXXX</td>
<td>M</td>
</tr>
<tr>
<td>1</td>
<td>XXXXXXXXXXXX</td>
</tr>
<tr>
<td>XXXXX</td>
<td>A1-1</td>
</tr>
<tr>
<td>XXXXXX</td>
<td>S3-3</td>
</tr>
<tr>
<td>XXXXX</td>
<td>A2-4  A3-3</td>
</tr>
<tr>
<td>XXX</td>
<td>S</td>
</tr>
<tr>
<td>X</td>
<td></td>
</tr>
<tr>
<td>XX</td>
<td>M  A3-2  S1-3</td>
</tr>
<tr>
<td>XXX</td>
<td>A2-1  A3-4  S3-2  S3-4</td>
</tr>
<tr>
<td>X</td>
<td>S2-2</td>
</tr>
<tr>
<td>-1</td>
<td>T+  S1-1</td>
</tr>
<tr>
<td>S</td>
<td></td>
</tr>
<tr>
<td>S2-1  S3-1</td>
<td></td>
</tr>
<tr>
<td>X</td>
<td>A3-1</td>
</tr>
<tr>
<td>-2</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>A1-3  S1-2</td>
</tr>
<tr>
<td></td>
<td>T</td>
</tr>
</tbody>
</table>

Less able persons | Less difficult items

Note. Each “X” equals 1 student. M = Mean; S = One standard deviation from the mean; T = Two standard deviations from the mean; A = authentic text; S = simplified text.
Table 28. Item Difficulty by Text Type, Text Difficulty, and Item Type for the Reading Comprehension Test (Time 2)

<table>
<thead>
<tr>
<th>Item</th>
<th>Measure</th>
<th>Simp-easy (S1)</th>
<th>Simp-mid (S2)</th>
<th>Simp-difficult (S3)</th>
<th>Auth-easy (A3)</th>
<th>Auth-mid (A2)</th>
<th>Auth-difficult (A1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>19</td>
<td>(A2-3)</td>
<td>2.47</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>synth</td>
</tr>
<tr>
<td>18</td>
<td>(A2-2)</td>
<td>1.54</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>rep</td>
</tr>
<tr>
<td>8</td>
<td>(S2-4)</td>
<td>1.53A</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>infer</td>
</tr>
<tr>
<td>14</td>
<td>(A1-2)</td>
<td>1.46A</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>rep</td>
</tr>
<tr>
<td>4</td>
<td>(S1-4)</td>
<td>1.10</td>
<td>Infer</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>(A1-4)</td>
<td>1.02A</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>infer</td>
</tr>
<tr>
<td>13</td>
<td>(A1-1)</td>
<td>0.77A</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>rep</td>
</tr>
<tr>
<td>11</td>
<td>(S3-3)</td>
<td>0.40</td>
<td></td>
<td>synth</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>(A3-3)</td>
<td>0.27</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>(A2-4)</td>
<td>0.20</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>infer</td>
</tr>
<tr>
<td>22</td>
<td>(A3-2)</td>
<td>-0.09</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>rep</td>
</tr>
<tr>
<td>3</td>
<td>(S1-3)</td>
<td>-0.17</td>
<td>synth</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>(S3-2)</td>
<td>-0.34</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>rep</td>
</tr>
<tr>
<td>24</td>
<td>(A3-4)</td>
<td>-0.34</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>infer</td>
</tr>
<tr>
<td>12</td>
<td>(S3-4)</td>
<td>-0.42</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>infer</td>
</tr>
<tr>
<td>17</td>
<td>(A2-1)</td>
<td>-0.42</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>rep</td>
</tr>
<tr>
<td>7</td>
<td>(S2-3)</td>
<td>-0.49A</td>
<td>synth</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>(S2-2)</td>
<td>-0.81A</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>rep</td>
</tr>
<tr>
<td>1</td>
<td>(S1-1)</td>
<td>-1.03</td>
<td>rep</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>(S3-1)</td>
<td>-1.44</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>rep</td>
</tr>
<tr>
<td>5</td>
<td>(S2-1)</td>
<td>-1.50A</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>rep</td>
</tr>
<tr>
<td>21</td>
<td>(A3-1)</td>
<td>-1.80</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>rep</td>
</tr>
<tr>
<td>2</td>
<td>(S1-2)</td>
<td>-2.60</td>
<td>rep</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>(A1-3)</td>
<td>-2.60</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>synth</td>
</tr>
</tbody>
</table>

Note. A = authentic text; S = simplified text; Simp = simplified text; Auth = authentic text; mid = medium level of difficulty; rep = repetitive (item type); synth = synthesis (item type); infer = inferencing (item type). All statistics are based on Rasch logits.

The effect of item type on item difficulty was less clear than in Time 1. Inferencing items were still generally more difficult than the other two types, and synthesis questions overall were a little more difficult than repetitive questions, which was not apparent in the Time 1 test. Item 15 was a synthesis question about the most difficult of the authentic texts, and yet turned out to be the easiest item on the test. A closer inspection
revealed that this is the same item 15 from the previous test, which was planned as an anchor item but left unanchored because of its high displacement value. It seems that the learning effect made an already easy item even easier.

The Rasch item reliability estimate for this test was .93 (Very Good), and the Rasch item separation statistic was 3.58 (Good), indicating that the items could be statistically separated into four distinct levels. The Rasch person reliability estimate was .67 (Fair), and the Rasch person separation statistic was 1.44 (Poor).

**Reading Comprehension Test (Time 3)**

The descriptive statistics of the Rasch analysis of the Time 3 reading comprehension test are displayed in Table 29. Of the eight items intended to be anchored to the Time 2 test, four had difficulty measures that exceeded the 0.50 logit displacement criteria, so only four items were retained as anchored items.

All of the items demonstrated good fit to the Rasch model. The infit MNSQ statistic ranged from 0.85 to 1.23, and the point-measure correlations were between .00 and .49.

Regarding dimensionality, variance explained by items was 16.9%, and variance explained by measures was 27.1%. Variance explained by the first contrast was 2.3 eigenvalue units (7.0%), which indicates good unidimensionality.

The item-person map in Figure 11 shows how well item difficulty and person ability matched in this test. There was a slight ceiling effect, although it was much less pronounced than in the previous two tests. Only one student had an ability level that
Table 29. Rasch Descriptive Statistics for the Reading Comprehension Test Items (Time 3)

<table>
<thead>
<tr>
<th>Item</th>
<th>Measure</th>
<th>SE</th>
<th>Infit MNSQ</th>
<th>Infit ZSTD</th>
<th>Outfit MNSQ</th>
<th>Outfit ZSTD</th>
<th>Pt-measure correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>14</td>
<td>2.18</td>
<td>0.28</td>
<td>0.90</td>
<td>-0.7</td>
<td>0.82</td>
<td>-0.7</td>
<td>.43</td>
</tr>
<tr>
<td>19</td>
<td>2.09</td>
<td>0.28</td>
<td>1.02</td>
<td>0.2</td>
<td>1.06</td>
<td>0.3</td>
<td>.27</td>
</tr>
<tr>
<td>7</td>
<td>1.81</td>
<td>0.26</td>
<td>1.01</td>
<td>0.1</td>
<td>1.02</td>
<td>0.2</td>
<td>.31</td>
</tr>
<tr>
<td>18</td>
<td>1.54A</td>
<td>0.25</td>
<td>1.23</td>
<td>2.2</td>
<td>1.26</td>
<td>1.7</td>
<td>.00</td>
</tr>
<tr>
<td>22</td>
<td>1.27</td>
<td>0.25</td>
<td>0.88</td>
<td>-1.4</td>
<td>0.87</td>
<td>-1.0</td>
<td>.48</td>
</tr>
<tr>
<td>24</td>
<td>1.09</td>
<td>0.25</td>
<td>1.12</td>
<td>1.5</td>
<td>1.11</td>
<td>0.9</td>
<td>.22</td>
</tr>
<tr>
<td>23</td>
<td>0.91</td>
<td>0.24</td>
<td>1.04</td>
<td>0.5</td>
<td>1.04</td>
<td>0.4</td>
<td>.31</td>
</tr>
<tr>
<td>8</td>
<td>0.87</td>
<td>0.24</td>
<td>1.14</td>
<td>1.7</td>
<td>1.16</td>
<td>1.5</td>
<td>.19</td>
</tr>
<tr>
<td>6</td>
<td>0.46</td>
<td>0.25</td>
<td>1.05</td>
<td>0.6</td>
<td>1.02</td>
<td>0.2</td>
<td>.31</td>
</tr>
<tr>
<td>3</td>
<td>0.27</td>
<td>0.25</td>
<td>0.88</td>
<td>-1.3</td>
<td>0.83</td>
<td>-1.4</td>
<td>.49</td>
</tr>
<tr>
<td>16</td>
<td>0.21</td>
<td>0.25</td>
<td>0.87</td>
<td>-1.4</td>
<td>0.80</td>
<td>-1.6</td>
<td>.51</td>
</tr>
<tr>
<td>20</td>
<td>0.20A</td>
<td>0.25</td>
<td>1.02</td>
<td>0.3</td>
<td>1.04</td>
<td>0.4</td>
<td>.35</td>
</tr>
<tr>
<td>15</td>
<td>0.02</td>
<td>0.26</td>
<td>0.92</td>
<td>-0.8</td>
<td>0.86</td>
<td>-1.0</td>
<td>.45</td>
</tr>
<tr>
<td>4</td>
<td>-0.19</td>
<td>0.26</td>
<td>0.96</td>
<td>-0.3</td>
<td>0.92</td>
<td>-0.4</td>
<td>.39</td>
</tr>
<tr>
<td>11</td>
<td>-0.26</td>
<td>0.27</td>
<td>0.95</td>
<td>-0.4</td>
<td>0.86</td>
<td>-0.7</td>
<td>.41</td>
</tr>
<tr>
<td>2</td>
<td>-0.33</td>
<td>0.27</td>
<td>1.01</td>
<td>0.1</td>
<td>1.19</td>
<td>1.0</td>
<td>.28</td>
</tr>
<tr>
<td>10</td>
<td>-0.34A</td>
<td>0.27</td>
<td>0.95</td>
<td>-0.3</td>
<td>0.95</td>
<td>-0.2</td>
<td>.32</td>
</tr>
<tr>
<td>17</td>
<td>-0.42A</td>
<td>0.28</td>
<td>0.92</td>
<td>-0.5</td>
<td>0.83</td>
<td>-0.8</td>
<td>.33</td>
</tr>
<tr>
<td>21</td>
<td>-0.45</td>
<td>0.28</td>
<td>0.96</td>
<td>-0.2</td>
<td>0.84</td>
<td>-0.7</td>
<td>.40</td>
</tr>
<tr>
<td>5</td>
<td>-0.48</td>
<td>0.28</td>
<td>1.02</td>
<td>0.2</td>
<td>1.24</td>
<td>1.1</td>
<td>.26</td>
</tr>
<tr>
<td>13</td>
<td>-1.01</td>
<td>0.32</td>
<td>1.01</td>
<td>0.1</td>
<td>0.97</td>
<td>0.0</td>
<td>.28</td>
</tr>
<tr>
<td>1</td>
<td>-1.45</td>
<td>0.37</td>
<td>1.01</td>
<td>0.1</td>
<td>1.06</td>
<td>0.3</td>
<td>.23</td>
</tr>
<tr>
<td>12</td>
<td>-1.48</td>
<td>0.37</td>
<td>0.85</td>
<td>-0.5</td>
<td>0.55</td>
<td>-1.3</td>
<td>.48</td>
</tr>
<tr>
<td>9</td>
<td>-2.18</td>
<td>0.48</td>
<td>1.10</td>
<td>0.4</td>
<td>0.92</td>
<td>0.1</td>
<td>.16</td>
</tr>
</tbody>
</table>

Note. MNSQ = mean square; ZSTD = Z-score standardized; Pt-measure = point-measure. All statistics are based on Rasch logits.

was higher than the most difficult items. However, five additional students had an ability level that matched the most difficult item. The difficulty estimates of the items ranged from -2.18 logits at the easiest to 2.18 logits at the most difficult, while the ability level of the students ranged from -2.18 logits at the lowest to 2.55 logits at the highest. The mean item difficulty measure was .18 logits ($SD = 1.11$), while the mean person ability measure was .80 logits ($SD = .83$).
Figure 11. Abridged Item-Person Map for the Reading Comprehension Test (Time 3)

More able persons | More difficult items

<table>
<thead>
<tr>
<th>X</th>
<th>T</th>
</tr>
</thead>
<tbody>
<tr>
<td>T</td>
<td></td>
</tr>
</tbody>
</table>
| | A1-2
| 2 |
| + |
| | A2-3
| | A3-2
| | | A3-4
| 1 | + |
| | A2-2
| | A3-3
| | | S2-3
| | | S2-2
| | S1-3
| | M |
| | | M
| 0 | + |
| | A1-3
| X |
| | S1-4
| | S1-2
| | S3-2
| | S3-3
| | A2-1
| | A3-1
| | S2-1
| | M
| | A1-4
| | A2-4
| | T |
| | S |
| -1 | + |
| | A1-1
| | S1-1
| | S3-4
| | -2 | + |
| | T
| | X
| | S3-1

Less able persons | Less difficult items

Note. Each “X” equals 1 student. M = Mean; S = One standard deviation from the mean; T = Two standard deviations from the mean; A = authentic text; S = simplified text.
Table 30 displays item difficulty by text type, text difficulty, and item type.

Questions about the authentic texts are more difficult than those about the simplified texts, with eight of the twelve most difficult items being about the authentic texts.

However, within the authentic texts and within the simplified texts, there was no

<table>
<thead>
<tr>
<th>Item</th>
<th>Measure</th>
<th>Simp-easy (S1)</th>
<th>Simp-mid (S2)</th>
<th>Simp-difficult (S3)</th>
<th>Auth-easy (A3)</th>
<th>Auth-mid (A2)</th>
<th>Auth-difficult (A1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>14</td>
<td>(A1-2)</td>
<td>2.18</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>rep</td>
</tr>
<tr>
<td>19</td>
<td>(A2-3)</td>
<td>2.09</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>synth</td>
</tr>
<tr>
<td>7</td>
<td>(S2-3)</td>
<td>1.81</td>
<td></td>
<td>synth</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>(A2-2)</td>
<td>1.54A</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>rep</td>
</tr>
<tr>
<td>22</td>
<td>(A3-2)</td>
<td>1.27</td>
<td></td>
<td></td>
<td>rep</td>
<td></td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>(A3-4)</td>
<td>1.09</td>
<td></td>
<td></td>
<td>infer</td>
<td></td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>(A3-3)</td>
<td>0.91</td>
<td></td>
<td></td>
<td>synth</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>(S2-4)</td>
<td>0.87</td>
<td></td>
<td>infer</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>(S2-2)</td>
<td>0.46</td>
<td></td>
<td>rep</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>(S1-3)</td>
<td>0.27</td>
<td>synth</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>(A1-4)</td>
<td>0.21</td>
<td></td>
<td></td>
<td>infer</td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>(A2-4)</td>
<td>0.20A</td>
<td></td>
<td></td>
<td>infer</td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>(A1-3)</td>
<td>0.02</td>
<td></td>
<td></td>
<td></td>
<td>synth</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>(S1-4)</td>
<td>-0.19</td>
<td>infer</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>(S3-3)</td>
<td>-0.26</td>
<td></td>
<td>synth</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>(S1-2)</td>
<td>-0.33</td>
<td>rep</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>(S3-2)</td>
<td>-0.34A</td>
<td></td>
<td>rep</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>(A2-1)</td>
<td>-0.42A</td>
<td></td>
<td></td>
<td>rep</td>
<td></td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>(A3-1)</td>
<td>-0.45</td>
<td></td>
<td></td>
<td>rep</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>(S2-1)</td>
<td>-0.48</td>
<td>rep</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>(A1-1)</td>
<td>-1.01</td>
<td></td>
<td></td>
<td>rep</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>(S1-1)</td>
<td>-1.45</td>
<td>rep</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>(S3-4)</td>
<td>-1.48</td>
<td></td>
<td>infer</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>(S3-1)</td>
<td>-2.18</td>
<td></td>
<td>rep</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Note.* A = authentic text; S = simplified text; Simp = simplified text; Auth = authentic text; mid = medium level of difficulty; rep = repetitive (item type); synth = synthesis (item type); infer = inferencing (item type). All statistics are based on Rasch logits.
relationship between text difficulty and item difficulty. Regarding item type, repetitive questions were easier than synthesis and inference questions, but there was no difference between synthesis and inference questions in terms of difficulty on this test. Item 12 was an inferencing question that was one of the easiest questions on the test. The reason for this seems to be threefold. First, the answer choices were high-frequency, one-word adjectives (shy, annoyed, bored, and happy), and the corresponding information in the text also consisted of high-frequency lexical items such as laughed and jumped. Second, the correct answer was a positive emotion, while the three distractors were transparently negative. Third, only light inferencing was required, as it is quite easy to infer happiness from laughter and jumping.

The Rasch item reliability estimate for this test was .93 (Very Good), and the Rasch item separation statistic was 3.68 (Good), indicating that the items could be divided into four distinct difficulty levels. Rasch person reliability estimate was .61 (Poor), and the Rasch person separation statistic was 1.25 (Poor). These are rather low, indicating that many participants performed so similarly that they could not be distinguished statistically.

**Reading Comprehension Test (Time 4)**

The descriptive statistics of the Rasch analysis of the Time 4 reading comprehension test are displayed in Table 31. Of the eight items intended to be anchored to the Time 3 test, three had difficulty measures that exceeded the 0.50 logit displacement criteria, so five items were retained as anchored items.
Table 31. Rasch Descriptive Statistics for the Reading Comprehension Test Items (Time 4)

<table>
<thead>
<tr>
<th>Item</th>
<th>Measure</th>
<th>SE</th>
<th>Infit MNSQ</th>
<th>Infit ZSTD</th>
<th>Outfit MNSQ</th>
<th>Outfit ZSTD</th>
<th>Pt-measure correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>7</td>
<td>2.43</td>
<td>0.31</td>
<td>1.15</td>
<td>0.8</td>
<td>1.28</td>
<td>1.0</td>
<td>.23</td>
</tr>
<tr>
<td>16</td>
<td>1.91</td>
<td>0.28</td>
<td>1.08</td>
<td>0.7</td>
<td>1.20</td>
<td>1.0</td>
<td>.29</td>
</tr>
<tr>
<td>2</td>
<td>1.10</td>
<td>0.25</td>
<td>1.09</td>
<td>1.0</td>
<td>1.21</td>
<td>1.7</td>
<td>.29</td>
</tr>
<tr>
<td>24</td>
<td>1.09A</td>
<td>0.25</td>
<td>0.95</td>
<td>-0.5</td>
<td>0.94</td>
<td>-0.4</td>
<td>.43</td>
</tr>
<tr>
<td>11</td>
<td>1.03</td>
<td>0.25</td>
<td>0.96</td>
<td>-0.4</td>
<td>0.96</td>
<td>-0.3</td>
<td>.42</td>
</tr>
<tr>
<td>23</td>
<td>0.91A</td>
<td>0.25</td>
<td>0.94</td>
<td>-0.7</td>
<td>0.92</td>
<td>-0.7</td>
<td>.47</td>
</tr>
<tr>
<td>8</td>
<td>0.87A</td>
<td>0.25</td>
<td>1.12</td>
<td>1.4</td>
<td>1.14</td>
<td>1.2</td>
<td>.26</td>
</tr>
<tr>
<td>4</td>
<td>0.85</td>
<td>0.25</td>
<td>1.14</td>
<td>1.6</td>
<td>1.18</td>
<td>1.6</td>
<td>.26</td>
</tr>
<tr>
<td>14</td>
<td>0.67</td>
<td>0.25</td>
<td>0.94</td>
<td>-0.7</td>
<td>0.91</td>
<td>-0.8</td>
<td>.45</td>
</tr>
<tr>
<td>22</td>
<td>0.36</td>
<td>0.25</td>
<td>1.14</td>
<td>1.5</td>
<td>1.17</td>
<td>1.4</td>
<td>.24</td>
</tr>
<tr>
<td>18</td>
<td>0.11</td>
<td>0.25</td>
<td>0.85</td>
<td>-1.6</td>
<td>0.80</td>
<td>-1.5</td>
<td>.51</td>
</tr>
<tr>
<td>10</td>
<td>-0.02</td>
<td>0.26</td>
<td>0.92</td>
<td>-0.8</td>
<td>0.91</td>
<td>-0.6</td>
<td>.43</td>
</tr>
<tr>
<td>6</td>
<td>-0.09</td>
<td>0.26</td>
<td>0.95</td>
<td>-0.5</td>
<td>0.87</td>
<td>-0.8</td>
<td>.42</td>
</tr>
<tr>
<td>21</td>
<td>-0.45A</td>
<td>0.27</td>
<td>0.91</td>
<td>-0.7</td>
<td>0.80</td>
<td>-1.0</td>
<td>.35</td>
</tr>
<tr>
<td>5</td>
<td>-0.48A</td>
<td>0.28</td>
<td>1.00</td>
<td>0.1</td>
<td>0.95</td>
<td>-0.2</td>
<td>.35</td>
</tr>
<tr>
<td>13</td>
<td>-0.52</td>
<td>0.28</td>
<td>1.03</td>
<td>0.3</td>
<td>1.08</td>
<td>0.4</td>
<td>.30</td>
</tr>
<tr>
<td>17</td>
<td>-0.52</td>
<td>0.28</td>
<td>0.96</td>
<td>-0.2</td>
<td>1.01</td>
<td>0.1</td>
<td>.35</td>
</tr>
<tr>
<td>20</td>
<td>-0.60</td>
<td>0.28</td>
<td>0.89</td>
<td>-0.8</td>
<td>0.74</td>
<td>-1.2</td>
<td>.45</td>
</tr>
<tr>
<td>12</td>
<td>-0.68</td>
<td>0.29</td>
<td>0.84</td>
<td>-1.1</td>
<td>0.66</td>
<td>-1.6</td>
<td>.49</td>
</tr>
<tr>
<td>15</td>
<td>-0.76</td>
<td>0.29</td>
<td>0.96</td>
<td>-0.2</td>
<td>1.06</td>
<td>0.3</td>
<td>.33</td>
</tr>
<tr>
<td>3</td>
<td>-1.35</td>
<td>0.34</td>
<td>1.10</td>
<td>0.5</td>
<td>1.30</td>
<td>.9</td>
<td>.13</td>
</tr>
<tr>
<td>19</td>
<td>-1.47</td>
<td>0.35</td>
<td>1.00</td>
<td>0.1</td>
<td>0.86</td>
<td>-0.3</td>
<td>.27</td>
</tr>
<tr>
<td>1</td>
<td>-1.60</td>
<td>0.37</td>
<td>1.04</td>
<td>0.2</td>
<td>1.28</td>
<td>0.8</td>
<td>.16</td>
</tr>
<tr>
<td>9</td>
<td>-2.53</td>
<td>0.52</td>
<td>0.92</td>
<td>0.0</td>
<td>0.50</td>
<td>-0.7</td>
<td>.29</td>
</tr>
</tbody>
</table>

Note. MNSQ = mean square; ZSTD = Z-score standardized; Pt-measure = point-measure. All statistics are based on Rasch logits.

Regarding model fit, all of the inﬁt MNSQ statistics fell within the 0.70–1.30 range. The inﬁt MNSQ statistic ranged from 0.84 to 1.15, and the point-measure correlations were between .13 and .51.

Regarding dimensionality, the variance explained by items was 17.9%, and the variance explained by measures was 27.9%. Variance explained by the ﬁrst contrast was 2.4 eigenvalue units (7.2%), a good indication of unidimensionality.
The item-person map in Figure 12 shows how well the item difficulty and person ability estimates matched in this test. There was a stronger ceiling effect, similar to that in the first and second tests, with 21 of the 79 students (26.6%) having ability estimates higher than 22 of the 24 items. The item difficulty estimates ranged from -2.53 logits at the easiest to 2.43 logits at the most difficult, while the students’ ability level ranged from -1.36 logits at the lowest to 4.98 logits at the highest. The mean item difficulty measure was .01 logits ($SD = 1.14$), while the mean person ability measure was .75 logits ($SD = 1.00$).

Table 32 displays item difficulty by text type, text difficulty, and item type. There was no discernable difference in item difficulty due to text type or text difficulty within each text type. Regarding item type, repetitive questions were again easier than synthesis and inference questions, with seven of the eight most difficult questions being synthesis and inference questions. However, there was no clear difference between synthesis and inference questions in terms of difficulty on this test. There was also a pocket of five synthesis and inference questions at the easy end of the scale.

Of particular interest is the fact that the difference in item difficulty between the simplified texts and authentic texts was less clear than in the previous three tests. A closer examination of a few items that exemplify this phenomenon was warranted. Item 2 was a repetitive item on the easiest of the simplified texts, but was the third most difficult item of the entire test. The question is *Where did the young man sit in the end?* In the text, the father of the main character said to the young man, *There is a place for you behind us*, after which the young man’s face turns red and he *moved angrily to the*
Figure 12. Abridged Item-Person Map for the Reading Comprehension Test (Time 4)

More able persons | More difficult items

| T | XXXX | S2-3 |
| T |

| 2 | XXXX + | A1-4 |
| XXXX | S |
| XXXX | S |
| XXXX | S |
| 1 | XXXX | A3-4 S1-2 |
| XXXXXXX | A3-3 S2-4 | S1-4 |
| XXXXXXX | M | A1-2 |

| XXXXXXX | A3-2 |
| XXXXXXX | A2-2 |
| 0 | XXXX +M | S3-2 |
| XXXXX | S |
| XXXXX | A3-1 |
| XXX | A1-1 A2-1 S2-1 |
| XX | A2-4 |
| S3-4 |
| A1-3 |
| XXX | -1 |
| -1 |
| T | S |

Less able persons | Less difficult items

| T |
| S1-3 |
| X |
| A2-3 |
| S1-1 |
| -2 |
| T |
| S3-1 |

Note. Each “X” equals 1 student. M = Mean; S = One standard deviation from the mean; T = Two standard deviations from the mean; A = authentic text; S = simplified text.
Table 32. Item Difficulty by Text Type, Text Difficulty, and Item Type for the Reading Comprehension Test (Time 4)

<table>
<thead>
<tr>
<th>Item</th>
<th>Measure</th>
<th>Simp-easy (S1)</th>
<th>Simp-mid (S2)</th>
<th>Simp-difficult (S3)</th>
<th>Auth-easy (A3)</th>
<th>Auth-mid (A2)</th>
<th>Auth-difficult (A1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>7</td>
<td>S2-3</td>
<td>2.43</td>
<td></td>
<td>synth</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>A1-4</td>
<td>1.91</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>S1-2</td>
<td>1.10</td>
<td>rep</td>
<td></td>
<td></td>
<td></td>
<td>infer</td>
</tr>
<tr>
<td>24</td>
<td>A3-4</td>
<td>1.09A</td>
<td></td>
<td></td>
<td>infer</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>S3-3</td>
<td>1.03</td>
<td></td>
<td>synth</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>A3-3</td>
<td>0.91A</td>
<td></td>
<td></td>
<td>synth</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>S2-4</td>
<td>0.87A</td>
<td>infer</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>S1-4</td>
<td>0.85</td>
<td>infer</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>A1-2</td>
<td>0.67</td>
<td></td>
<td></td>
<td></td>
<td>rep</td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>A3-2</td>
<td>0.36</td>
<td></td>
<td></td>
<td>rep</td>
<td></td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>A2-2</td>
<td>0.11</td>
<td></td>
<td></td>
<td>rep</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>S3-2</td>
<td>-0.02</td>
<td></td>
<td>rep</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>S2-2</td>
<td>-0.09</td>
<td>rep</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>A3-1</td>
<td>-0.45A</td>
<td></td>
<td>rep</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>S2-1</td>
<td>-0.48A</td>
<td></td>
<td>rep</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>A1-1</td>
<td>-0.52</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>rep</td>
</tr>
<tr>
<td>17</td>
<td>A2-1</td>
<td>-0.52</td>
<td></td>
<td></td>
<td>rep</td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>A2-4</td>
<td>-0.60</td>
<td></td>
<td></td>
<td>infer</td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>S3-4</td>
<td>-0.68</td>
<td></td>
<td>infer</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>A1-3</td>
<td>-0.76</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>synth</td>
</tr>
<tr>
<td>3</td>
<td>S1-3</td>
<td>-1.35</td>
<td>synth</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>A2-3</td>
<td>-1.47</td>
<td></td>
<td></td>
<td>synth</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>S1-1</td>
<td>-1.60</td>
<td>rep</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>S3-1</td>
<td>-2.53</td>
<td></td>
<td>rep</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. A = authentic text; S = simplified text; Simp = simplified text; Auth = authentic text; mid = medium level of difficulty; rep = repetitive (item type); synth = synthesis (item type); infer = inferencing (item type). All statistics are based on Rasch logits.

empty seat. The correct answer choice was farther from the stage, but the change in wording from behind the family to farther from the stage might have been cognitively difficult for many students. In addition, one of the two distractors that was often chosen incorrectly was next to Alexis and his family, so some students might not have realized that next to was close in meaning to near. Item 19, on the other hand, was a synthesis
question from the medium-level authentic text. Upon closer inspection, this item can indeed be considered a synthesis question, as answering correctly required carefully sifting through, and understanding the meaning of, four symptoms described over the course of five lines of text. This was an exclusion question, *Which of these symptoms did the author NOT mention?*, but some of the distractors (the symptoms that WERE described in the text) had wording considerably different than that in the text, so finding the correct answer would have required more thought than simply finding a single repeated word. Therefore, it is unclear why this question was so easy.

The Rasch item reliability for this test was .93 (Very Good), and the Rasch item separation was 3.69 (Good), indicating that the items were again spread over about four difficulty levels. The Rasch person reliability estimate was .69 (Fair), and the Rasch person separation statistic was 1.48 (Poor).

**Vocabulary Size Tests**

The vocabulary size tests were administered four times. Each test consisted of five items each from the first through seventh most frequent 1,000-word bands, for a total of 35 items. However, on the Time 4 test, two of the 35 items had to be discarded (due to an error in the correct answer choice that made one item difficult to answer correctly and a printing error that caused two of the four answer choices in another item not to be printed), resulting in only 33 items being analyzed here.

In order to avoid the learning effect that would likely occur with four administrations of the same test, while also allowing for the analysis of validity and
reliability across the test forms, on the second, third, and fourth administrations 10 of the 35 items were retained from previous administrations to be used as anchor items. The determination of which items to use as anchors was made after analyzing the item difficulty estimates and fit statistics of the previous administration. Anchored items were carefully selected based on their difficulty estimates in order to ensure a good range of difficulty amount the items while avoiding using misfitting items as anchors and avoiding using items that were too easy or too difficult. The resulting anchoring plan is presented in Table 33.

<table>
<thead>
<tr>
<th>Time 1</th>
<th>Time 2</th>
<th>Time 3</th>
<th>Time 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 (2000-1) period</td>
<td>1 (1000-1) jump</td>
<td>1 (1000-1) joke</td>
<td>1 (1000-1) extra</td>
</tr>
<tr>
<td>2 (1000-1) see</td>
<td>2 (1000-2) shoe</td>
<td>2 (1000-2) shoe*</td>
<td>2 (1000-2) visit</td>
</tr>
<tr>
<td>3 (1000-2) poor</td>
<td>3 (1000-3) see* (Time 1, Item 2)</td>
<td>3 (1000-3) fine</td>
<td>3 (1000-3) each</td>
</tr>
<tr>
<td>7 (2000-3) circle</td>
<td>7 (1000-4) stone</td>
<td>7 (1000-4) eye</td>
<td>7 (1000-4) true</td>
</tr>
<tr>
<td>8 (1000-4) figure</td>
<td>8 (2000-4) maintain</td>
<td>8 (2000-4) guarantee</td>
<td>8 (2000-4) decent</td>
</tr>
<tr>
<td>9 (2000-4) soldier</td>
<td>9 (5000-1) microphone* (Time 1, Item 4)</td>
<td>9 (5000-1) accessory</td>
<td>9 (5000-1) enzyme</td>
</tr>
<tr>
<td>10 (3000-1) restore</td>
<td>10 (3000-1) pave* (Time 1, Item 12)</td>
<td>10 (3000-1) authority</td>
<td>10 (3000-1) myth</td>
</tr>
<tr>
<td>12 (3000-2) pave</td>
<td>12 (3000-2) remedy</td>
<td>12 (3000-2) remedy* (Time 2, Item 12)</td>
<td>12 (3000-2) hypothesis</td>
</tr>
<tr>
<td>13 (2000-5) pro</td>
<td>13 (4000-1) strap* (Time 1, Item 15)</td>
<td>13 (4000-1) authentic</td>
<td>13 (4000-1) fracture</td>
</tr>
</tbody>
</table>

Table 33 continues.
<table>
<thead>
<tr>
<th>Table 33 (continued)</th>
<th>14 (3000-3) restore* (Time 1, Item 10)</th>
<th>14 (3000-3) compound* (Time 1, Item 14)</th>
<th>14 (3000-3) scandal</th>
</tr>
</thead>
<tbody>
<tr>
<td>15 (4000-1) strap</td>
<td>15 (3000-4) bacterium</td>
<td>15 (3000-4) input</td>
<td>15 (3000-4) allege* (Time 1, Item 16)</td>
</tr>
<tr>
<td>16 (3000-4) allege</td>
<td>16 (5000-2) rove</td>
<td>16 (5000-2) compost</td>
<td>16 (5000-2) compost* (Time 3, Item 16)</td>
</tr>
<tr>
<td>17 (5000-2) crab</td>
<td>17 (1000-5) dig* (Time 1, Item 11)</td>
<td>17 (1000-5) system</td>
<td>17 (1000-5) plant</td>
</tr>
<tr>
<td>18 (5000-3) dinosaur</td>
<td>18 (5000-3) miniature</td>
<td>18 (5000-3) awe</td>
<td>18 (5000-3) nun* (Time 1, Item 23)</td>
</tr>
<tr>
<td>19 (3000-5) deficit</td>
<td>19 (3000-5) peasantry</td>
<td>19 (3000-5) latter</td>
<td>19 (3000-5) latter* (Time 3, Item 19)</td>
</tr>
<tr>
<td>20 (4000-2) weep</td>
<td>20 (4000-2) dash</td>
<td>20 (4000-2) threshold</td>
<td>20 (4000-2) threshold* (Time 3, Item 20)</td>
</tr>
<tr>
<td>21 (4000-3) haunt</td>
<td>21 (4000-3) peel</td>
<td>21 (4000-3) peel* (Time 2, Item 21)</td>
<td>21 (4000-3) patience</td>
</tr>
<tr>
<td>22 (4000-4) cube</td>
<td>22 (4000-4) thesis</td>
<td>22 (4000-4) thesis* (Time 2, Item 22)</td>
<td>22 (4000-4) chaos</td>
</tr>
<tr>
<td>23 (5000-4) nun</td>
<td>23 (5000-4) scrub</td>
<td>23 (5000-4) eclipse</td>
<td>23 (5000-4) scrub (Time 2, Item 23)</td>
</tr>
<tr>
<td>24 (6000-1) strangle</td>
<td>24 (6000-1) palette</td>
<td>24 (6000-1) strangle* (Time 1, Item 24)</td>
<td>24 (6000-1) unravel</td>
</tr>
<tr>
<td>25 (7000-1) marrow</td>
<td>25 (7000-1) puritan</td>
<td>25 (7000-1) ruble</td>
<td>25 (7000-1) reptile</td>
</tr>
<tr>
<td>26 (6000-2) jug</td>
<td>26 (6000-2) jug* (Time 1, Item 26)</td>
<td>26 (6000-2) multitudinous</td>
<td>26 (6000-2) multitudinous* (Time 3, Item 26)</td>
</tr>
<tr>
<td>27 (7000-2) cavalier</td>
<td>27 (7000-2) erratic</td>
<td>27 (7000-2) haze</td>
<td>27 (7000-2) whim</td>
</tr>
<tr>
<td>28 (6000-3) malign</td>
<td>28 (6000-3) quiz</td>
<td>28 (6000-3) malign* (Time 1, Item 28)</td>
<td>28 (6000-3) gall</td>
</tr>
<tr>
<td>29 (4000-5) olive</td>
<td>29 (4000-5) olive* (Time 1, Item 29)</td>
<td>29 (4000-5) demography</td>
<td>29 (4000-5) diagnosis</td>
</tr>
<tr>
<td>30 (6000-4) quilt</td>
<td>30 (6000-4) bloc</td>
<td>30 (6000-4) bloc* (Time 2, Item 30)</td>
<td>30 (6000-4) curtail</td>
</tr>
<tr>
<td>31 (5000-5) shudder</td>
<td>31 (5000-5) vocabulary</td>
<td>31 (5000-5) feasible</td>
<td>31 (5000-5) asylum</td>
</tr>
<tr>
<td>32 (7000-3) veer</td>
<td>32 (7000-3) veer* (Time 1, Item 32)</td>
<td>32 (7000-3) maxim</td>
<td>32 (7000-3) avalanche</td>
</tr>
<tr>
<td>33 (6000-5) bristle</td>
<td>33 (6000-5) mumble</td>
<td>33 (6000-5) saga</td>
<td>33 (6000-5) ambush</td>
</tr>
<tr>
<td>34 (7000-4) yogurt</td>
<td>34 (7000-4) null</td>
<td>34 (7000-4) egalitarian</td>
<td>34 (7000-4) estuary</td>
</tr>
<tr>
<td>35 (7000-5) candid</td>
<td>35 (7000-5) stealth</td>
<td>35 (7000-5) stealth* (Time 2, Item 35)</td>
<td>35 (7000-5) hooker</td>
</tr>
</tbody>
</table>

Note. Asterisks indicate anchored items. Boldfaced items were discarded due to errors in the test.
Vocabulary Size Test (Time 1)

The Rasch descriptive statistics for the 35 items administered at Time 1 are reported in Table 34. Using 0.70–1.30 as the acceptable range of item fit for low-stakes tests (Wright et al., 1994), all of the items demonstrated acceptable fit to the Rasch model. The infit MNSQ statistic ranged from 0.84–1.14, and the point-measure correlations ranged from -.03–.53.

The dimensionality of this test was investigated using the Linacre (2010) criteria described above. The variance explained by the items was 27.5%, which was more than four times the variance accounted for by the first contrast (4.5%), indicating good unidimensionality. The variance explained by measures was 34.0%. The unexplained variance explained by the first contrast was 2.2 eigenvalue units (4.5%), indicating a high likelihood of unidimensionality.

The item-person map for Time 1 (Figure 13) shows how well the item difficulty estimates matched the student ability estimates. The students’ ability estimates are displayed on the left side, with each student represented as an X, and the items are on the right side. Students with higher ability are shown at the top, and those with lower ability are at the bottom. Similarly, more difficult items are at the top and easier items are at the bottom. Overall, the range of item difficulty on this test covered the ability range of the students well, with four of the items at a higher level than the highest-performing student, and 11 items easier than the lowest-performing students. The mean of the person ability estimates was 0.38 logits ($SD = 0.60$), and the mean of the item difficulty estimates was -0.31 logits ($SD = 1.96$). Generally speaking, the less frequent
Table 34. Rasch Descriptive Statistics for the Vocabulary Size Test Items (Time 1)

<table>
<thead>
<tr>
<th>Item</th>
<th>Measure</th>
<th>SE</th>
<th>Infit MNSQ</th>
<th>Infit ZSTD</th>
<th>Outfit MNSQ</th>
<th>Outfit ZSTD</th>
<th>Pt-measure correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>33</td>
<td>3.76</td>
<td>0.59</td>
<td>1.00</td>
<td>0.2</td>
<td>0.91</td>
<td>0.1</td>
<td>.13</td>
</tr>
<tr>
<td>31</td>
<td>2.43</td>
<td>0.34</td>
<td>1.04</td>
<td>0.2</td>
<td>1.06</td>
<td>0.3</td>
<td>.12</td>
</tr>
<tr>
<td>32</td>
<td>2.31</td>
<td>0.33</td>
<td>1.03</td>
<td>0.2</td>
<td>1.09</td>
<td>0.4</td>
<td>.13</td>
</tr>
<tr>
<td>16</td>
<td>1.76</td>
<td>0.28</td>
<td>1.01</td>
<td>0.1</td>
<td>0.95</td>
<td>-0.2</td>
<td>.23</td>
</tr>
<tr>
<td>24</td>
<td>1.69</td>
<td>0.28</td>
<td>1.08</td>
<td>0.6</td>
<td>1.17</td>
<td>0.9</td>
<td>.08</td>
</tr>
<tr>
<td>27</td>
<td>1.54</td>
<td>0.27</td>
<td>1.12</td>
<td>0.9</td>
<td>1.21</td>
<td>1.2</td>
<td>.02</td>
</tr>
<tr>
<td>4</td>
<td>1.36</td>
<td>0.26</td>
<td>1.02</td>
<td>0.2</td>
<td>1.04</td>
<td>0.3</td>
<td>.21</td>
</tr>
<tr>
<td>14</td>
<td>1.02</td>
<td>0.24</td>
<td>0.95</td>
<td>-0.5</td>
<td>0.95</td>
<td>-0.4</td>
<td>.34</td>
</tr>
<tr>
<td>15</td>
<td>0.90</td>
<td>0.24</td>
<td>0.95</td>
<td>-0.6</td>
<td>0.92</td>
<td>-0.8</td>
<td>.36</td>
</tr>
<tr>
<td>35</td>
<td>0.79</td>
<td>0.24</td>
<td>0.95</td>
<td>-0.6</td>
<td>0.96</td>
<td>-0.5</td>
<td>.34</td>
</tr>
<tr>
<td>23</td>
<td>0.68</td>
<td>0.24</td>
<td>0.99</td>
<td>0.0</td>
<td>0.98</td>
<td>-0.2</td>
<td>.29</td>
</tr>
<tr>
<td>21</td>
<td>0.66</td>
<td>0.23</td>
<td>0.94</td>
<td>-1.0</td>
<td>0.95</td>
<td>-0.7</td>
<td>.37</td>
</tr>
<tr>
<td>12</td>
<td>0.62</td>
<td>0.24</td>
<td>1.03</td>
<td>0.6</td>
<td>1.04</td>
<td>0.5</td>
<td>.22</td>
</tr>
<tr>
<td>19</td>
<td>0.62</td>
<td>0.24</td>
<td>0.95</td>
<td>-0.8</td>
<td>0.95</td>
<td>-0.7</td>
<td>.36</td>
</tr>
<tr>
<td>25</td>
<td>0.51</td>
<td>0.23</td>
<td>1.14</td>
<td>2.1</td>
<td>1.15</td>
<td>2.1</td>
<td>.05</td>
</tr>
<tr>
<td>30</td>
<td>0.51</td>
<td>0.23</td>
<td>0.84</td>
<td>-2.6</td>
<td>0.82</td>
<td>-2.6</td>
<td>.53</td>
</tr>
<tr>
<td>10</td>
<td>-0.06</td>
<td>0.24</td>
<td>0.97</td>
<td>-0.4</td>
<td>0.98</td>
<td>-0.1</td>
<td>.31</td>
</tr>
<tr>
<td>28</td>
<td>-0.10</td>
<td>0.24</td>
<td>0.97</td>
<td>-0.4</td>
<td>0.94</td>
<td>-0.6</td>
<td>.34</td>
</tr>
<tr>
<td>20</td>
<td>-0.16</td>
<td>0.24</td>
<td>0.98</td>
<td>-0.2</td>
<td>0.96</td>
<td>-0.4</td>
<td>.31</td>
</tr>
<tr>
<td>7</td>
<td>-0.17</td>
<td>0.24</td>
<td>1.11</td>
<td>1.3</td>
<td>1.10</td>
<td>1.0</td>
<td>.11</td>
</tr>
<tr>
<td>13</td>
<td>-0.35</td>
<td>0.25</td>
<td>0.87</td>
<td>-1.4</td>
<td>0.87</td>
<td>-1.1</td>
<td>.47</td>
</tr>
<tr>
<td>26</td>
<td>-0.40</td>
<td>0.25</td>
<td>1.00</td>
<td>0.1</td>
<td>1.00</td>
<td>0.0</td>
<td>.26</td>
</tr>
<tr>
<td>1</td>
<td>-1.03</td>
<td>0.28</td>
<td>1.03</td>
<td>0.2</td>
<td>1.03</td>
<td>0.2</td>
<td>.19</td>
</tr>
<tr>
<td>8</td>
<td>-1.03</td>
<td>0.28</td>
<td>1.14</td>
<td>0.9</td>
<td>1.30</td>
<td>1.5</td>
<td>-.03</td>
</tr>
<tr>
<td>29</td>
<td>-1.47</td>
<td>0.32</td>
<td>0.97</td>
<td>-0.1</td>
<td>0.92</td>
<td>-0.2</td>
<td>.27</td>
</tr>
<tr>
<td>17</td>
<td>-1.57</td>
<td>0.33</td>
<td>1.07</td>
<td>0.4</td>
<td>1.04</td>
<td>0.2</td>
<td>.09</td>
</tr>
<tr>
<td>2</td>
<td>-1.81</td>
<td>0.36</td>
<td>0.99</td>
<td>0.1</td>
<td>1.13</td>
<td>0.5</td>
<td>.16</td>
</tr>
<tr>
<td>6</td>
<td>-1.81</td>
<td>0.36</td>
<td>0.96</td>
<td>-0.1</td>
<td>0.94</td>
<td>-0.1</td>
<td>.26</td>
</tr>
<tr>
<td>22</td>
<td>-1.82</td>
<td>0.36</td>
<td>0.95</td>
<td>-0.1</td>
<td>0.84</td>
<td>-0.4</td>
<td>.28</td>
</tr>
<tr>
<td>11</td>
<td>-2.11</td>
<td>0.40</td>
<td>0.95</td>
<td>0.0</td>
<td>0.85</td>
<td>-0.3</td>
<td>.26</td>
</tr>
<tr>
<td>9</td>
<td>-2.28</td>
<td>0.43</td>
<td>0.96</td>
<td>0.0</td>
<td>0.84</td>
<td>-0.2</td>
<td>.25</td>
</tr>
<tr>
<td>34</td>
<td>-2.28</td>
<td>0.43</td>
<td>1.03</td>
<td>0.2</td>
<td>0.91</td>
<td>-0.1</td>
<td>.14</td>
</tr>
<tr>
<td>18</td>
<td>-2.72</td>
<td>0.52</td>
<td>1.02</td>
<td>0.2</td>
<td>1.07</td>
<td>0.3</td>
<td>.08</td>
</tr>
<tr>
<td>3</td>
<td>-5.38</td>
<td>1.83</td>
<td>Minimum Measure</td>
<td>.00</td>
<td>Minimum Measure</td>
<td>.00</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>-5.38</td>
<td>1.83</td>
<td>Minimum Measure</td>
<td>.00</td>
<td>Minimum Measure</td>
<td>.00</td>
<td></td>
</tr>
</tbody>
</table>

Note. MNSQ = mean square; ZSTD = Z-score standardized; Pt-measure = point-measure. Minimum Measure means that all students answered the item correctly. All statistics are based on Rasch logits.
**Figure 13. Abridged Item-Person Map for the Vocabulary Size Test (Time 1)**

<table>
<thead>
<tr>
<th>More able persons</th>
<th>More difficult items</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>5000–5</td>
</tr>
<tr>
<td></td>
<td>7000–3</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3000–4</td>
</tr>
<tr>
<td></td>
<td>6000–1</td>
</tr>
<tr>
<td></td>
<td>S 7000–2</td>
</tr>
<tr>
<td></td>
<td>5000–1</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>XXXX S+ 3000–3</td>
</tr>
<tr>
<td></td>
<td>XXXXXXXX 4000–1</td>
</tr>
<tr>
<td></td>
<td>XXXXXXXXX 7000–5</td>
</tr>
<tr>
<td></td>
<td>XXXXXXXXXXXXXXX 3000–2000–3 4000–2 6000–3</td>
</tr>
<tr>
<td></td>
<td>M</td>
</tr>
<tr>
<td></td>
<td>XXXXXXXXXX 4000–1</td>
</tr>
<tr>
<td></td>
<td>XXXXXXXX 7000–5</td>
</tr>
<tr>
<td></td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>M 3000–1</td>
</tr>
<tr>
<td></td>
<td>XXXX 2000–3 4000–2 6000–3</td>
</tr>
<tr>
<td></td>
<td>XXXX S 2000–5 6000–2</td>
</tr>
<tr>
<td></td>
<td>XXXXX</td>
</tr>
<tr>
<td></td>
<td>XXX</td>
</tr>
<tr>
<td></td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>XX T</td>
</tr>
<tr>
<td></td>
<td>-1</td>
</tr>
<tr>
<td></td>
<td>X 1000–4 2000–1</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>-2</td>
</tr>
<tr>
<td></td>
<td>1000–5</td>
</tr>
<tr>
<td></td>
<td>2000–4 7000–4</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>-3</td>
</tr>
<tr>
<td></td>
<td>+ 1000–2 1000–3</td>
</tr>
</tbody>
</table>

**Less able persons** | **Less difficult items**

**Note.** Each “X” equals 1 student. M = Mean; S = One standard deviation from the mean; T = Two standard deviations from the mean.
words were more difficult, but there were a number of notable exceptions: _dinosaur_ (5th 1,000-word list) and _yogurt_ (7th 1,000-word list) were nearly uniformly known, likely because they are commonly used as foreign loanwords in Japanese.

The Rasch item reliability estimate for this instrument was .92 (Very Good), and the Rasch item separation statistic was 3.51 (Good), indicating that the items were spread over four distinct levels of difficulty. The Rasch person reliability estimate was .41 (Poor), and the Rasch person separation statistic was .84 (Poor). These are rather low, indicating that many participants performed so similarly that they could not be distinguished statistically.

**Vocabulary Size Test (Time 2)**

The descriptive statistics of the Rasch analysis of the Time 2 vocabulary size test are displayed in Table 35. Of the ten items intended to be anchored to the Time 1 test, five had difficulty measures that exceeded the .50 logit displacement criteria, so five items were retained as anchored items.

Using 0.70–1.30 as the acceptable range of item fit for low-stakes tests (Wright et al., 1994), all of the items demonstrated acceptable fit to the Rasch model. The infit MNSQ statistic ranged from 0.82–1.18, and the point-measure correlations ranged from -.03 to .42.

Regarding dimensionality, the variance explained by the items was 28.3%, which was more than four times the variance accounted for by the first contrast (5.2%), indicating good unidimensionality. The variance explained by measures was 33.8%.
Table 35. Rasch Descriptive Statistics for the Vocabulary Size Test Items (Time 2)

<table>
<thead>
<tr>
<th>Item</th>
<th>Measure</th>
<th>SE</th>
<th>Infit MNSQ</th>
<th>Infit ZSTD</th>
<th>Outfit MNSQ</th>
<th>Outfit ZSTD</th>
<th>Pt-measure correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>16</td>
<td>3.61</td>
<td>0.47</td>
<td>0.97</td>
<td>0.1</td>
<td>0.86</td>
<td>-0.2</td>
<td>.42</td>
</tr>
<tr>
<td>32</td>
<td>2.31A</td>
<td>0.29</td>
<td>0.93</td>
<td>-0.4</td>
<td>0.97</td>
<td>-0.1</td>
<td>.25</td>
</tr>
<tr>
<td>19</td>
<td>2.18</td>
<td>0.28</td>
<td>1.00</td>
<td>0.0</td>
<td>1.02</td>
<td>0.2</td>
<td>.30</td>
</tr>
<tr>
<td>10</td>
<td>1.69</td>
<td>0.25</td>
<td>1.06</td>
<td>0.6</td>
<td>1.07</td>
<td>0.6</td>
<td>.22</td>
</tr>
<tr>
<td>33</td>
<td>1.57</td>
<td>0.25</td>
<td>0.94</td>
<td>-0.6</td>
<td>0.97</td>
<td>-0.2</td>
<td>.34</td>
</tr>
<tr>
<td>35</td>
<td>1.28</td>
<td>0.24</td>
<td>1.02</td>
<td>0.3</td>
<td>1.01</td>
<td>0.2</td>
<td>.25</td>
</tr>
<tr>
<td>9</td>
<td>1.17</td>
<td>0.23</td>
<td>0.99</td>
<td>-0.2</td>
<td>0.99</td>
<td>-0.1</td>
<td>.29</td>
</tr>
<tr>
<td>27</td>
<td>1.17</td>
<td>0.23</td>
<td>1.07</td>
<td>1.0</td>
<td>1.09</td>
<td>1.2</td>
<td>.18</td>
</tr>
<tr>
<td>18</td>
<td>1.11</td>
<td>0.23</td>
<td>0.97</td>
<td>-0.4</td>
<td>0.98</td>
<td>-0.3</td>
<td>.30</td>
</tr>
<tr>
<td>34</td>
<td>1.00</td>
<td>0.23</td>
<td>1.03</td>
<td>0.6</td>
<td>1.06</td>
<td>0.9</td>
<td>.21</td>
</tr>
<tr>
<td>13</td>
<td>0.90A</td>
<td>0.23</td>
<td>1.06</td>
<td>1.0</td>
<td>1.05</td>
<td>0.8</td>
<td>.20</td>
</tr>
<tr>
<td>22</td>
<td>0.79</td>
<td>0.23</td>
<td>0.98</td>
<td>-0.4</td>
<td>1.00</td>
<td>0.0</td>
<td>.27</td>
</tr>
<tr>
<td>11</td>
<td>0.70</td>
<td>0.23</td>
<td>0.96</td>
<td>-0.8</td>
<td>0.94</td>
<td>-0.9</td>
<td>.31</td>
</tr>
<tr>
<td>21</td>
<td>0.57</td>
<td>0.23</td>
<td>1.00</td>
<td>0.1</td>
<td>0.99</td>
<td>-0.1</td>
<td>.24</td>
</tr>
<tr>
<td>14</td>
<td>0.47</td>
<td>0.23</td>
<td>1.18</td>
<td>2.8</td>
<td>1.24</td>
<td>3.0</td>
<td>.00</td>
</tr>
<tr>
<td>30</td>
<td>0.47</td>
<td>0.23</td>
<td>0.96</td>
<td>-0.6</td>
<td>0.97</td>
<td>-0.4</td>
<td>.28</td>
</tr>
<tr>
<td>25</td>
<td>-0.10</td>
<td>0.25</td>
<td>0.89</td>
<td>-1.1</td>
<td>0.83</td>
<td>-1.4</td>
<td>.36</td>
</tr>
<tr>
<td>5</td>
<td>-0.38</td>
<td>0.26</td>
<td>0.95</td>
<td>-0.4</td>
<td>0.93</td>
<td>-0.4</td>
<td>.26</td>
</tr>
<tr>
<td>26</td>
<td>-0.40A</td>
<td>0.27</td>
<td>1.08</td>
<td>0.6</td>
<td>1.05</td>
<td>0.4</td>
<td>.25</td>
</tr>
<tr>
<td>12</td>
<td>-0.45</td>
<td>0.27</td>
<td>0.97</td>
<td>-0.2</td>
<td>0.89</td>
<td>-0.6</td>
<td>.25</td>
</tr>
<tr>
<td>23</td>
<td>-0.68</td>
<td>0.29</td>
<td>0.98</td>
<td>-0.1</td>
<td>1.09</td>
<td>0.5</td>
<td>.18</td>
</tr>
<tr>
<td>6</td>
<td>-1.14</td>
<td>0.33</td>
<td>0.98</td>
<td>0.0</td>
<td>0.90</td>
<td>-0.3</td>
<td>.19</td>
</tr>
<tr>
<td>24</td>
<td>-1.25</td>
<td>0.34</td>
<td>0.96</td>
<td>-0.1</td>
<td>0.82</td>
<td>-0.5</td>
<td>.22</td>
</tr>
<tr>
<td>15</td>
<td>-1.38</td>
<td>0.36</td>
<td>0.96</td>
<td>0.0</td>
<td>0.80</td>
<td>-0.6</td>
<td>.22</td>
</tr>
<tr>
<td>29</td>
<td>-1.47A</td>
<td>0.37</td>
<td>0.82</td>
<td>-0.6</td>
<td>0.95</td>
<td>0.0</td>
<td>.01</td>
</tr>
<tr>
<td>20</td>
<td>-1.51</td>
<td>0.38</td>
<td>0.99</td>
<td>0.1</td>
<td>0.93</td>
<td>-0.1</td>
<td>.15</td>
</tr>
<tr>
<td>4</td>
<td>-1.66</td>
<td>0.40</td>
<td>0.95</td>
<td>0.0</td>
<td>0.87</td>
<td>-0.2</td>
<td>.19</td>
</tr>
<tr>
<td>8</td>
<td>-1.66</td>
<td>0.40</td>
<td>1.00</td>
<td>0.1</td>
<td>1.06</td>
<td>0.3</td>
<td>.11</td>
</tr>
<tr>
<td>28</td>
<td>-2.03</td>
<td>0.47</td>
<td>1.02</td>
<td>0.2</td>
<td>0.97</td>
<td>0.1</td>
<td>.08</td>
</tr>
<tr>
<td>17</td>
<td>-2.11A</td>
<td>0.48</td>
<td>1.08</td>
<td>0.3</td>
<td>1.56</td>
<td>1.1</td>
<td>.03</td>
</tr>
<tr>
<td>2</td>
<td>-2.58</td>
<td>0.59</td>
<td>0.98</td>
<td>0.2</td>
<td>0.83</td>
<td>-0.1</td>
<td>.12</td>
</tr>
<tr>
<td>31</td>
<td>-2.58</td>
<td>0.59</td>
<td>0.99</td>
<td>0.2</td>
<td>1.40</td>
<td>0.8</td>
<td>.05</td>
</tr>
<tr>
<td>3</td>
<td>-3.00</td>
<td>0.72</td>
<td>1.01</td>
<td>0.3</td>
<td>1.84</td>
<td>1.1</td>
<td>.03</td>
</tr>
<tr>
<td>1</td>
<td>-1.92</td>
<td>1.83</td>
<td>Minimum Measure</td>
<td>Minimum measure</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
</tr>
<tr>
<td>7</td>
<td>-4.92</td>
<td>1.83</td>
<td>Minimum Measure</td>
<td>Minimum measure</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
</tr>
</tbody>
</table>

\textit{Note.} MNSQ = mean square; ZSTD = Z-score standardized; Pt-measure = point-measure. Minimum Measure means that all students answered the item correctly. All statistics are based on Rasch logits.

The unexplained variance explained by the first contrast was 2.6 eigenvalue units (5.2%), indicating acceptable unidimensionality.
The item-person map for Time 2 (Figure 14) shows how well the item difficulty estimates matched the student ability estimates. Overall, the range of item difficulty on this test covered the ability range of the students well, with three of the items at a higher level than the highest-performing student, and 18 items easier than the lowest-performing students. The mean of the person ability estimates was 0.86 logits ($SD = 0.77$), and the mean of the item difficulty estimates was -0.38 logits ($SD = 1.91$). There were again a number of low-frequency words that most students knew: *vocabulary* (5th 1,000-word list), *quiz* (6th 1,000-word list), and *palette* (6th 1,000-word list), all of which are commonly used in Japan as loanwords.

The Rasch item reliability estimate for this instrument was .92 (Very Good), and the Rasch item separation statistic was 3.28 (Good), indicating that the items were spread over three distinct levels of difficulty. The Rasch person reliability estimate was .56 (Poor), and the Rasch person separation statistic was 1.13 (Poor).

**Vocabulary Size Test (Time 3)**

The descriptive statistics of the Rasch analysis of the Time 3 vocabulary size test are displayed in Table 36. Of the ten items intended to be anchored to previous tests, four had difficulty measures that exceeded the 0.50 logit displacement criteria, so six items were retained as anchored items.

Using 0.70–1.30 as the acceptable range of item fit for low-stakes tests (Wright, et al., 1994), all of the items demonstrated acceptable fit to the Rasch model. The infit
Figure 14. Abridged Item-Person Map for the Vocabulary Size Test (Time 2)

<table>
<thead>
<tr>
<th>More able persons</th>
<th>More difficult items</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
</tr>
<tr>
<td></td>
<td># +</td>
</tr>
<tr>
<td></td>
<td># T</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>. +</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>-1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>+</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>-2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>+</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>-3</td>
<td></td>
</tr>
<tr>
<td></td>
<td>+</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Less able persons | Less difficult items

Note. Each “#” equals 2 students. Each “.” equals 1 student. M = Mean; S = One standard deviation from the mean; T = Two standard deviations from the mean.
Table 36. Rasch Descriptive Statistics for the Vocabulary Size Test Items (Time 3)

<table>
<thead>
<tr>
<th>Item</th>
<th>Measure</th>
<th>SE</th>
<th>Infit MNSQ</th>
<th>Infit ZSTD</th>
<th>Outfit MNSQ</th>
<th>Outfit ZSTD</th>
<th>Pt-measure correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>16</td>
<td>2.33</td>
<td>0.34</td>
<td>0.98</td>
<td>0.0</td>
<td>1.23</td>
<td>0.9</td>
<td>.08</td>
</tr>
<tr>
<td>25</td>
<td>1.76</td>
<td>0.28</td>
<td>0.98</td>
<td>-0.1</td>
<td>0.94</td>
<td>-0.3</td>
<td>.22</td>
</tr>
<tr>
<td>24</td>
<td>1.69A</td>
<td>0.28</td>
<td>1.03</td>
<td>0.3</td>
<td>1.23</td>
<td>1.3</td>
<td>.01</td>
</tr>
<tr>
<td>20</td>
<td>1.46</td>
<td>0.26</td>
<td>1.05</td>
<td>0.4</td>
<td>1.08</td>
<td>0.6</td>
<td>.03</td>
</tr>
<tr>
<td>31</td>
<td>1.45</td>
<td>0.26</td>
<td>1.02</td>
<td>0.2</td>
<td>0.99</td>
<td>0.0</td>
<td>.13</td>
</tr>
<tr>
<td>35</td>
<td>1.28A</td>
<td>0.25</td>
<td>0.96</td>
<td>-0.4</td>
<td>0.96</td>
<td>-0.3</td>
<td>.11</td>
</tr>
<tr>
<td>29</td>
<td>1.20</td>
<td>0.25</td>
<td>0.97</td>
<td>-0.3</td>
<td>1.02</td>
<td>0.2</td>
<td>.22</td>
</tr>
<tr>
<td>32</td>
<td>1.20</td>
<td>0.25</td>
<td>1.00</td>
<td>0.1</td>
<td>1.02</td>
<td>0.2</td>
<td>.15</td>
</tr>
<tr>
<td>26</td>
<td>1.14</td>
<td>0.25</td>
<td>1.01</td>
<td>0.2</td>
<td>1.02</td>
<td>0.3</td>
<td>.14</td>
</tr>
<tr>
<td>13</td>
<td>1.08</td>
<td>0.24</td>
<td>0.99</td>
<td>-0.1</td>
<td>0.97</td>
<td>-0.3</td>
<td>.21</td>
</tr>
<tr>
<td>14</td>
<td>1.02A</td>
<td>0.24</td>
<td>1.05</td>
<td>0.6</td>
<td>1.10</td>
<td>1.0</td>
<td>.08</td>
</tr>
<tr>
<td>34</td>
<td>1.02</td>
<td>0.24</td>
<td>0.95</td>
<td>-0.6</td>
<td>0.92</td>
<td>-0.8</td>
<td>.31</td>
</tr>
<tr>
<td>18</td>
<td>0.79</td>
<td>0.24</td>
<td>1.02</td>
<td>0.4</td>
<td>1.03</td>
<td>0.5</td>
<td>.13</td>
</tr>
<tr>
<td>6</td>
<td>0.77</td>
<td>0.24</td>
<td>1.03</td>
<td>0.6</td>
<td>1.05</td>
<td>0.8</td>
<td>.10</td>
</tr>
<tr>
<td>23</td>
<td>0.65</td>
<td>0.23</td>
<td>1.03</td>
<td>0.6</td>
<td>1.03</td>
<td>0.5</td>
<td>.11</td>
</tr>
<tr>
<td>27</td>
<td>0.63</td>
<td>0.23</td>
<td>0.99</td>
<td>-0.2</td>
<td>0.98</td>
<td>-0.3</td>
<td>.22</td>
</tr>
<tr>
<td>4</td>
<td>0.57</td>
<td>0.23</td>
<td>0.93</td>
<td>-1.6</td>
<td>0.92</td>
<td>-1.6</td>
<td>.37</td>
</tr>
<tr>
<td>33</td>
<td>0.52</td>
<td>0.23</td>
<td>1.08</td>
<td>1.8</td>
<td>1.08</td>
<td>1.5</td>
<td>.00</td>
</tr>
<tr>
<td>30</td>
<td>0.47A</td>
<td>0.23</td>
<td>0.98</td>
<td>-0.5</td>
<td>0.97</td>
<td>-0.7</td>
<td>.27</td>
</tr>
<tr>
<td>5</td>
<td>0.47</td>
<td>0.23</td>
<td>1.00</td>
<td>0.0</td>
<td>0.99</td>
<td>-0.2</td>
<td>.19</td>
</tr>
<tr>
<td>22</td>
<td>0.41</td>
<td>0.23</td>
<td>0.95</td>
<td>-1.3</td>
<td>0.94</td>
<td>-1.2</td>
<td>.32</td>
</tr>
<tr>
<td>10</td>
<td>0.39</td>
<td>0.23</td>
<td>1.01</td>
<td>0.3</td>
<td>1.00</td>
<td>0.1</td>
<td>.17</td>
</tr>
<tr>
<td>28</td>
<td>0.10A</td>
<td>0.23</td>
<td>0.90</td>
<td>-2.1</td>
<td>0.88</td>
<td>-2.1</td>
<td>.39</td>
</tr>
<tr>
<td>9</td>
<td>0.09</td>
<td>0.23</td>
<td>0.96</td>
<td>-0.8</td>
<td>0.95</td>
<td>-0.9</td>
<td>.30</td>
</tr>
<tr>
<td>11</td>
<td>0.09</td>
<td>0.23</td>
<td>1.01</td>
<td>0.2</td>
<td>1.00</td>
<td>0.1</td>
<td>.17</td>
</tr>
<tr>
<td>19</td>
<td>0.09</td>
<td>0.23</td>
<td>1.01</td>
<td>0.1</td>
<td>1.01</td>
<td>0.2</td>
<td>.18</td>
</tr>
<tr>
<td>21</td>
<td>-0.07</td>
<td>0.24</td>
<td>1.04</td>
<td>0.6</td>
<td>1.06</td>
<td>0.9</td>
<td>.08</td>
</tr>
<tr>
<td>8</td>
<td>-0.24</td>
<td>0.24</td>
<td>1.05</td>
<td>0.6</td>
<td>1.07</td>
<td>0.8</td>
<td>.07</td>
</tr>
<tr>
<td>12</td>
<td>-0.45A</td>
<td>0.25</td>
<td>1.15</td>
<td>1.5</td>
<td>1.15</td>
<td>1.4</td>
<td>.08</td>
</tr>
<tr>
<td>15</td>
<td>-0.68</td>
<td>0.26</td>
<td>0.90</td>
<td>-0.8</td>
<td>0.87</td>
<td>-0.9</td>
<td>.43</td>
</tr>
<tr>
<td>1</td>
<td>-1.04</td>
<td>0.28</td>
<td>0.97</td>
<td>-0.1</td>
<td>0.99</td>
<td>0.0</td>
<td>.22</td>
</tr>
<tr>
<td>3</td>
<td>-2.40</td>
<td>0.47</td>
<td>0.98</td>
<td>0.1</td>
<td>0.83</td>
<td>-0.3</td>
<td>.22</td>
</tr>
<tr>
<td>17</td>
<td>-2.94</td>
<td>0.59</td>
<td>1.02</td>
<td>0.2</td>
<td>1.18</td>
<td>0.5</td>
<td>-.02</td>
</tr>
<tr>
<td>2</td>
<td>-4.07</td>
<td>1.01</td>
<td>1.01</td>
<td>0.3</td>
<td>1.04</td>
<td>0.4</td>
<td>.01</td>
</tr>
<tr>
<td>7</td>
<td>-4.07</td>
<td>1.01</td>
<td>1.00</td>
<td>0.3</td>
<td>0.90</td>
<td>0.3</td>
<td>.06</td>
</tr>
</tbody>
</table>

Note. MNSQ = mean square; ZSTD = Z-score standardized; Pt-measure = point-measure. All statistics are based on Rasch logits.

MNSQ statistic ranged from 0.90–1.15, and the point-measure correlations ranged from -.02 to .43.
Regarding dimensionality, the variance explained by the items was 19.0%, and the variance explained by measures was 23.0%. The unexplained variance explained by the first contrast was 3.0 eigenvalue units (6.5%), indicating good unidimensionality. The item-person map for Time 3 (Figure 15) shows how well the difficulty of the items matches the ability of the students. Overall, the range of item difficulty on this test covered the students’ ability range well, with ten items at a higher level than the highest-performing student, and eight items easier than the lowest-performing students. The cluster of students and cluster of items around the mean reveals a higher degree of uniformity in which items the students knew, relative to the previous two tests. The mean of the person ability estimates was 0.36 logits ($SD = .40$), and the mean of the item difficulty estimates was 0.19 logits ($SD = 1.48$). There were no low-frequency words below the mean item difficulty as with the first two tests; overall, item difficulty corresponded to frequency level as expected.

The Rasch item reliability estimate for this instrument was .94 (Very Good), and the Rasch item separation statistic was 3.96 (Good), indicating that the items were spread over four distinct levels of difficulty. The Rasch person reliability estimate was .04 (Poor), and the Rasch person separation statistic was .20 (Poor), indicating that student ability was more concentrated around the mean than in the previous administrations.
Figure 15. Abridged Item-Person Map for the Vocabulary Size Test (Time 3)

<table>
<thead>
<tr>
<th>More able persons</th>
<th>More difficult items</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>5000-2</td>
</tr>
<tr>
<td>2</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td>S 6000-1 7000-1</td>
</tr>
<tr>
<td></td>
<td>. 4000-2 5000-5</td>
</tr>
<tr>
<td></td>
<td>7000-5</td>
</tr>
<tr>
<td></td>
<td>. T 4000-1 4000-5 6000-2 7000-3</td>
</tr>
<tr>
<td>1</td>
<td>.# + 3000-3 7000-4</td>
</tr>
<tr>
<td></td>
<td># 5000-3</td>
</tr>
<tr>
<td></td>
<td>### S 2000-3 5000-4</td>
</tr>
<tr>
<td></td>
<td>######</td>
</tr>
<tr>
<td></td>
<td>#######</td>
</tr>
<tr>
<td></td>
<td>######## M 2000-5 3000-5 5000-1 6000-3</td>
</tr>
<tr>
<td>0</td>
<td>.### S+ 4000-3 2000-4</td>
</tr>
<tr>
<td></td>
<td>T 3000-2</td>
</tr>
<tr>
<td></td>
<td>. 3000-4</td>
</tr>
<tr>
<td></td>
<td>.</td>
</tr>
<tr>
<td>-1</td>
<td>. + 1000-1</td>
</tr>
<tr>
<td></td>
<td>S</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>-2</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td>1000-3</td>
</tr>
<tr>
<td></td>
<td>T</td>
</tr>
<tr>
<td>-3</td>
<td>+ 1000-5</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>-4</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td>1000-2 1000-4</td>
</tr>
<tr>
<td>Less able persons</td>
<td>Less difficult items</td>
</tr>
</tbody>
</table>

Note. Each “#” equals 2 students. Each “.” equals 1 student. M = Mean; S = One standard deviation from the mean; T = Two standard deviations from the mean.
Vocabulary Size Test (Time 4)

The descriptive statistics of the Rasch analysis of the Time 4 vocabulary size test are displayed in Table 37. Of the 35 test items, two items had to be eliminated. In item 19, there was an error in the correct answer choice that made the item difficult to answer correctly. In item 27, a printing error caused two of the four answer choices (including the correct answer) not to be printed. Therefore, only 33 of the 35 items were analyzed. Of the ten items intended to be anchored to previous tests, one was item 19, and four of the remaining nine had difficulty measures that exceeded the 0.50 logit displacement criteria, so five items were retained as anchored items.

Using 0.70–1.30 as the acceptable range of item fit for low-stakes tests (Wright, Linacre, Gustafson, & Martin-Lof, 1994), all of the items demonstrated acceptable fit to the Rasch model. The infit MNSQ statistic ranged from 0.84–1.11, and the point-measure correlations ranged from -.06 to .43.

Regarding dimensionality, the variance explained by the items was 29.9%, which was more than 4 times the variance accounted for by the first contrast (5.4%), indicating good unidimensionality. The variance explained by measures was 36.2%. The unexplained variance explained by the first contrast was 2.7 eigenvalue units (5.4%), indicating acceptable unidimensionality.

The item-person map for Time 4 (Figure 16) shows how well the difficulty of the items matches the ability of the students. Overall, the range of item difficulty on this test covered the ability range of the students well, with eleven of the items at a higher level than the highest-performing student, and six items easier than the lowest-
Table 37. Rasch Descriptive Statistics for the Vocabulary Size Test Items (Time 4)

<table>
<thead>
<tr>
<th>Item</th>
<th>Measure</th>
<th>SE</th>
<th>Infit MNSQ</th>
<th>Infit ZSTD</th>
<th>Outfit MNSQ</th>
<th>Outfit ZSTD</th>
<th>Pt-measure correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>16</td>
<td></td>
<td>2.33A</td>
<td>0.35</td>
<td>1.07</td>
<td>0.4</td>
<td>1.09</td>
<td>0.4</td>
</tr>
<tr>
<td>19</td>
<td></td>
<td>2.27</td>
<td>0.34</td>
<td>1.03</td>
<td>0.2</td>
<td>1.10</td>
<td>0.4</td>
</tr>
<tr>
<td>33</td>
<td></td>
<td>2.15</td>
<td>0.33</td>
<td>1.01</td>
<td>0.1</td>
<td>1.03</td>
<td>0.2</td>
</tr>
<tr>
<td>32</td>
<td></td>
<td>1.95</td>
<td>0.31</td>
<td>1.01</td>
<td>0.1</td>
<td>0.99</td>
<td>0.0</td>
</tr>
<tr>
<td>24</td>
<td></td>
<td>1.86</td>
<td>0.30</td>
<td>1.06</td>
<td>0.4</td>
<td>1.21</td>
<td>0.9</td>
</tr>
<tr>
<td>13</td>
<td></td>
<td>1.77</td>
<td>0.29</td>
<td>0.95</td>
<td>-0.2</td>
<td>0.93</td>
<td>-0.3</td>
</tr>
<tr>
<td>9</td>
<td></td>
<td>1.68</td>
<td>0.29</td>
<td>0.89</td>
<td>-0.7</td>
<td>0.77</td>
<td>-1.1</td>
</tr>
<tr>
<td>26</td>
<td></td>
<td>1.60</td>
<td>0.28</td>
<td>1.11</td>
<td>0.8</td>
<td>1.38</td>
<td>1.8</td>
</tr>
<tr>
<td>28</td>
<td></td>
<td>1.60</td>
<td>0.28</td>
<td>1.00</td>
<td>0.1</td>
<td>0.95</td>
<td>-0.2</td>
</tr>
<tr>
<td>23</td>
<td></td>
<td>1.45</td>
<td>0.27</td>
<td>1.02</td>
<td>0.2</td>
<td>0.99</td>
<td>0.0</td>
</tr>
<tr>
<td>15</td>
<td></td>
<td>1.38</td>
<td>0.27</td>
<td>1.02</td>
<td>0.2</td>
<td>1.13</td>
<td>0.8</td>
</tr>
<tr>
<td>29</td>
<td></td>
<td>1.24</td>
<td>0.26</td>
<td>1.06</td>
<td>0.5</td>
<td>1.07</td>
<td>0.5</td>
</tr>
<tr>
<td>30</td>
<td></td>
<td>0.87</td>
<td>0.24</td>
<td>1.06</td>
<td>0.8</td>
<td>1.09</td>
<td>0.9</td>
</tr>
<tr>
<td>31</td>
<td></td>
<td>0.87</td>
<td>0.24</td>
<td>1.06</td>
<td>0.8</td>
<td>1.07</td>
<td>0.7</td>
</tr>
<tr>
<td>20</td>
<td></td>
<td>0.81</td>
<td>0.24</td>
<td>0.97</td>
<td>-0.3</td>
<td>0.99</td>
<td>-0.1</td>
</tr>
<tr>
<td>18</td>
<td></td>
<td>0.68A</td>
<td>0.24</td>
<td>0.94</td>
<td>-0.8</td>
<td>0.94</td>
<td>-0.7</td>
</tr>
<tr>
<td>5</td>
<td></td>
<td>0.47A</td>
<td>0.24</td>
<td>0.95</td>
<td>-0.9</td>
<td>0.93</td>
<td>-0.9</td>
</tr>
<tr>
<td>11</td>
<td></td>
<td>0.35</td>
<td>0.23</td>
<td>0.98</td>
<td>-0.3</td>
<td>0.97</td>
<td>-0.4</td>
</tr>
<tr>
<td>25</td>
<td></td>
<td>0.24</td>
<td>0.23</td>
<td>0.96</td>
<td>-0.7</td>
<td>0.97</td>
<td>-0.4</td>
</tr>
<tr>
<td>8</td>
<td></td>
<td>0.02</td>
<td>0.24</td>
<td>0.96</td>
<td>-0.7</td>
<td>0.95</td>
<td>-0.7</td>
</tr>
<tr>
<td>12</td>
<td></td>
<td>-0.20</td>
<td>0.24</td>
<td>1.02</td>
<td>0.3</td>
<td>1.02</td>
<td>0.3</td>
</tr>
<tr>
<td>1</td>
<td></td>
<td>-0.63</td>
<td>0.25</td>
<td>1.05</td>
<td>0.5</td>
<td>1.05</td>
<td>0.4</td>
</tr>
<tr>
<td>21</td>
<td></td>
<td>-0.97</td>
<td>0.27</td>
<td>0.96</td>
<td>-0.2</td>
<td>0.89</td>
<td>-0.6</td>
</tr>
<tr>
<td>27</td>
<td></td>
<td>-0.97</td>
<td>0.27</td>
<td>0.96</td>
<td>-0.2</td>
<td>0.98</td>
<td>-0.1</td>
</tr>
<tr>
<td>6</td>
<td></td>
<td>-1.14A</td>
<td>.28</td>
<td>.84</td>
<td>-1.0</td>
<td>.77</td>
<td>-1.2</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>-1.20</td>
<td>.29</td>
<td>1.00</td>
<td>.1</td>
<td>1.01</td>
<td>.1</td>
</tr>
<tr>
<td>10</td>
<td></td>
<td>-1.38</td>
<td>.30</td>
<td>1.02</td>
<td>.2</td>
<td>1.07</td>
<td>.4</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>-1.66A</td>
<td>.33</td>
<td>1.11</td>
<td>.6</td>
<td>1.24</td>
<td>.9</td>
</tr>
<tr>
<td>22</td>
<td></td>
<td>-1.92</td>
<td>.36</td>
<td>.96</td>
<td>.0</td>
<td>.90</td>
<td>-.2</td>
</tr>
<tr>
<td>7</td>
<td></td>
<td>-3.13</td>
<td>.59</td>
<td>.99</td>
<td>.2</td>
<td>.73</td>
<td>-.2</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>-3.56</td>
<td>.72</td>
<td>1.00</td>
<td>.2</td>
<td>.71</td>
<td>-.1</td>
</tr>
<tr>
<td>14</td>
<td></td>
<td>-3.56</td>
<td>.72</td>
<td>1.02</td>
<td>.3</td>
<td>.99</td>
<td>.3</td>
</tr>
<tr>
<td>17</td>
<td></td>
<td>-5.49</td>
<td>1.83</td>
<td>Minimum measure</td>
<td>.00</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. MNSQ = mean square; ZSTD = Z-score standardized; Pt-measure = point-measure. Minimum Measure means that all students answered the item correctly. All statistics are based on Rasch logits.

performing students. The mean of the person ability estimates was 0.24 logits ($SD = .56$), and the mean of the item difficulty estimates was -0.01 logits ($SD = 1.91$).

Frequency level and item difficulty seemed to correlate well, although there was one
Figure 16. Abridged Item-Person Map for the Vocabulary Size Test (Time 4)

<table>
<thead>
<tr>
<th>More able persons</th>
<th>More difficult items</th>
</tr>
</thead>
<tbody>
<tr>
<td>5000-2</td>
<td></td>
</tr>
<tr>
<td>4000-2</td>
<td></td>
</tr>
<tr>
<td>7000-5</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>7000-4</td>
</tr>
<tr>
<td>5000-1</td>
<td>6000-3   6000-4</td>
</tr>
<tr>
<td>4000-1</td>
<td></td>
</tr>
<tr>
<td>6000-1</td>
<td></td>
</tr>
<tr>
<td>T</td>
<td>3000-4</td>
</tr>
<tr>
<td>XX</td>
<td>5000-5</td>
</tr>
<tr>
<td>XXXXX</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>6000-5   7000-3</td>
</tr>
<tr>
<td>XXXXXXXX</td>
<td>4000-3</td>
</tr>
<tr>
<td>S</td>
<td>5000-3</td>
</tr>
<tr>
<td>XXXXXXXXXX</td>
<td>2000-2</td>
</tr>
<tr>
<td>XXXXXXX</td>
<td>2000-5</td>
</tr>
<tr>
<td>M</td>
<td>6000-2</td>
</tr>
<tr>
<td>XXXXXXXXXX</td>
<td>M</td>
</tr>
<tr>
<td>0</td>
<td>2000-4</td>
</tr>
<tr>
<td>XXXXXXX</td>
<td>3000-2</td>
</tr>
<tr>
<td>S</td>
<td></td>
</tr>
<tr>
<td>XX</td>
<td>1000-1</td>
</tr>
<tr>
<td>XXXXX</td>
<td></td>
</tr>
<tr>
<td>XX T</td>
<td></td>
</tr>
<tr>
<td>-1</td>
<td>4000-4   4000-5</td>
</tr>
<tr>
<td>2000-3</td>
<td></td>
</tr>
<tr>
<td>1000-3</td>
<td></td>
</tr>
<tr>
<td>X</td>
<td>3000-1</td>
</tr>
<tr>
<td>S</td>
<td>2000-1</td>
</tr>
<tr>
<td></td>
<td>5000-4</td>
</tr>
<tr>
<td>-2</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>-3</td>
<td></td>
</tr>
<tr>
<td>T 1000-4</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>Less able persons</td>
<td>Less difficult items</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>1000-2</td>
<td>3000-3</td>
</tr>
</tbody>
</table>

Note. Each "X" equals 1 student. M = Mean; S = One standard deviation from the mean; T = Two standard deviations from the mean.
word from the 5th 1,000-word level, *scrub*, that was easy for most of the students. This was a repeated item (intended anchor) from Time 2, but the difficulty estimate of this item was much lower than at Time 2. That administration was approximately five months prior, so a strong learning effect for this specific item seems unlikely.

The Rasch item reliability estimate for this instrument was .94 (Very Good), and the Rasch item separation statistic was 4.00 (Very Good), indicating that the items were spread over four distinct levels of difficulty. The Rasch person reliability estimate was .32 (Poor), and the Rasch person separation statistic was .69 (Poor), indicating that student ability range was not wide.

**Summary of Instrument Validation**

In this chapter, Rasch analysis was used to generate and examine evidence of the validity and reliability of the measurements used to investigate the hypotheses presented in this study. The raters and rating scale used for the oral reading prosody measurement were assessed by conducting multifaceted Rasch analyses using FACETS (Linacre, 2011a), and the reading comprehension tests and vocabulary size tests were assessed for the degree of fit to the Rasch model by using Winsteps (Linacre, 2011b).

For the rating scale, which was used for eight oral prosody measurements (four simplified and four authentic), Wright maps were generated in order to visualize the interplay of student performance (ability), rater strictness, and criterion difficulty on a single scale and with a single unit of measurement (Rasch logits). Analysis of these maps indicated a good spread of student ability across the rating scale (indicating full
use of the scale by the raters) in all cases, and strong agreement among raters in their application of the scale in seven out of the eight sets of ratings. Next, infit and outfit statistics, along with person and item separation and reliability statistics, were generated and analyzed for both the raters and the rating scale criteria. The infit and outfit statistics generally provided good evidence of intra-rater reliability (consistency) while revealing a few cases of inconsistency. The person and item measurement reliability of the raters was Fair or Good in seven out of the eight analyses, and Poor in one of the eight analyses (Time 1, authentic text), indicating good inter-rater reliability using Fisher’s (2007) criteria. The person and item separation statistic was Fair or Good in five out of the eight analyses, and Poor in three of the eight analyses, indicating difficulty in distinguishing between the levels of severity of the raters in some cases. For the rating scale criteria, both the separation and reliability statistics were nearly always in the Excellent range. Finally, Rasch-Andrich thresholds were reported as an indication whether the four levels of the rating scale were clearly distinguished. In all cases, the thresholds met the minimum separation criteria suggested by Wolfe and Smith (2007).

For the reading comprehension and vocabulary size tests, measures of item misfit (infit and outfit) were first reported in order to identify unexpected responses. For the reading comprehension test, only two items out of a total of 96 (24 times four administrations) fell outside the acceptable range. For the vocabulary size tests, all 138 items (35 times four administrations, minus two discarded items) were within the acceptable range.
Next, Rasch principal components analyses were conducted and reported as a measure of unidimensionality (the notion that the test was measuring a single construct—in this case, reading comprehension or vocabulary knowledge). These analyses indicated good likelihood of unidimensionality in all cases.

Next, Wright item-person maps were generated and analyzed to assess how well the test items matched the participants’ abilities. For the reading comprehension tests, these maps revealed a ceiling effect in most cases, meaning that there were not enough difficult items to separate the higher-ability students well. These maps also allowed for a detailed analysis of the influence of text difficulty and item type on item difficulty. For the vocabulary size tests, the maps revealed that the range of item difficulty matched the range of student ability well, and detailed analyses were conducted for instances where most students answered correctly for low-frequency words.

Finally, item and person reliability and separation statistics were reported and interpreted. For both the reading comprehension tests and the vocabulary size tests, the item reliability and separation statistics indicated that the items were spread over four difficulty levels in nearly all cases. The person reliability levels were usually rather low, indicating that many participants performed so similarly that they could not be distinguished statistically.

In addition to their use in analyzing the validity and reliability of the measurements used in this study, the Rasch interval measures (logits) were used instead of raw scores in the statistical analyses reported in the next chapter.
CHAPTER 5

RESULTS

Preliminary Results

The general English proficiency of the students in this study was measured via the CASEC test at the beginning of the school year, which corresponds approximately to Time 1 of this study. As reported in Table 38, the mean CASEC score was 581.33 (SD = 66.56), which corresponds to a TOEIC reference score of 523.97 (SD = 102.90) or a TOEFL-PBT score of 456.83 (SD = 26.08). This mean score falls in CASEC Level C (450-599) on the CASEC Proficiency Scale, which ranges from E (0-389) to AA (880-1000). According to the CASEC website (global.casec.com), those with Level C proficiency can generally communicate their intended meaning in daily life, and can comprehend the overall gist when listening to a native speaker speak.

<table>
<thead>
<tr>
<th></th>
<th>IR-before-ERFT group</th>
<th>ERFT-before-IR group</th>
<th>All participants</th>
</tr>
</thead>
<tbody>
<tr>
<td>M</td>
<td>609.68</td>
<td>552.54</td>
<td>581.33</td>
</tr>
<tr>
<td>SE</td>
<td>11.07</td>
<td>8.37</td>
<td>7.54</td>
</tr>
<tr>
<td>95% CI</td>
<td>[587.25, 632.12]</td>
<td>[535.59, 569.48]</td>
<td>[566.33, 596.34]</td>
</tr>
<tr>
<td>SD</td>
<td>68.25</td>
<td>52.27</td>
<td>66.56</td>
</tr>
<tr>
<td>Skewness</td>
<td>-0.14</td>
<td>-0.43</td>
<td>0.09</td>
</tr>
<tr>
<td>SES</td>
<td>0.38</td>
<td>0.38</td>
<td>0.27</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>-0.60</td>
<td>-0.35</td>
<td>-0.26</td>
</tr>
<tr>
<td>SEK</td>
<td>0.75</td>
<td>0.74</td>
<td>0.54</td>
</tr>
</tbody>
</table>

Note. All students read the same amount in the IR class, so the amount reported here is not a mean. It is the estimated number of standard words in the texts they read, and is thus the same for all students. CI = confidence interval; SES = standard error of skewness; SEK = standard error of kurtosis.
Table 39 shows the amount of text read by the students in the extensive reading plus fluency training (ERFT) and intensive reading (IR) semesters, respectively. In keeping with the definition and goals of extensive reading and intensive reading, students read the planned 20,258.32 standard words (estimated) (See Chapter 3 for a detailed explanation of how this amount was estimated) during the intensive reading semester, and the mean amount read during the extensive reading with fluency training semester was 113,444.73 (SD = 10,457.26) standard words (using exact counts).

<table>
<thead>
<tr>
<th>Table 39. Amount Read (Standard Words)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ERFT</td>
</tr>
<tr>
<td>M</td>
</tr>
<tr>
<td>SE</td>
</tr>
<tr>
<td>95% CI</td>
</tr>
<tr>
<td>SD</td>
</tr>
<tr>
<td>Skewness</td>
</tr>
<tr>
<td>SES</td>
</tr>
<tr>
<td>Kurtosis</td>
</tr>
<tr>
<td>SEK</td>
</tr>
</tbody>
</table>

Note. *All students read the same amount in the IR class, so the amount reported here is not a mean. It is the estimated number of standard words in the texts they read, and is thus the same for all students. ERFT = extensive reading with reading fluency group; IR = intensive reading group; CI = confidence interval; SES = standard error of skewness; SEK = standard error of kurtosis.

Table 40 reports the amount of time spent reading under the two conditions. The exact time spent reading during the extensive reading with fluency training semester, as recorded by Xreading.com, was a mean of 23.86 (SD = 8.03) hours (2.39 hours/week). Mean time spent reading during the intensive reading semester, as self-reported by the students, was 21.43 (SD = 6.96) hours (2.14 hours/week).
Table 40. Time Spent Reading (Hours)

<table>
<thead>
<tr>
<th></th>
<th>ERFT</th>
<th>IR</th>
</tr>
</thead>
<tbody>
<tr>
<td>M</td>
<td>23.86</td>
<td>21.43</td>
</tr>
<tr>
<td>SE</td>
<td>0.91</td>
<td>0.79</td>
</tr>
<tr>
<td>95% CI</td>
<td>[22.05, 25.67]</td>
<td>[19.86, 23.00]</td>
</tr>
<tr>
<td>SD</td>
<td>8.03</td>
<td>6.96</td>
</tr>
<tr>
<td>Skewness</td>
<td>1.05</td>
<td>0.21</td>
</tr>
<tr>
<td>SES</td>
<td>0.27</td>
<td>0.27</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>1.73</td>
<td>-0.42</td>
</tr>
<tr>
<td>SEK</td>
<td>0.54</td>
<td>0.54</td>
</tr>
</tbody>
</table>

Note. ERFT = extensive reading with reading fluency group; IR = intensive reading group; CI = confidence interval; SES = standard error of skewness; SEK = standard error of kurtosis.

Hypothesis 1: Reading Rate and Fluency Gains

Hypothesis 1 stated that compared with intensive reading, extensive reading with fluency training would yield significantly greater gains in silent reading rate, oral reading rate, and oral reading fluency. To test this hypothesis, repeated-measures ANOVAs with two levels (extensive reading with fluency training and intensive reading) and three dependent variables (silent reading rate change, oral reading rate change, and oral reading fluency change) were conducted. Changes were calculated by subtracting pretest from posttest results. ANOVAs were conducted separately for the IR-before-ERFT students and the ERFT-before-IR students, and for simplified and authentic texts.

Descriptive statistics for the IR-before-ERFT ANOVA for simplified texts are reported in Table 41.

With the Bonferroni adjustment required due to three dependent variables, the $p$-level required for significance was set at .017. Sphericity is not an issue in a repeated-measures ANOVA with only two treatment levels (Field, 2009). The effect of treatment
Table 41. Descriptive Statistics for Change in Silent Reading Rate, Oral Reading Rate, and Oral Reading Fluency During ERFT and IR Treatments, for IR-Before-ERFT Group, for Simplified Texts (N = 38)

<table>
<thead>
<tr>
<th></th>
<th>ERFT-SRR (change in swpm)</th>
<th>IR-SRR (change in swpm)</th>
<th>ERFT-ORR (change in wcpm)</th>
<th>IR-ORR (change in wcpm)</th>
<th>ERFT-ORF (change in logits)</th>
<th>IR-ORF (change in logits)</th>
</tr>
</thead>
<tbody>
<tr>
<td>M</td>
<td>2.99</td>
<td>2.69</td>
<td>*9.79</td>
<td>*3.05</td>
<td>0.27</td>
<td>-1.36</td>
</tr>
<tr>
<td>SE</td>
<td>2.70</td>
<td>2.71</td>
<td>1.29</td>
<td>1.26</td>
<td>0.54</td>
<td>0.55</td>
</tr>
<tr>
<td>95% CI</td>
<td>[-2.47, 8.45]</td>
<td>[-2.81, 8.19]</td>
<td>[7.18, 12.39]</td>
<td>[.50, 5.59]</td>
<td>[-.83, 1.36]</td>
<td>[-2.48, -.25]</td>
</tr>
<tr>
<td>SD</td>
<td>16.62</td>
<td>16.73</td>
<td>7.93</td>
<td>7.74</td>
<td>3.33</td>
<td>3.40</td>
</tr>
<tr>
<td>Skewness</td>
<td>0.35</td>
<td>1.18</td>
<td>-0.88</td>
<td>0.14</td>
<td>-0.52</td>
<td>0.13</td>
</tr>
<tr>
<td>SES</td>
<td>0.38</td>
<td>0.38</td>
<td>0.38</td>
<td>0.38</td>
<td>0.38</td>
<td>0.38</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>0.06</td>
<td>1.50</td>
<td>1.43</td>
<td>-0.02</td>
<td>0.27</td>
<td>0.44</td>
</tr>
<tr>
<td>SEK</td>
<td>0.75</td>
<td>0.75</td>
<td>0.75</td>
<td>0.75</td>
<td>0.75</td>
<td>0.75</td>
</tr>
</tbody>
</table>

Note. * indicates significant treatment effect; ERFT = extensive reading with reading fluency; IR = intensive reading; CI = confidence interval; SES = standard error of skewness; SEK = standard error of kurtosis.

on silent reading rate was not significant, $F(3, 35) = 5.75, p = .92$, partial $\eta^2 = .00$. The effect of treatment on oral reading rate was significant, $F(3, 35) = 5.75, p < .001$, partial $\eta^2 = .31$. The effect of treatment on oral reading fluency was not significant, $F(3, 35) = 5.75, p = .049$, partial $\eta^2 = .10$. This result indicated that for students who did intensive reading first and extensive reading with fluency training second, the extensive reading with fluency training approach yielded significantly greater gains in oral reading rate on simplified texts over one semester when compared to the intensive reading approach.

Descriptive statistics for the ERFT-before-IR ANOVA for simplified texts are reported in Table 42.

With the Bonferroni adjustment required due to three dependent variables, the $p$-level required for significance was set at .017. Sphericity is not an issue in a repeated-measures ANOVA with only two treatment levels (Field, 2009). The effect of treatment on silent reading rate was significant, $F(3, 36) = 31.22, p < .001$, partial $\eta^2 = .68$. The
Table 42. Descriptive Statistics for Change in Silent Reading Rate, Oral Reading Rate, and Oral Reading Fluency During ERFT and IR Treatments, for ERFT-Before-IR Group, for Simplified Texts (N = 39)

<table>
<thead>
<tr>
<th></th>
<th>ERFT-SRR (change in swpm)</th>
<th>IR-SRR (change in swpm)</th>
<th>ERFT-ORR (change in wcpm)</th>
<th>IR-ORR (change in wcpm)</th>
<th>ERFT-ORF (change in logits)</th>
<th>IR-ORF (change in logits)</th>
</tr>
</thead>
<tbody>
<tr>
<td>M</td>
<td>*19.22</td>
<td>*-2.50</td>
<td>*12.55</td>
<td>*4.81</td>
<td>2.04</td>
<td>0.45</td>
</tr>
<tr>
<td>SE</td>
<td>1.91</td>
<td>1.82</td>
<td>1.45</td>
<td>1.37</td>
<td>0.63</td>
<td>0.73</td>
</tr>
<tr>
<td>95% CI</td>
<td>[15.26,23.18]</td>
<td>[-6.28,1.28]</td>
<td>[9.53,15.57]</td>
<td>[1.95,7.66]</td>
<td>[.74,3.34]</td>
<td>[-1.08,1.97]</td>
</tr>
<tr>
<td>SD</td>
<td>12.21</td>
<td>11.51</td>
<td>9.19</td>
<td>8.68</td>
<td>3.96</td>
<td>4.63</td>
</tr>
<tr>
<td>Skewness</td>
<td>1.65</td>
<td>0.39</td>
<td>0.61</td>
<td>0.99</td>
<td>0.48</td>
<td>0.47</td>
</tr>
<tr>
<td>SES</td>
<td>0.38</td>
<td>0.38</td>
<td>0.38</td>
<td>0.38</td>
<td>0.38</td>
<td>0.38</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>5.78</td>
<td>-0.01</td>
<td>0.21</td>
<td>1.82</td>
<td>2.13</td>
<td>1.07</td>
</tr>
<tr>
<td>SEK</td>
<td>0.74</td>
<td>0.74</td>
<td>0.74</td>
<td>0.74</td>
<td>0.74</td>
<td>0.74</td>
</tr>
</tbody>
</table>

Note. * indicates significant treatment effect; ERFT = extensive reading with reading fluency; IR = intensive reading; CI = confidence interval; SES = standard error of skewness; SEK = standard error of kurtosis.

The effect of treatment on oral reading rate was significant, \( F(3, 36) = 31.22, p < .001 \), partial \( \eta^2 = .31 \). The effect of treatment on oral reading fluency was not significant, \( F(3, 36) = 31.22, p = .069 \), partial \( \eta^2 = .08 \). This result indicated that for students who took extensive reading with fluency training first and intensive reading second, the extensive reading with fluency training approach yielded significantly greater gains in silent and oral reading rate on simplified texts over one semester when compared to the intensive reading approach.

Descriptive statistics for the IR-before-ERFT ANOVA for authentic texts are reported in Table 43.

With the Bonferroni adjustment required due to three dependent variables, the \( p \)-level required for significance was set at .017. Sphericity is not an issue in a repeated-measures ANOVA with only two treatment levels (Field, 2009). The effect of treatment on silent reading rate was significant, \( F(1, 37) = 17.87, p < .001 \), partial \( \eta^2 = .33 \). The
Table 43. Descriptive Statistics for Change in Silent Reading Rate, Oral Reading Rate, and Oral Reading Fluency During ERFT and IR Treatments, for IR-Before-ERFT Group, for Authentic Texts (N = 38)

<table>
<thead>
<tr>
<th></th>
<th>ERFT-SRR (change in swpm)</th>
<th>IR-SRR (change in swpm)</th>
<th>ERFT-ORR (change in wcpm)</th>
<th>IR-ORR (change in wcpm)</th>
<th>ERFT-ORF (change in logits)</th>
<th>IR-ORF (change in logits)</th>
</tr>
</thead>
<tbody>
<tr>
<td>M</td>
<td>-2.86</td>
<td>8.26</td>
<td>*15.18</td>
<td>*4.37</td>
<td>1.84</td>
<td>1.73</td>
</tr>
<tr>
<td>SE</td>
<td>2.25</td>
<td>2.34</td>
<td>1.22</td>
<td>1.49</td>
<td>0.86</td>
<td>0.91</td>
</tr>
<tr>
<td>95% CI</td>
<td>[-7.41, 1.70]</td>
<td>[3.51, 13.00]</td>
<td>[12.72, 17.64]</td>
<td>[1.35, 7.40]</td>
<td>[0.10, 3.59]</td>
<td>[-0.12, 3.58]</td>
</tr>
<tr>
<td>SD</td>
<td>13.85</td>
<td>14.43</td>
<td>7.49</td>
<td>9.20</td>
<td>5.32</td>
<td>5.63</td>
</tr>
<tr>
<td>Skewness</td>
<td>-0.39</td>
<td>-0.19</td>
<td>0.61</td>
<td>0.16</td>
<td>-0.81</td>
<td>0.62</td>
</tr>
<tr>
<td>SES</td>
<td>0.38</td>
<td>0.38</td>
<td>0.38</td>
<td>0.38</td>
<td>0.38</td>
<td>0.38</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>-0.30</td>
<td>1.37</td>
<td>-0.42</td>
<td>-0.08</td>
<td>2.23</td>
<td>0.61</td>
</tr>
<tr>
<td>SEK</td>
<td>0.75</td>
<td>0.75</td>
<td>0.75</td>
<td>0.75</td>
<td>0.75</td>
<td>0.75</td>
</tr>
</tbody>
</table>

Note. * indicates significant treatment effect; ERFT = extensive reading with reading fluency; IR = intensive reading; CI = confidence interval; SES = standard error of skewness; SEK = standard error of kurtosis.

effect of treatment on oral reading rate was significant, $F(1, 37) = 27.43, p < .001$, partial $\eta^2 = .43$. The effect of treatment on oral reading fluency was not significant, $F(1, 37) = .01, p = .920$, partial $\eta^2 = .00$. This indicates that for students who took intensive reading first and extensive reading with fluency training second, the intensive reading approach yielded a significantly greater gain in silent reading rate on authentic texts over one semester when compared to the extensive reading with fluency training approach. However, the extensive reading with fluency training approach yielded a significantly greater gain in oral reading rate on authentic texts over one semester when compared to the intensive reading approach.

Descriptive statistics for the ERFT-before-IR ANOVA for authentic texts are reported in Table 44.

With the Bonferroni adjustment required due to three dependent variables, the $p$-level required for significance was set at .017. Sphericity is not an issue in a repeated-measures ANOVA with only two treatment levels (Field, 2009). The effect of treatment
Table 44. Descriptive Statistics for Change in Silent Reading Rate, Oral Reading Rate, and Oral Reading Fluency During ERFT and IR Treatments, for the ERFT-Before-IR Group, for Authentic Texts ($N = 39$)

<table>
<thead>
<tr>
<th></th>
<th>ERFT-SRR (change in swpm)</th>
<th>IR-SRR (change in swpm)</th>
<th>ERFT-ORR (change in wcpm)</th>
<th>IR-ORR (change in wcpm)</th>
<th>ERFT-ORF (change in logits)</th>
<th>IR-ORF (change in logits)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$M$</td>
<td>*21.86</td>
<td>*-7.64</td>
<td>*6.70</td>
<td>*13.06</td>
<td>1.09</td>
<td>-1.23</td>
</tr>
<tr>
<td>$SE$</td>
<td>1.98</td>
<td>1.91</td>
<td>1.40</td>
<td>1.29</td>
<td>0.98</td>
<td>0.95</td>
</tr>
<tr>
<td>95% CI</td>
<td>[17.85,25.87]</td>
<td>[-11.51,-3.76]</td>
<td>[3.86,9.54]</td>
<td>[10.46,15.66]</td>
<td>[-.88,3.07]</td>
<td>[-3.14,0.69]</td>
</tr>
<tr>
<td>$SD$</td>
<td>12.37</td>
<td>11.95</td>
<td>8.76</td>
<td>8.03</td>
<td>6.10</td>
<td>5.91</td>
</tr>
<tr>
<td>Skewness</td>
<td>1.19</td>
<td>-0.40</td>
<td>0.10</td>
<td>0.38</td>
<td>-0.29</td>
<td>0.62</td>
</tr>
<tr>
<td>SES</td>
<td>0.38</td>
<td>0.38</td>
<td>0.38</td>
<td>0.38</td>
<td>0.38</td>
<td>0.38</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>4.87</td>
<td>0.62</td>
<td>-0.74</td>
<td>0.14</td>
<td>-0.08</td>
<td>-0.12</td>
</tr>
<tr>
<td>SEK</td>
<td>0.74</td>
<td>0.74</td>
<td>0.74</td>
<td>0.74</td>
<td>0.74</td>
<td>0.74</td>
</tr>
</tbody>
</table>

*Note.* * indicates significant treatment effect; ERFT = extensive reading with reading fluency; IR = intensive reading; CI = confidence interval; SES = standard error of skewness; SEK = standard error of kurtosis.

on silent reading rate was significant, $F(1, 38) = 100.15, p < .001$, partial $\eta^2 = .73$. The effect of treatment on oral reading rate was significant, $F(1, 38) = 12.48, p = .001$, partial $\eta^2 = .25$. The effect of treatment on oral reading fluency was not significant, $F(1, 38) = 3.98, p = .053$, partial $\eta^2 = .10$. This finding indicated that for students who took extensive reading with fluency training first and intensive reading second, the extensive reading with fluency training approach yielded a significantly greater gain in silent reading rate on authentic texts over one semester when compared to the intensive reading approach. However, the intensive reading approach yielded a significantly greater gain in oral reading rate on authentic texts over one semester when compared to the extensive reading with fluency training approach.

**Hypothesis 2: Reading Comprehension and Vocabulary Knowledge**

Hypothesis 2 stated that compared with the extensive reading with fluency training approach, the intensive reading approach would yield greater reading
comprehension gains when reading authentic texts as well as greater written receptive vocabulary size gains. To test this hypothesis, a repeated-measures ANOVA with two levels (extensive reading with fluency training and intensive reading) and three dependent variables (reading comprehension change with simplified texts, reading comprehension change with authentic texts, and vocabulary size change) were conducted. Changes were calculated by subtracting pretest from posttest results. All three measures are reported and analyzed as Rasch ability estimates (logits). Separate ANOVAs were conducted separately for the IR-before-ERFT students and the ERFT-before-IR students.

Descriptive statistics for the IR-before-ERFT ANOVA are reported in Table 45.

Table 45. Descriptive Statistics for Change in Reading Comprehension on Simplified Texts, Reading Comprehension on Authentic Texts, and Vocabulary Size During ERFT and IR Treatments, for IR-Before-ERFT Group (N = 38)

<table>
<thead>
<tr>
<th></th>
<th>ERFT-RCsimp (change in logits)</th>
<th>IR-RCsimp (change in logits)</th>
<th>ERFT-RCauth (change in logits)</th>
<th>IR-RCauth (change in logits)</th>
<th>ERFT-Vsize (change in logits)</th>
<th>IR-Vsize (change in logits)</th>
</tr>
</thead>
<tbody>
<tr>
<td>M</td>
<td>-0.10</td>
<td>-0.26</td>
<td>*0.17</td>
<td>*1.00</td>
<td>*-0.19</td>
<td>*0.36</td>
</tr>
<tr>
<td>SE</td>
<td>0.23</td>
<td>0.20</td>
<td>0.24</td>
<td>0.20</td>
<td>0.09</td>
<td>0.11</td>
</tr>
<tr>
<td>SD</td>
<td>1.41</td>
<td>1.26</td>
<td>1.49</td>
<td>1.24</td>
<td>0.57</td>
<td>0.70</td>
</tr>
<tr>
<td>Skewness</td>
<td>-0.38</td>
<td>-0.04</td>
<td>0.39</td>
<td>0.61</td>
<td>-0.10</td>
<td>0.28</td>
</tr>
<tr>
<td>SES</td>
<td>0.38</td>
<td>0.38</td>
<td>0.38</td>
<td>0.38</td>
<td>0.38</td>
<td>0.38</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>-0.12</td>
<td>-1.03</td>
<td>0.33</td>
<td>0.50</td>
<td>-0.82</td>
<td>-0.10</td>
</tr>
<tr>
<td>SEK</td>
<td>0.75</td>
<td>0.75</td>
<td>0.75</td>
<td>0.75</td>
<td>0.75</td>
<td>0.75</td>
</tr>
</tbody>
</table>

Note. * indicates significant treatment effect; RC = reading comprehension; simp = simplified text; auth = authentic text; Vsize = vocabulary size test; ERFT = extensive reading with reading fluency; IR = intensive reading; CI = confidence interval; SES = standard error of skewness; SEK = standard error of kurtosis.

With the Bonferroni adjustment required due to three dependent variables, the \( p \)-level required for significance was set at .017. Sphericity is not an issue in a repeated-
measures ANOVA with only two treatment levels (Field, 2009). The effect of treatment on reading comprehension (simplified text) was not significant, $F(1, 37) = .35, p = .560$, partial $\eta^2 = .01$. The effect of treatment on reading comprehension (authentic text) was significant, $F(1, 37) = 7.42, p = .010$, partial $\eta^2 = .17$. The effect of treatment on vocabulary size was significant, $F(1, 37) = 13.91, p = .001$, partial $\eta^2 = .27$. This finding indicated that the intensive reading approach yielded significantly greater gains in reading comprehension (with authentic texts) and vocabulary size over one semester when compared to the extensive reading with fluency training approach.

Descriptive statistics for the ERFT-before-IR ANOVA are reported in Table 46.

Table 46. Descriptive Statistics for Change in Reading Comprehension on Simplified Texts, Reading Comprehension on Authentic Texts, and Vocabulary Size During ERFT and IR Treatments, for ERFT-Before-IR Group (N = 39)

<table>
<thead>
<tr>
<th></th>
<th>ERFT-RCsimp (change in logits)</th>
<th>IR-RCsimp (change in logits)</th>
<th>ERFT-RCauth (change in logits)</th>
<th>IR-RCauth (change in logits)</th>
<th>ERFT-Vsize (change in logits)</th>
<th>IR-Vsize (change in logits)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$M$</td>
<td>*-0.09</td>
<td>*-0.99</td>
<td>*1.05</td>
<td>*0.11</td>
<td>*0.44</td>
<td>*-0.07</td>
</tr>
<tr>
<td>$SE$</td>
<td>0.19</td>
<td>0.18</td>
<td>0.22</td>
<td>0.20</td>
<td>0.08</td>
<td>0.11</td>
</tr>
<tr>
<td>$SD$</td>
<td>1.22</td>
<td>1.12</td>
<td>1.34</td>
<td>1.26</td>
<td>0.47</td>
<td>0.69</td>
</tr>
<tr>
<td>Skewness</td>
<td>-0.03</td>
<td>-0.54</td>
<td>0.79</td>
<td>-0.08</td>
<td>-0.15</td>
<td>-0.29</td>
</tr>
<tr>
<td>$SES$</td>
<td>0.38</td>
<td>0.38</td>
<td>0.38</td>
<td>0.38</td>
<td>0.38</td>
<td>0.38</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>0.29</td>
<td>-0.41</td>
<td>3.19</td>
<td>-0.43</td>
<td>-0.21</td>
<td>0.64</td>
</tr>
<tr>
<td>$SEK$</td>
<td>0.74</td>
<td>0.74</td>
<td>0.74</td>
<td>0.74</td>
<td>0.74</td>
<td>0.74</td>
</tr>
</tbody>
</table>

Note. * indicates significant treatment effect; RC = reading comprehension; simp = simplified text; auth = authentic text; Vsize = vocabulary size test; ERFT = extensive reading with reading fluency; IR = intensive reading; CI = confidence interval; $SES$ = standard error of skewness; $SEK$ = standard error of kurtosis.

With the Bonferroni adjustment required due to three dependent variables, the $p$-level required for significance was set at .017. Sphericity is not an issue in a repeated-measures ANOVA with only two treatment levels (Field, 2009). The effect of treatment
on reading comprehension (simplified text) was significant, $F(1, 38) = 12.27, p = .001$, partial $\eta^2 = .24$. The effect of treatment on reading comprehension (authentic text) was significant, $F(1, 38) = 10.07, p = .003$, partial $\eta^2 = .21$. The effect of treatment on vocabulary size was significant, $F(1, 38) = 13.91, p = .001$, partial $\eta^2 = .27$. This result indicated that the extensive reading with fluency training approach yielded significantly greater gains in reading comprehension (with authentic texts) and vocabulary size over one semester when compared to the intensive reading approach.

**Hypothesis 3: The Relationship Between Silent Reading Rate, Oral Reading Rate, Oral Reading Fluency, and Reading Comprehension**

Hypothesis 3 proposed that for the ERFT treatment, changes in silent reading rate would correlate more highly with reading comprehension changes than changes in oral reading rate and fluency did. To test this hypothesis, Pearson product-moment correlation coefficient tables showing the correlations among silent reading rate change, reading comprehension change, oral reading rate change, and oral reading fluency change during the extensive reading with fluency training treatment were generated. This analysis was conducted separately for the IR-before-ERFT group and the ERFT-before-IR group, and for both simple and authentic texts.

The descriptive statistics for simplified texts for the IR-before-ERFT group are reported in Table 47, and the correlation coefficients are reported in Table 48.
Table 47. Descriptive Statistics for the ERFT Treatment for Reading Comprehension Change, Silent Reading Rate Change, Oral Reading Rate Change, and Oral Reading Fluency Change for Simplified Texts, for IR-Before-ERFT Group (N = 38)

<table>
<thead>
<tr>
<th></th>
<th>RCsimp (change in logits)</th>
<th>SRR (change in swpm)</th>
<th>ORR (change in wcpm)</th>
<th>ORF (change in logits)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$M$</td>
<td>-0.10</td>
<td>2.99</td>
<td>9.79</td>
<td>0.27</td>
</tr>
<tr>
<td>$SE$</td>
<td>0.23</td>
<td>2.70</td>
<td>1.29</td>
<td>0.54</td>
</tr>
<tr>
<td>95% CI</td>
<td>[-.56, .36]</td>
<td>[-2.47, 8.45]</td>
<td>[7.18, 12.39]</td>
<td>[-.83, 1.36]</td>
</tr>
<tr>
<td>$SD$</td>
<td>1.41</td>
<td>16.62</td>
<td>7.93</td>
<td>3.33</td>
</tr>
<tr>
<td>Skewness</td>
<td>-0.38</td>
<td>0.35</td>
<td>-0.88</td>
<td>-0.52</td>
</tr>
<tr>
<td>SES</td>
<td>0.38</td>
<td>0.38</td>
<td>0.38</td>
<td>0.38</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>-0.12</td>
<td>0.06</td>
<td>1.4</td>
<td>0.27</td>
</tr>
<tr>
<td>SEK</td>
<td>0.75</td>
<td>0.75</td>
<td>0.75</td>
<td>0.75</td>
</tr>
</tbody>
</table>

Note. simp = simplified text; auth = authentic text; ERFT = extensive reading with reading fluency; IR = intensive reading; CI = confidence interval; SES = standard error of skewness; SEK = standard error of kurtosis.

Table 48. Pearson Correlation Coefficients for the ERFT Treatment for Reading Comprehension Change, Silent Reading Rate Change, Oral Reading Rate Change, and Oral Reading Fluency Change for Simplified Texts, for IR-Before-ERFT Group (N = 38)

<table>
<thead>
<tr>
<th></th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Reading comprehension change</td>
<td></td>
</tr>
<tr>
<td>2. Silent reading rate change</td>
<td>.035</td>
</tr>
<tr>
<td>3. Oral reading rate change</td>
<td>.054</td>
</tr>
<tr>
<td>4. Oral reading fluency change</td>
<td>-.043</td>
</tr>
</tbody>
</table>

*p < .05

Note. ERFT = extensive reading with reading fluency; IR = intensive reading.

The $p$-level required for significance was set at .006 due to eight correlations being computed. All three of the rate and fluency measures correlated weakly and non-significantly with reading comprehension of simplified texts.

The descriptive statistics for simplified texts for the ERFT-before-IR group are reported in Table 49, and the correlation coefficients are reported in Table 50.

The $p$-level required for significance was set at .006 due to eight correlations being computed. Oral reading rate correlated slightly more highly ($r = .172, p = .296$)
with reading comprehension than silent reading rate \((r = .123, p = .454)\), but none of the correlations met the criterion for significance.

Table 49. Descriptive Statistics for the ERFT Treatment for Reading Comprehension Change, Silent Reading Rate Change, Oral Reading Rate Change, and Oral Reading Fluency Change for Simplified Texts, for ERFT-Before-IR Group (\(N = 39\))

<table>
<thead>
<tr>
<th></th>
<th>RCsimp (change in logits)</th>
<th>SRR (change in swpm)</th>
<th>ORR (change in wcpm)</th>
<th>ORF (change in logits)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(M)</td>
<td>-0.09</td>
<td>19.22</td>
<td>12.55</td>
<td>2.04</td>
</tr>
<tr>
<td>(SE)</td>
<td>0.19</td>
<td>1.96</td>
<td>1.49</td>
<td>0.64</td>
</tr>
<tr>
<td>95% CI</td>
<td>[-.49,.30]</td>
<td>[15.26,23.17]</td>
<td>[9.53,15.57]</td>
<td>[.74,3.34]</td>
</tr>
<tr>
<td>(SD)</td>
<td>1.22</td>
<td>12.21</td>
<td>9.31</td>
<td>4.01</td>
</tr>
<tr>
<td>Skewness</td>
<td>-0.03</td>
<td>1.65</td>
<td>0.61</td>
<td>0.48</td>
</tr>
<tr>
<td>(SES)</td>
<td>0.38</td>
<td>0.38</td>
<td>0.38</td>
<td>0.38</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>0.29</td>
<td>5.78</td>
<td>0.21</td>
<td>2.13</td>
</tr>
<tr>
<td>(SEK)</td>
<td>0.74</td>
<td>0.74</td>
<td>0.74</td>
<td>0.74</td>
</tr>
</tbody>
</table>

*Note.* simp = simplified text; ERFT = extensive reading with reading fluency; IR = intensive reading; CI = confidence interval; SES = standard error of skewness; SEK = standard error of kurtosis.

Table 50. Pearson Correlation Coefficients for the ERFT Treatment for Reading Comprehension Change, Silent Reading Rate Change, Oral Reading Rate Change, and Oral Reading Fluency Change for Simplified Texts, for ERFT-Before-IR Group (\(N = 39\))

<table>
<thead>
<tr>
<th></th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Reading comprehension change</td>
<td>—</td>
</tr>
<tr>
<td>2. Silent reading rate change</td>
<td>.123</td>
</tr>
<tr>
<td>3. Oral reading rate change</td>
<td>.172</td>
</tr>
<tr>
<td>4. Oral reading fluency change</td>
<td>.028</td>
</tr>
</tbody>
</table>

*\(p < .05\)

*Note.* ERFT = extensive reading with reading fluency; IR = intensive reading.

The descriptive statistics for authentic texts for the IR-before-ERFT group are reported in Table 51, and the correlation coefficients are reported in Table 52.

The *p*-level required for significance was set at .006 due to eight correlations being computed. None of the rate/fluency measures correlated positively with reading comprehension, and none of the correlations met the criterion for significance.
Table 51. Descriptive Statistics for the ERFT Treatment for Reading Comprehension Change, Silent Reading Rate Change, Oral Reading Rate Change, and Oral Reading Fluency Change for Authentic Texts, for IR-Before-ERFT Group (N = 38)

<table>
<thead>
<tr>
<th></th>
<th>RCauth (change in logits)</th>
<th>SRR (change in swpm)</th>
<th>ORR (change in wcpm)</th>
<th>ORF (change in logits)</th>
</tr>
</thead>
<tbody>
<tr>
<td>M</td>
<td>0.17</td>
<td>-2.86</td>
<td>15.18</td>
<td>1.84</td>
</tr>
<tr>
<td>SE</td>
<td>0.24</td>
<td>2.25</td>
<td>1.22</td>
<td>0.86</td>
</tr>
<tr>
<td>95% CI</td>
<td>[-.32,.66]</td>
<td>[-7.41,1.70]</td>
<td>[12.72,17.64]</td>
<td>[.10,3.59]</td>
</tr>
<tr>
<td>SD</td>
<td>1.49</td>
<td>13.85</td>
<td>7.49</td>
<td>5.32</td>
</tr>
<tr>
<td>Skewness</td>
<td>0.39</td>
<td>-0.39</td>
<td>0.61</td>
<td>-0.81</td>
</tr>
<tr>
<td>SES</td>
<td>0.38</td>
<td>0.38</td>
<td>0.38</td>
<td>0.38</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>0.33</td>
<td>-0.30</td>
<td>-0.42</td>
<td>2.23</td>
</tr>
<tr>
<td>SEK</td>
<td>0.75</td>
<td>0.75</td>
<td>0.75</td>
<td>0.75</td>
</tr>
</tbody>
</table>

Note. auth = authentic text; ERFT = extensive reading with reading fluency; IR = intensive reading; CI = confidence interval; SES = standard error of skewness; SEK = standard error of kurtosis.

Table 52. Pearson Correlation Coefficients for the ERFT Treatment for Reading Comprehension Change, Silent Reading Rate Change, Oral Reading Rate Change, and Oral Reading Fluency Change for Authentic Texts, for IR-Before-ERFT Group (N = 38)

<table>
<thead>
<tr>
<th></th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Reading comprehension change</td>
<td></td>
</tr>
<tr>
<td>2. Silent reading rate change</td>
<td>-.124</td>
</tr>
<tr>
<td>3. Oral reading rate change</td>
<td>-.303</td>
</tr>
<tr>
<td>4. Oral reading fluency change</td>
<td>-.052</td>
</tr>
</tbody>
</table>

*p < .05

Note. ERFT = extensive reading with reading fluency; IR = intensive reading.

The descriptive statistics for authentic texts for the ERFT-before-IR group are reported in Table 53, and the correlation coefficients are reported in Table 54.

The p-level required for significance was set at .006 due to eight correlations being computed. Oral reading rate correlated more highly ($r = .312, p = .054$) with reading comprehension than silent reading rate ($r = .174, p = .288$), but none of the correlations met the criterion for significance.
Table 53. Descriptive Statistics for the ERFT Treatment for Reading Comprehension Change, Silent Reading Rate Change, Oral Reading Rate Change, and Oral Reading Fluency Change for Authentic Texts, for ERFT-Before-IR Group (N = 39)

<table>
<thead>
<tr>
<th></th>
<th>RCauth (change in logits)</th>
<th>SRR (change in swpm)</th>
<th>ORR (change in wcpm)</th>
<th>ORF (change in logits)</th>
</tr>
</thead>
<tbody>
<tr>
<td>M</td>
<td>1.05</td>
<td>21.86</td>
<td>6.70</td>
<td>1.09</td>
</tr>
<tr>
<td>SE</td>
<td>0.22</td>
<td>1.98</td>
<td>1.40</td>
<td>0.98</td>
</tr>
<tr>
<td>95% CI</td>
<td>[.61,1.49]</td>
<td>[17.85,25.87]</td>
<td>[3.86,9.54]</td>
<td>[-.88,3.07]</td>
</tr>
<tr>
<td>SD</td>
<td>1.34</td>
<td>12.37</td>
<td>8.76</td>
<td>6.10</td>
</tr>
<tr>
<td>Skewness</td>
<td>0.78</td>
<td>1.19</td>
<td>0.10</td>
<td>-0.29</td>
</tr>
<tr>
<td>SES</td>
<td>0.38</td>
<td>0.38</td>
<td>0.38</td>
<td>0.38</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>3.19</td>
<td>4.87</td>
<td>-0.74</td>
<td>-0.08</td>
</tr>
<tr>
<td>SEK</td>
<td>0.74</td>
<td>0.74</td>
<td>0.74</td>
<td>0.74</td>
</tr>
</tbody>
</table>

Note. auth = authentic text; ERFT = extensive reading with reading fluency; IR = intensive reading; CI = confidence interval; SES = standard error of skewness; SEK = standard error of kurtosis.

Table 54. Pearson Correlation Coefficients for the ERFT Treatment for Reading Comprehension Change, Silent Reading Rate Change, Oral Reading Rate Change, and Oral Reading Fluency Change for Authentic Texts, for ERFT-Before-IR Group (N = 39)

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Reading comprehension change</td>
<td>—</td>
<td>.174</td>
<td>.312</td>
<td>.227</td>
</tr>
</tbody>
</table>

*p < .05

Note. ERFT = extensive reading with reading fluency; IR = intensive reading.

Hypothesis 4: The Relationship Between Initial Reading Comprehension, Initial Reading Rate, and Reading Rate Changes

The fourth hypothesis proposed that for the extensive reading with fluency training treatment, there would be a negative relationship between initial reading ability (measured by reading comprehension or reading rate) reading rate increases during the extensive reading with fluency training treatment. To test this hypothesis, Pearson product-moment correlation coefficient tables showing the correlations among initial reading comprehension, initial silent reading rate, and silent reading rate change during
the extensive reading with fluency training treatment were generated. This analysis was conducted separately for the IR-before-ERFT group and the ERFT-before-IR group, and for both simple and authentic texts.

The descriptive statistics for simplified texts for the IR-before-ERFT group are reported in Table 55, and the correlation coefficients are reported in Table 56.

Table 55. Descriptive Statistics for the ERFT Treatment for Initial Reading Comprehension, Initial Silent Reading Rate, and Silent Reading Rate Change for Simplified Texts for IR-Before-ERFT Group (N = 38)

<table>
<thead>
<tr>
<th></th>
<th>Pre reading comprehension (logits)</th>
<th>Pre silent reading rate (swpm)</th>
<th>Silent reading rate change (swpm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>M</td>
<td>0.83</td>
<td>116.83</td>
<td>2.99</td>
</tr>
<tr>
<td>SE</td>
<td>0.21</td>
<td>3.49</td>
<td>2.70</td>
</tr>
<tr>
<td>95% CI</td>
<td>[.41, 1.25]</td>
<td>[109.76, 123.89]</td>
<td>[-2.47, 8.45]</td>
</tr>
<tr>
<td>SD</td>
<td>1.27</td>
<td>21.50</td>
<td>16.62</td>
</tr>
<tr>
<td>Skewness</td>
<td>-0.61</td>
<td>0.61</td>
<td>0.35</td>
</tr>
<tr>
<td>SES</td>
<td>0.38</td>
<td>0.38</td>
<td>0.38</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>0.97</td>
<td>0.06</td>
<td>0.06</td>
</tr>
<tr>
<td>SEK</td>
<td>0.75</td>
<td>0.75</td>
<td>0.75</td>
</tr>
</tbody>
</table>

Note. ERFT = extensive reading with reading fluency; IR = intensive reading; CI = confidence interval; SES = standard error of skewness; SEK = standard error of kurtosis.

Table 56. Pearson Correlation Coefficients for the ERFT Treatment for Initial Reading Comprehension, Initial Silent Reading Rate, and Silent Reading Rate Change for Simplified Texts for IR-Before-ERFT Group (N = 38)

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Silent reading rate change</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>2. Pre reading comprehension</td>
<td>.249</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>3. Pre Silent reading rate</td>
<td>.301</td>
<td>.220</td>
<td>—</td>
</tr>
</tbody>
</table>

*p < .05

Note. ERFT = extensive reading with reading fluency; IR = intensive reading.

The p-level required for significance was set at .008 due to six correlations being computed. Initial silent reading rate showed moderate correlation with silent reading rate change (r = .301, p = .066), but this did not meet the significance criterion. The
The correlation between initial reading comprehension and silent reading rate change was also not significant.

The descriptive statistics for simplified texts for the ERFT-before-IR group are reported in Table 57, and the correlation coefficients are reported in Table 58.

**Table 57. Descriptive Statistics for the ERFT Treatment for Initial Reading Comprehension, Initial Silent Reading Rate, and Silent Reading Rate Change for Simplified Texts for ERFT-Before-IR Group (N = 39)**

<table>
<thead>
<tr>
<th></th>
<th>Pre reading comprehension (logits)</th>
<th>Pre silent reading rate (swpm)</th>
<th>Silent reading rate change (swpm)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>M</strong></td>
<td>1.31</td>
<td>84.77</td>
<td>19.22</td>
</tr>
<tr>
<td><strong>SE</strong></td>
<td>0.18</td>
<td>2.48</td>
<td>1.96</td>
</tr>
<tr>
<td>95% CI</td>
<td>[.95,1.68]</td>
<td>[79.73,89.80]</td>
<td>[15.26,23.17]</td>
</tr>
<tr>
<td><strong>SD</strong></td>
<td>1.12</td>
<td>15.52</td>
<td>12.21</td>
</tr>
<tr>
<td><strong>Skewness</strong></td>
<td>0.36</td>
<td>0.95</td>
<td>1.65</td>
</tr>
<tr>
<td><strong>SES</strong></td>
<td>0.38</td>
<td>0.38</td>
<td>0.38</td>
</tr>
<tr>
<td><strong>Kurtosis</strong></td>
<td>-0.11</td>
<td>0.87</td>
<td>5.78</td>
</tr>
<tr>
<td><strong>SEK</strong></td>
<td>0.74</td>
<td>0.74</td>
<td>0.74</td>
</tr>
</tbody>
</table>

*Note. ERFT = extensive reading with reading fluency; IR = intensive reading; CI = confidence interval; SES = standard error of skewness; SEK = standard error of kurtosis.*

**Table 58. Pearson Correlation Coefficients for the ERFT Treatment for Initial Reading Comprehension, Initial Silent Reading Rate, and Silent Reading Rate Change for Simplified Texts for ERFT-Before-IR Group (N = 39)**

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Silent reading rate change</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>2. Pre reading comprehension</td>
<td>-.167</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>3. Pre silent reading rate</td>
<td>-.422**</td>
<td>.197</td>
<td>—</td>
</tr>
</tbody>
</table>

**p < .01

*Note. ERFT = extensive reading with reading fluency; IR = intensive reading.*

The $p$-level required for significance was set at .008 due to six correlations being computed. Initial silent reading rate showed a moderate-to-high and significant negative correlation with silent reading rate change ($r = -.422, p = .007$), which is in line with Hypothesis 4. There was also a weaker negative correlation between initial reading
comprehension and silent reading rate change \( (r = -.167, p = .310) \), although this was not significant.

The descriptive statistics for authentic texts for the IR-before-ERFT group are reported in Table 59, and the correlation coefficients are reported in Table 60.

Table 59. Descriptive Statistics for the ERFT Treatment for Initial Reading Comprehension, Initial Silent Reading Rate, and Silent Reading Rate Change for Authentic Texts for IR-Before-ERFT Group (\( N = 38 \))

<table>
<thead>
<tr>
<th></th>
<th>Pre reading comprehension (logits)</th>
<th>Pre silent reading rate (swpm)</th>
<th>Silent reading rate change (swpm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( M )</td>
<td>0.82</td>
<td>114.59</td>
<td>-2.86</td>
</tr>
<tr>
<td>( SE )</td>
<td>0.20</td>
<td>4.70</td>
<td>2.25</td>
</tr>
<tr>
<td>95% CI</td>
<td>[.43,1.22]</td>
<td>[105.07,124.11]</td>
<td>[-7.41,1.70]</td>
</tr>
<tr>
<td>( SD )</td>
<td>1.21</td>
<td>28.96</td>
<td>13.85</td>
</tr>
<tr>
<td>Skewness</td>
<td>-0.35</td>
<td>1.39</td>
<td>-0.39</td>
</tr>
<tr>
<td>SES</td>
<td>0.38</td>
<td>0.38</td>
<td>0.38</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>0.24</td>
<td>2.03</td>
<td>-0.30</td>
</tr>
<tr>
<td>SEK</td>
<td>0.75</td>
<td>0.75</td>
<td>0.75</td>
</tr>
</tbody>
</table>

*Note.* ERFT = extensive reading with reading fluency; IR = intensive reading; CI = confidence interval; SES = standard error of skewness; SEK = standard error of kurtosis.

Table 60. Pearson Correlation Coefficients for the ERFT Treatment for Initial Reading Comprehension, Initial Silent Reading Rate, and Silent Reading Rate Change for Authentic Texts, for IR-Before-ERFT Group (\( N = 38 \))

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Silent reading rate change</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>2. Pre reading comprehension</td>
<td>.085</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>3. Pre silent reading rate</td>
<td>-.233</td>
<td>.562***</td>
<td>—</td>
</tr>
</tbody>
</table>

***\( p < .001 \)

*Note.* ERFT = extensive reading with reading fluency; IR = intensive reading.

The \( p \)-level required for significance was set at \( .008 \) due to six correlations being computed. There was a small-to-moderate correlation between initial silent reading rate and silent reading rate change \( (r = -.233, p = .159) \), but this did not meet the criterion for significance. The correlation between initial reading comprehension and silent
reading rate change was small and not significant. The correlation between initial reading comprehension and initial silent reading rate, however, was large and significant \((r = .562, p < .001)\).

The descriptive statistics for authentic texts for the ERFT-before-IR group are reported in Table 61, and the correlation coefficients are reported in Table 62.

Table 61. Descriptive Statistics for the ERFT Treatment for Initial Reading Comprehension, Initial Silent Reading Rate, and Silent Reading Rate Change for Authentic Texts for ERFT-Before-IR Group \((N = 39)\)

<table>
<thead>
<tr>
<th></th>
<th>Pre reading comprehension (logits)</th>
<th>Pre silent reading rate (swpm)</th>
<th>Silent reading rate change (swpm)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>M</strong></td>
<td>-.14</td>
<td>74.15</td>
<td>21.86</td>
</tr>
<tr>
<td><strong>SE</strong></td>
<td>0.14</td>
<td>1.90</td>
<td>1.98</td>
</tr>
<tr>
<td><strong>95% CI</strong></td>
<td>[.43,.16]</td>
<td>[70.30,78.01]</td>
<td>[17.85,25.87]</td>
</tr>
<tr>
<td><strong>SD</strong></td>
<td>0.90</td>
<td>11.90</td>
<td>12.37</td>
</tr>
<tr>
<td><strong>Skewness</strong></td>
<td>0.89</td>
<td>1.04</td>
<td>1.19</td>
</tr>
<tr>
<td><strong>SES</strong></td>
<td>0.38</td>
<td>0.38</td>
<td>0.38</td>
</tr>
<tr>
<td><strong>Kurtosis</strong></td>
<td>1.20</td>
<td>2.82</td>
<td>4.87</td>
</tr>
<tr>
<td><strong>SEK</strong></td>
<td>0.74</td>
<td>0.74</td>
<td>0.74</td>
</tr>
</tbody>
</table>

*Note.* ERFT = extensive reading with reading fluency; IR = intensive reading; CI = confidence interval; SES = standard error of skewness; SEK = standard error of kurtosis.

Table 62. Pearson Correlation Coefficients for the ERFT Treatment for Initial Reading Comprehension, Initial Silent Reading Rate, and Silent Reading Rate Change for Authentic Texts for ERFT-Before-IR Group \((N = 39)\)

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Silent reading rate change</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>2. Pre reading comprehension</td>
<td>.083</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>3. Pre silent reading rate</td>
<td>-.212</td>
<td>.214</td>
<td>—</td>
</tr>
</tbody>
</table>

*Note.* ERFT = extensive reading with reading fluency; IR = intensive reading.

The \(p\)-level required for significance was set at .008 due to six correlations being computed. There was small-to-moderate negative correlation between initial silent reading rate and silent reading rage change \((r = -.212, p = .196)\), but this was not
significant. The correlation between initial reading comprehension and silent reading rate change ($r = .083, p = .614$) was also small and not significant.

**Hypothesis 5: Silent Reading Rate Change Patterns**

Hypothesis 5 was an exploration of the pattern of silent reading rate change during a semester of extensive reading with fluency training. During that semester, there were three silent reading rate measurements (simplified text) spaced evenly over the course of the semester, in addition to the pre-course and post-course measurements. To obtain a finer-grained analysis of the patterns of reading rate change during the semester, rate change graphs for all participants were generated and examined (Appendix G). Using Chung and Nation (2006) and Tran (2012) as a starting point, data-driven categories of change patterns were identified, and the number of participants in each pattern category was tallied and reported.

Six patterns were visually identified: gradual increase, up-down increase, jagged increase, down-up increase, plateau increase, and decrease. The definition of each pattern was established as follows, so that participant data graphs could be placed into categories and tallied in a systematic way.

If there was an overall increase and the rate decreased only one time between measurements, as in Figure 17, this was considered a gradual increase.
If there was an overall increase and the rate increase 2-3 times followed by 1-2 decreases, forming the shape of either a mountain or hill, as in Figure 18, this was termed an up-down increase.
If there was an overall increase with non-consecutive increases and decreases, forming jagged spikes, as in Figure 19, this was termed a jagged increase.
If there was an overall increase and the rate decreased 1-2 times followed by 2-3 increases, forming the shape of either a hole or a trench, as in Figure 20, this was termed a down-up increase.
If there was an overall increase and an initial 1-2 increases followed by a plateau, as in Figure 21, it was termed a plateau increase. To fit this definition, the final three data points must fall within a range of 10 wpm.
The overall tabulation of the number of participants in each pattern category is shown in Table 63.

Based on a few previous research studies (Chung & Nation, 2006; Tran, 2012), I hypothesized that students would display a variety of patterns of silent reading rate increase during the extensive reading with fluency training semester, and that a gradual increase would be the most common pattern. However, the results of this study indicate
Table 63. *Silent Reading Rate Change Patterns During the ERFT Treatment (N = 77)*

<table>
<thead>
<tr>
<th>Pattern</th>
<th>Cases</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gradual increase</td>
<td>12</td>
</tr>
<tr>
<td>Up-down increase</td>
<td>16</td>
</tr>
<tr>
<td>Jagged increase</td>
<td>22</td>
</tr>
<tr>
<td>Down-up increase</td>
<td>1</td>
</tr>
<tr>
<td>Plateau increase</td>
<td>6</td>
</tr>
<tr>
<td>Decrease</td>
<td>20</td>
</tr>
</tbody>
</table>

that the jagged increase and even the up-down increase were more common than the gradual increase.
CHAPTER 6
DISCUSSION

In this chapter, the results of the data analysis for each hypothesis are summarized and discussed in turn, with reference to previous literature. It should be clearly noted here that from the start of this study, I had no particular predilection about whether extensive reading or intensive reading is more or less effective for developing EFL reading ability in general. It is true that intensive reading is the more traditional and entrenched approach in Japan and other EFL environments, and extensive reading can be viewed as a movement as much as an approach, and it is a movement designed to challenge or at least supplement the traditional approach by adding something that is missing in the traditional approach (namely, a focus on improving students’ enjoyment of and confidence in reading, getting them to read more extensively, and improving reading rate and probably fluency). However, this by no means implies that I have a preference for one approach over the other in general. Rather, I think there is a likelihood that each approach is particularly suited for developing particular aspects of reading ability. This possibility is investigated in Hypothesis 1 and 2 of this study and discussed in this section. These hypotheses test the notion that extensive reading more effective for improving reading rate and fluency, but that intensive reading might be more effective for improving reading comprehension (accuracy) and receptive vocabulary knowledge.
Hypothesis 1: Reading Rate and Fluency Gains

Hypothesis 1 posited that during the extensive reading with fluency training semester, students would demonstrate greater gains in silent and oral reading rate and fluency, when reading both simplified and authentic texts, than during the IR semester.

The data in Tables 41–44 support the hypothesis in many cases, but provide contraindicative evidence in some cases as well. Overall, there were five instances of data supporting the hypothesis and two instances of data contradicting the hypothesis. The results were more supportive of the hypothesis when the extensive reading with fluency training semester was before the IR semester. This might be because overall the students were more motivated during the first semester of college than the second. Extensive reading was likely a new concept for them after six years of secondary education with a primarily intensive reading-oriented approach. Starting off their first semester with a novel approach, and with reading easier books and choosing books they wanted to read, might have motivated them more than the other students, who could have felt that they were doing more of the same with the intensive reading approach. In other words, the differences between the two approaches were likely more pronounced during the first semester. Another factor might be that since the group that did intensive reading during the first semester included the advanced class and had a much higher mean initial silent reading rate (116.83 vs. 84.77 swpm; see Tables 55 and 56), they had less “room to grow” in terms of rate increase.

Evidence supporting the hypothesis is seen in the significantly greater increase in oral reading rate while reading simplified texts during the extensive reading with
fluency training semester regardless of the order of treatment, the significantly greater increase in silent reading rate while reading simplified texts during the extensive reading with fluency training semester when the order was ERFT-before-IR, the significantly greater increase in oral reading rate while reading authentic texts during the extensive reading with fluency training semester when the order was IR-before-ERFT, and the significantly greater increase in silent reading rate while reading authentic texts during the extensive reading with fluency training when the order was ERFT-before-IR.

Evidence contradicting the hypothesis is seen in the significantly greater increase in silent reading rate while reading authentic texts during the intensive reading semester when the order was IR-before-ERFT, and the significantly greater increase in oral reading rate while reading authentic texts during the intensive reading semester when the order was ERFT-before-IR.

In all five instances where the difference between extensive reading with fluency training and intensive reading was non-significant (and therefore not mentioned above), the results favored the extensive reading with fluency training treatment.

In terms of the general superiority of extensive reading with fluency training over intensive reading for improving reading rate and fluency, these results are largely consistent with previous studies (Beglar et al., 2012; Huffman, 2014; McLean & Rouault, 2017; Suk, 2017), adding to this expanding body of evidence. In the current study, the average silent reading rate change experienced by students during one semester of extensive reading with fluency training (for the ERFT-before-IR group) was
21.72 wpm greater than that experienced with the intensive reading treatment. Beglar et al. (2012) reported a comparable difference of 16.85 wpm, Huffman (2014) reported a difference of 21.35 wpm, McLean and Rouault (2017) reported a difference of 25.81 wpm, and Suk (2017) reported a difference of 19.6 wpm. So the evidence that ER leads to significantly greater silent reading rate gains seems well-established at this point.

Development of reading fluency involves increased speed, increased decoding accuracy, and the automatization and synchronization of multiple subskills such as syntactic parsing and chunking (Grabe, 2009; Kuhn & Stahl, 2003; LaBerge & Samuels, 1974; Pikulski & Chard, 2005). This happens more efficiently when students read texts that are easy for them, and it happens more quickly when students are reading a large volume of text over a long period of time. Therefore, a class that encourages students to meet these conditions, with both carrots and sticks (setting reading goals for the students, tying their grade to achievement of those goals, encouraging them to read books that they enjoy, and raising their awareness of the importance of reading a lot and improving their reading speed) seems logically more likely to result in improvement of those skills.

One area where this study expands on previous research is that it provides a fuller battery of reading speed and fluency measurements, going beyond the established “silent reading rate with evidence of comprehension” formula and measuring changes in both oral reading rate (a nearly-objective measurement) and oral reading fluency (a subjective, rater-based measurement). The silent reading rate measurement showed a significant difference in three of the four analyses, one of which was counterevidence.
The oral reading rate measurement showed a significant difference in all four analyses one of which was counterevidence. The oral reading fluency measurement did not reveal a difference in any of the four analyses. These results indicate that the oral reading rate measurement can likely be used as an alternative or complement to the silent reading rate measurement when measuring reading rate, but that the larger and more complex construct of reading fluency is more problematic to measure using the rating rubric employed in this study. Alternatively, this might indicate that the simpler, more objective silent reading rate and oral reading rate measurements are accurate measurements of rate but do not quite capture the fluency construct.

Looking at simplified texts versus authentic texts is something that has not been done in most previous studies examining the effect of ER and IR on reading rate or fluency (Macalister [2010] and Tran [2012] are exceptions). Of the five instances where significant difference supported Hypothesis 1, three of these showed extensive reading with fluency training yielding greater increases in reading rate on simplified texts and two of them showed greater rate increases on authentic texts. Of the two instances of counterevidence, both involved authentic texts. Taken on the whole, these results seem to provide some (but not particularly clear) evidence that a semester of reading simplified texts helps more students improve their reading rate and fluency to a greater extent when reading simplified texts than with authentic texts, and that both the extensive reading with fluency training and intensive reading treatments can yield some reading rate and fluency increases when reading authentic texts. In the same way that practicing piano would give someone at least some advantage when learning to play the
organ, due to overlap of the skills involved, it makes logical sense that fluency
development that occurs with simplified texts would transfer to authentic texts.

Looking at the possibility of an order effect, the overall verdict is unclear. Three
of the five instances of significant difference supporting Hypothesis 1 occurred in the
ERFT-before-IR group, while two occurred in the IR-before-ERFT group. The two
instances of counterevidence were split evenly between the ERFT-before-IR and IR-
before-ERFT group. However, if only silent reading rate changes are examined, the
ERFT-before-IR group demonstrated a clear benefit for extensive reading with fluency
training, while the IR-before-ERFT group did not (Tables 40 and 41).

In light of previous research on the link between extensive reading and reading
rate gains (Beglar et al., 2012; Huffman, 2014; McLean & Rouault, 2017), the lack of
clarity and consistency of evidence for that link in these results was somewhat
surprising. In addition to the possibility of motivational and order effects discussed
above, another possibility is that the duration and the amount of text read in the
extensive reading semester was simply not enough to provide measurable reading rate
and fluency gains for many students. Students read an average of around 113,000
standard words during one semester in this study, but Burrows (2012) found 150,000
words to be a minimum threshold for extensive reading to yield large gains in reading
comprehension, Beglar and Hunt (2014) determined that only students who read around
200,000 words during an academic year made substantial gains in reading rate. It is
likely that the students in this current study would see greater reading fluency gains if
they continued with the approach for a full year.
Hypothesis 2: Reading Comprehension and Vocabulary Knowledge

The second hypothesis suggested that while extensive reading with fluency training would likely result in greater reading rate and fluency gains, and while extensive reading with fluency training should yield moderate reading comprehension gains when reading simplified texts, the intensive reading approach would yield greater reading comprehension gains when reading authentic texts as well as greater written receptive vocabulary size gains.

Overall, the results of this analysis do not clearly support or contradict the hypothesis. In fact, they seem to suggest that both the extensive reading with fluency training approach and the intensive reading approach have similar effects on reading comprehension and written receptive vocabulary size gains. For the IR-before-ERFT group, the intensive reading approach yielded significantly greater gains in both reading comprehension (authentic texts) and vocabulary size. However, for the ERFT-before-IR group, the extensive reading with fluency training approach yielded significantly greater gains in reading comprehension (authentic texts) and vocabulary size. Moreover, the magnitude of the gains was similar for both groups. Gains were not seen for either group on the measure of reading comprehension of simplified text, probably because the simplified text portion of the reading comprehension test showed moderate to strong ceiling effects, as mentioned in Chapter 4 and seen in Figures 9–12.

Whichever approach was conducted during the first semester was the more effective approach, meaning that both approaches seemed to be equally effective in improving reading comprehension and vocabulary knowledge. This would make sense
if motivational factors related to first semester are strong, as mentioned in the
discussion of Hypothesis 1. The students in this study were first-year university students
coming into a highly-regarded university, into which they had worked hard to be
accepted. A qualitative component to this study could have shed further light on their
feelings and motivations, but it is likely that they had high hopes and intentions and
expectations for themselves and their studies, particularly due to the fact that the
university is well known within the field of nursing for its emphasis on English and
international education. However, the first semester and first year are rigorous, in both
their English courses and other courses. The amount of reading in the extensive reading
class, the difficulty level of the reading and the extensive vocabulary study in the
intensive reading class, and the various pretests and posttests in both class might have
combined to make many students feel a heavy burden and great stress in relation to their
studies. This might have sapped some of their energy and excitement gradually over the
course of the first semester and especially second semester. By the end of the academic
year, they might have felt far less motivation toward the posttests, particularly since the
posttests analyzed in this study were not directly linked to the students’ grades.

The theoretical evidence that extensive reading improves reading
comprehension is strong. Both Nathan and Stanovich (1991) and Adams (1994), for
example, point out that when readers have to expend a large portion of their cognitive
capacity on the mechanics of decoding and word recognition, little cognitive capacity is
available for global comprehension of the text. When the lower-level skills become
rapid, effortless, and automatized, however, comprehension improves. Empirical
evidence that reading comprehension improves due to extensive reading has also accumulated (Homan at al., 1993; Pichette, 2005; Suk, 2017). However, logic also dictates that a semester spent focusing specifically on comprehension, including answering and discussing comprehension questions during class time, would result in improved performance on comprehension tests. It could even be argued that the class content and activities prepare the students specifically for these tests. The intense vocabulary study element of the intensive reading semester would theoretically improve reading comprehension, as vocabulary knowledge is a strong predictor of reading comprehension. Anderson and Freebody (1981) proposed a direct causal relationship between vocabulary knowledge and reading comprehension, and this has been empirically supported in many studies (e.g., Hu & Nation, 2000; Jeon & Yamashita, 2014; Qian, 1999; Schmitt et al., 2011).

Regarding vocabulary learning in particular, the two approaches investigated in this study engaged the students in different learning strategies. The extensive reading approach used only implicit vocabulary learning strategies, while the intensive reading approach included an explicit approach to vocabulary learning as well. Of the vocabulary learning strategies and processes introduced in Nation (2001), the word card study in the intensive reading semester of this study involved noticing and choosing (students were instructed to make word cards only for the words they were unfamiliar with from a larger list of words selected by the teacher), spaced retrieval and repetition (students were instructed to apply this concept and keep a record demonstrating it), and
making use of dictionaries and other reference sources (to write the definition and example sentence on their word cards).

Nation (2001) also discusses incidental learning from context as vocabulary learning strategy. This refers to the vocabulary learning that occurs naturally through extensive reading or listening, conversations in the target language, etc. He points out that learners in EFL environments face difficulty meeting the conditions necessary for this sort of learning to take place: namely, exposure to large amounts of text and a high percentage of coverage (in terms of learners’ knowledge) of the vocabulary encountered in the text. In order for learners to guess the meaning of vocabulary from context while reading, they need 95-98% coverage of the tokens in the text (Hu & Nation, 2000), a condition which was likely met in this study based on the text difficulty, students’ initial vocabulary size, and the fact that students were encouraged to read books at a level where they already knew 95% or more of the words in the text. Nation and Waring (2020) have asserted that extensive reading is not primarily viewed as a vocabulary learning task, but that it is likely to enhance vocabulary knowledge in the areas of fluency and automaticity. Students in the extensive reading semester of this study were told that they should avoid using dictionaries as much as possible in order to maintain reading flow, but that they could use dictionaries when necessary for unknown words that are key to understanding what is happening in the story. This position is consistent with that espoused by Nation and Waring (2020).

Based on the prior theoretical and empirical evidence, as well as the results of the data analysis in this study, it seems likely that both ER and IR contribute in different
ways to reading comprehension improvement, but that the magnitude of that contribution might be similar.

**Hypothesis 3: The Relationship Between Silent Reading Rate, Oral Reading Rate, Oral Reading Fluency, and Reading Comprehension**

This hypothesis focused on how the different methods of measuring reading rate and fluency correlate with reading comprehension. Specifically, it was hypothesized that for Japanese EFL learners in the extensive reading with fluency training treatment, changes in silent reading rate would correlate more highly with reading comprehension changes than changes in oral reading rate and fluency do.

Unfortunately, the data analyses did not support the hypothesis, nor did they provide evidence for any claims about whether silent reading rate, oral reading rate, or oral reading fluency are more or less correlated with reading comprehension in Japanese EFL learners at the university level.

In L1 contexts, an empirically-supported high correlation between oral reading rate and reading comprehension has been well established, so much so that oral reading rate is often used as an easy and efficient proxy for reading comprehension (Fuchs et al., 2001). However, some researchers in EFL contexts have questioned whether this correlation extends to EFL environments, and these questions are grounded in both theory and empirical analyses. Dhanapala and Yamada (2014) surmised that Japanese EFL students read out loud more slowly than when they read silently, and both Lems (2006) and Jeon (2012) suggested that students with L1 orthographies different from
English might be much slower at oral reading, which involves phonological recoding (pronouncing) words than they are at silent decoding and comprehension. Both Lems (2006) and Jeon (2012) provided empirical support for this hypothesis.

However, the simple correlational analyses conducted in this study did not support this hypothesis. None of the reading rate and fluency measures used in this study correlated significantly with reading comprehension. The lack of correlation between the rate/fluency measures and reading comprehension of simplified texts can be explained by the fact that ceiling effects in the test (mentioned above) prevented possible increases in comprehension from being revealed by the measure. The lack of correlation with authentic text comprehension is less easy to explain, but it could be that for lower-level EFL learners, students are forced to make a choice between reading accuracy and fluency. However, this is not consistent with studies reviewed in the literature that show a positive correlation between reading comprehension and oral reading fluency in L1 learners.

**Hypothesis 4: The Relationship Between Initial Reading Comprehension, Initial Reading Rate, and Reading Rate Changes**

Hypothesis 4 posited that during the extensive reading with fluency training semester, students with lower initial reading ability (operationalized as reading comprehension ability or silent reading rate) would experience higher silent reading rate increases during the semester than their counterparts with higher initial reading ability, and vice versa. It was hypothesized, therefore, that initial reading comprehension and
initial silent reading rate would correlate inversely with the change in silent reading rate that occurred during the semester.

Three of the four analyses conducted in this section did not support the hypothesis, finding no significant correlations between initial reading rate and comprehension on the one hand, and silent reading rate change on the other. However, a significant, moderately high, inverse correlation ($r = -0.422$) was found between initial silent reading rate and silent reading rate change for the ERFT-before-IR group while reading simplified texts. That result provides some support for the hypothesis, but is ultimately outweighed by the lack of correlation found in the other three analyses.

This hypothesis was based on theory (the general concept of learning curves that is based on the power law, described below) rather than on previous empirical studies, and in fact I am not aware of any previous studies investigating this correlation. The theoretical basis was simply that for a given student’s learning curve, whether reading rate or reading comprehension, and if a one-semester window of that curve is examined, the slope of that curve would be steeper for a student with a lower initial ability than for one with a higher initial ability. This is because the process of learning a new activity follows the power law, meaning that gains are largest in the early stages and diminish with continued practice (Logan, 1997). Reading fluently in English might be akin to a new activity for low-proficiency EFL readers in their first year of college, but overall the evidence found in this study do not clearly support this conclusion.

Although the overall results do not warrant a strong claim on this hypothesis, the strong inverse correlation found between initial silent reading rate and silent reading
rate change for the ERFT-before-IR group while reading simplified texts would justify further exploration of this phenomenon in future studies. This would be consistent with Takase’s (2012) finding that ER effectively motivates reluctant readers, and it might also lead teachers to place ER at the beginning stages of reading curricula, such as in the first semester or year of university education, or earlier, in secondary education contexts. Ideally, this is where the biggest gains in rate and fluency and comprehension (in EFL contexts, of course) can and should occur.

**Hypothesis 5: Silent Reading Rate Change Patterns**

The purpose of this hypothesis was to investigate, in a fine-grained and individualized manner, the silent reading rate changes of individual students during the extensive reading with fluency training semester. It was hypothesized that a variety of rate change patterns would emerge, but that a gradual, more-or-less linear increase would be the most common pattern.

Rate change graphs were generated for all participants, and six patterns were identified and defined. The results contradicted the hypothesis, with 22 of the students (N = 77) demonstrating a jagged increase, 16 an up-down increase, and only 12 the hypothesized gradual increase. Twenty students saw their silent reading rate decrease during the extensive reading with fluency training semester. Most of those students were in the IR-before-ERFT group, which had a mean reading rate gain of only 3 swpm during the extensive reading with fluency training semester, which was their second
semester. It seems likely that lack of motivation during the posttest, which was at the end of a stressful year and was not linked to their grades, was a contributing factor.

Because the pattern categories in this study were data-driven, they cannot be compared precisely with those of the two previous studies that have performed similar analyses (Chung & Nation, 2006; Tran, 2012). The jagged increase in this study corresponds loosely with the erratic increase in those two studies, but was defined somewhat differently. One reason for this was that both previous studies had 20 measurements, whereas this study only had 5 measurement points. This study also had a relatively large number of students demonstrating the up-down increase pattern, which did not occur often in the previous studies and which necessitated adding that category.

In both Chung and Nation (2006) and Tran (2012), most patterns were categorized as gradual increase (75.00% and 81.97%, respectively). In this study, however, only 15.58% were gradual increase, while 28.57% were jagged increase and 20.78% were up-down increase. This departure from the findings of previous studies could be due to variations in student motivation, variations in the difficulty of the texts that are hidden in the Flesch-Kincaid readability calculations, or some combination thereof. A possible reason for the prevalence of the up-down increase is that students’ motivation toward the class in general, and toward the battery of tests in particular, seemed to wane toward the end of the semester and especially the end of the year.
Summary of the Findings

The data analyses conducted in this study showed that, for first-year EFL students at a high-level nursing college in Japan, the extensive reading with fluency training approach generally yielded greater gains in silent and oral reading rate, especially while reading simplified texts, than the intensive reading approach. Regardless of order of treatment, for example, students’ oral reading rate gains for simplified texts were significantly greater during the extensive reading with fluency training semester. The ERFT-before-IR group also experienced greater gains in silent reading rate for simplified and authentic texts during the extensive reading semester, and the IR-before-ERFT group achieved greater gains in oral reading rate for authentic texts during the extensive reading semester.

However, some data showed that intensive reading yielded greater reading rate gains than extensive reading. For authentic texts, the IR-before-ERFT group experienced greater silent reading rate gains, and the ERFT group experienced greater oral reading rate gains during the intensive reading semester than the extensive reading semester.

Overall, the evidence also pointed to an order effect: the extensive reading with fluency training approach outperformed the intensive reading approach on the reading
rate measures particularly well when the extensive reading with fluency training approach was done first semester and the intensive reading approach second semester.

This study also investigated whether ERFT or IR leads to greater reading comprehension and vocabulary size gains, with the mixed results showing that both approaches likely contribute in different ways to reading comprehension improvement. The order effect appeared important here as with the first hypothesis, with the approach used in the first semester often yielding greater reading comprehension and vocabulary size gains. The IR-before-ERFT group experienced significantly greater gains in both reading comprehension of authentic texts and vocabulary size during the intensive reading semester, while for the ERFT-before-IR group the gains were greater during the extensive reading with fluency training semester.

Regarding the question of whether silent or oral reading rate correlates more highly with reading comprehension in EFL learners, the results did not support a clear conclusion. The L1 reading research demonstrates a strong and reliable relationship between oral reading rate and reading comprehension, but some researchers have suggested that this might not hold true for adult EFL learners. However, the tests and analyses in this study did not reveal any significant correlations between reading comprehension and the three measures of reading rate and fluency employed.

Although not clear enough to warrant substantial claims, limited support was found for the hypothesis that students with lower initial reading rate experience larger rate gains during an extensive reading course. Specifically, a significant, moderately high, inverse correlation was found between initial silent reading rate and silent reading
rate gains (for simplified texts) for the ERFT-before-IR group. However, a significant correlation was not found for the IR-first group, nor for authentic texts.

Finally, it was found that students in a one-semester extensive reading with fluency training course demonstrate a variety of different patterns of reading rate change during the course, with 22 students showing a jagged increase, 16 students an up-down increase, and 12 students a gradual increase pattern. However, 20 of the 77 students (most of them in the IR-before-ERFT group) saw their silent reading rate decrease during the extensive reading with fluency training semester.

**Limitations**

This study was specifically designed to avoid the pitfalls and limitations evident in previous research on reading fluency gains in extensive reading courses. However, it was not without limitations of its own.

The primary limitation was the fact that all participants were administered the four measurements in the same order, but half of the students had ERFT before IR while the other half had IR before ERFT. This necessitated analyzing the data for these two groups separately (due to non-equivalent test forms), which cut the sample size in half and might have negatively affected the clarity of the results. However, this limitation also made it possible to investigate the effect of changing the order of the approaches. Some evidence was found that doing the extensive reading with fluency training approach before the intensive reading approach yields better results.
Another limitation was the fact that the students participating in this study had a limited range of English proficiency, as demonstrated by their proficiency scores. As seen in Table 38, the mean CASEC score was 581.33 ($SD = 66.56$) (this corresponds to a TOEIC score of 523.97), and the 95% confidence interval was 566.33 to 596.34. This limited range of proficiency is common in Japan, due to the entrance exam system. English is a heavy component of these exams, so students attending a given university tend to fall within a narrow range of English proficiency levels.

A related limitation was that an advanced class (based on initial English proficiency scores) of 17 students was included in the 38-student IR-before-ER group. The higher proficiency level of this group might have acted as a confounding variable, weakening any claim that the results were attributable primarily to the reading approach used in the two conditions. The advanced students might have had an advantage in terms of their proficiency, or alternatively their initially high reading ability might have made it difficult for them to show high gains during the semester.

Another unknown variable that might have affected the results, and one that is common in longitudinal in-classroom studies like this one, is the learning activities that took place outside the context of the reading course investigated here. Students were enrolled in other English courses, specifically a writing course and an oral communication course, during the two semesters. Sixteen of the 77 participants joined a three-week study abroad program to Canada during the summer between first and second semester. Other individual students might have studied abroad or joined extracurricular language study programs or schools. Any or all of these activities might
have helped individual students increase their reading rate, reading comprehension ability and/or vocabulary size in ways that were neither controllable nor considered in this study.

The treatment period and amount of text read in the extensive reading semester might not have been enough for many students to begin realizing measurable reading rate, fluency, and comprehension gains. There is some evidence that 150,000 (Burrows, 2012) to 200,000 words (Beglar & Hunt, 2014) words of text needs to read before substantial rate and comprehension gains become apparent.

In the silent reading rate, oral reading rate and fluency, and reading comprehension measures employed in this study, the texts used were excerpted portions of longer stories. However, when reading a story, comprehension is a contextual and cumulative process, with everything that has happened previously aiding in the comprehension of a given sentence or paragraph or section of text. As described in the discussion of Kintsch’s (1988) construction-integration model in Chapter 2, reading is akin to putting together a puzzle, where the previously placed pieces aid in the act of placing each successive piece. Using excerpts detached from their original contexts, therefore, might have reduced the ecological validity of the tests, because the students were not doing exactly the same thing they do when reading a book or text in real-life conditions. This concern is higher for the oral reading rate/fluency texts, which were 40 standard words each, than for the longer excerpts used in the silent reading rate and reading comprehension tests.
A final possible limitation concerns the possibility of researcher bias. I have previously mentioned (see the beginning of Chapter 6 specifically) that my feeling at the start of this study was that extensive reading and intensive reading are both valuable and effective approaches that both provide different benefits to EFL learners, and that these benefits likely complement each other in a synergistic way. However, it is also true that I view intensive reading as the default approach to reading instruction in Japan, and the approach that most students are most familiar with. The opposite is also true: I think students are less likely to be familiar with, to know about, and to have experienced an intentional, well-designed extensive reading course. For these reasons, I might have spent more time and energy explaining, and even came across as more enthusiastic about, extensive reading. Even though I explicitly taught the students about the likely benefits of both approaches, I probably gave many more details, and focused on the evidence-based nature of, extensive reading than intensive reading. For these reasons, researcher bias might have acted as a hidden variable affecting the results.

**Suggestions for Future Research**

Publication bias is a well-known and much-discussed problem in all fields of research, and it is apparent even in the literature review of this dissertation. Generally speaking, only positive results are published, so a review of previously published literature excludes a portion of research from the start. Concerning common-sense and perhaps even obvious notion that extensive reading is likely to yield greater reading rate gains than intensive reading, for example, a review of the existing literature might lead
one to believe that the case is closed and the connection is clear. The reality, as shown in this study, is messier. The IR approach yielded rate gains in a number of analyses. The oral reading fluency rating did not significantly favor either ERFT or IR. Even if we accept that ER yields greater reading fluency benefits than IR, the correlation between rate and comprehension was not supported by this data. These discrepancies provide hints for how future research should proceed in this general area of applied linguistics.

There is still a need for more and stronger evidence of the link between ER and reading rate and fluency improvement. There have still been only a handful of studies (particularly if those with serious limitations are discounted) establishing this link, and we must assume that there could be an equally large set of unpublished counterevidence as well. Based on the similarity of results for the silent reading rate and oral reading rate measurements, future researchers should consider incorporating oral reading rate measurements more often as well.

Ultimately, the results of this study indicate a variety of benefits from both ER and IR, so another possible direction for future research would be to investigate whether integrating the ER and IR approaches within a single curriculum, so that students do them simultaneously, provides additional benefits beyond those seen in a curriculum such as the one in this study, where the two approaches are undergone in separate semesters. This might be more challenging to design, but could be achieved by keeping the approaches separate one year (or semester) and then integrating them the next.
Longitudinal, in-classroom research on the effectiveness of educational approaches should adopt a mixed-methods approach whenever possible, so that quantitative evidence can be supported or otherwise explained with qualitative evidence sources such as questionnaires or interviews. For example, in this study it would have been useful to have known the extent to which motivational factors did indeed affect students’ performance at the end of the semester or year.

Another direction for future research that came to light in this study is to investigate various types of “ER plus” in addition to the ER approach that has become more and more standardized of the past few decades. The varying effects of the standard approach versus an ER plus repeated readings approach, or ER plus IR, or ER plus extensive listening, or an ER plus vocabulary learning approach could be investigated. The goals of empirical research (controlling, isolating, and standardizing) are often at odds with the goals of pedagogy (flexible, student-centered integration of approaches), so researchers must be persistently creative in order to meet these competing demands.

Finally, future researchers need to keep in mind the likelihood that the effect of individual differences often outweighs the effect of particular instructional or learning approaches. Longitudinal, fine-grained, individualized approaches to understanding how students’ abilities progress over a semester or year, especially if they are accompanied by qualitative research components, would likely reveal patterns that shed more insight on how reading fluency and the components thereof develop over time.
Final Conclusions

Reading ability remains an essential key to success and fulfilment in academia, in research, and in life. Reading instruction will therefore continue to occupy a prominent position and play an important role in higher education curricula in EFL environments. Students will continue to ask educators about the most effective ways to improve their English proficiency in general and their reading skills in particular. Educators and researchers will continue to search for clearer answers to these questions.

One takeaway from this study is that reading ability, reading fluency, and reading comprehension are complex constructs consisting of an array of subskills, the interplay among which is not easily nailed down. Understanding which approaches are most effective in developing which skills remains an important goal in applied linguistics, because it allows educators to tailor their curricula to individual contexts, classes, and even students. This study takes us one more step toward that understanding, providing educators with evidence on which to base their decisions: whether or not to include extensive reading and/or intensive reading in their curricula, whether to integrate them or keep them separate, which order they should be conducted in, which students should be encouraged to use one or the other for independent learning.

Ultimately, the results of this study, taken in the context of previous literature, show that ER and IR are both important components of a university EFL English curriculum, and that they both contribute to students’ reading fluency and comprehension development in complex and complementary ways. ER is more effective for increasing reading rate, fluency, enjoyment, and confidence at the earlier
stages of university and secondary education, while IR is likely essential for higher level comprehension ability and vocabulary knowledge increases, but in the end the division between the two is not always clear, nor necessary, in pedagogical settings.
REFERENCES


Huffman, J. (2014). Reading rate gains during a one-semester extensive reading course. *Reading in a Foreign Language, 26*(2), 17–33.


Linacre, J. M. (2011b). WINSTEPS Rasch measurement computer program (Version 3.73) [Computer software]. Winsteps.com


APPENDICES
APPENDIX A

SILENT READING RATE TEXTS AND QUESTIONS

[Form A (Time 1), Texts 1-3: Simplified, general-topic narrative fiction]

Text 1

It was early June in 1751 when I shut the door of our house behind me for the last time. All
my life I had lived in the quiet little village of Essendean, in the Lowlands of Scotland. My
father had been the dominie, or schoolteacher. But now that he and my mother were both dead, I
had to leave the house. The new dominie would soon arrive, and he would teach at the school
and live in the dominie’s house. So, although I was only seventeen, there was nowhere for me to
live. And there was also no reason for me to stay in Essendean.

But my heart was beating with excitement as I walked down the road, because in my hand I
carried the letter that my father had given me just before he died. ‘Davie,’ he had said, ‘when I
am dead, take this to the house of Shaws, near Cramond. That’s where I came from, and that’s
where you must go. Put this letter into the hands of Ebenezer Balfour.’

Balfour! The same name as my own! It was the first time I had heard of any of our family
outside Essendean.

So I decided to walk to Cramond, hoping that perhaps this Mr. Balfour would receive me
kindly, and help me to become a rich man one day. With my plaid over my shoulder, I walked
fast up the hill away from the village.

TIME: ____________

1. Why did the narrator (the person telling the story) have to leave his house?
   A. His father lost his job.
   B. His father moved to a different job.
   C. His father died.
   D. His father and mother divorced.

2. Why did the narrator go to Cramond?
   A. He had a family member living there.
   B. He found a job there.
   C. He wanted to go to school there.
   D. He wanted to find a wife and start a family there.

3. What did the narrator hope Mr. Balfour in Cramond would help him do?
   A. get into a good school
   B. find a good wife
   C. make a lot of money
   D. travel the world
Buck did not read the newspapers, so he did not know that trouble was coming for every big dog in California. Men had found gold in the Yukon, and these men needed strong dogs to work in the snow in the north.

Buck lived in Mr. Miller’s big house in the sunny Santa Clara valley. There were large gardens and fruit trees around the house, and a river nearby. In a big place like this, of course, there were many dogs. There were house dogs and farm dogs, but Buck was the chief dog. He was born here, and this was his place. He was four years old and weighed sixty kilograms. He went swimming with Mr. Miller’s sons, and walking with his daughters. He carried the grandchildren on his back, and he sat at Mr. Miller’s feet by the fire during winter.

But this was 1897, and Buck did not know that men and dogs were hurrying northwest to Canada to look for gold. And he did not know that Manuel, one of Mr. Miller’s gardeners, needed money for his large family. One day, when Mr. Miller was out, Manuel and Buck left the garden together. It was just an evening walk, Buck thought. No one saw them go, and only one man saw them arrive at the railway station. This man talked to Manuel, and gave him some money.

TIME: ____________

1. Why did men want to take strong dogs north to Canada?
   A. to help them look for oil
   B. to help them look for food
   C. to help them look for gold
   D. to help them look for a new place to live

2. What did Buck often do with Mr. Miller’s children?
   A. swim and walk
   B. hunt and play
   C. run and work
   D. work and play

3. What did the gardener want to do with Buck?
   A. take him north to find gold
   B. make him work on his farm
   C. let him play with his children
   D. sell him to make money
Minty Cane had known she was a witch, or something like it, since she was small. She had woken at night to see shadowy people moving silently across the floor of her room, or heard invisible feet. She did not talk about these things because she did not think them strange. She had once spoken of a dark visitor to her mother, Kate, but Kate had talked about lights from cars in the street making shadows. In the past year, Minty sometimes heard her father’s voice. And she knew that was strange, because he was dead.

Kate worked all day at the hospital, so for the summer holidays she decided that Minty should go to the village of Belton and stay with an old friend of Kate’s family. Aunt Mary lived in a little stone house opposite Belton House, which was large and beautiful and once belonged to Lord Brownlow. None of the Brownlow family lived there now, and the house was open to visitors from April to October.

‘I always think of Belton as a place where things happen,’ said Kate. ‘When I was small and stayed with Aunt Mary, I had the front bedroom. I could see the top of the church, and bits of the garden at Belton House.’

‘Were there ghosts?’ said Minty.
‘Perhaps. But I never actually saw anything.’

1. Whose voice did Minty sometimes hear?
   A. the voice of her mother
   B. the voice of her father
   C. the voice of her aunt
   D. the voice of an unknown person

2. What was Aunt Mary’s house like?
   A. small and made of stone
   B. normal-sized and made of wood
   C. big and beautiful
   D. like a church

3. What did Kate say about ghosts at Belton House?
   A. She doesn’t believe in ghosts.
   B. She saw a ghost there once.
   C. She thinks there are ghosts because she heard strange sounds.
   D. She thinks there might be ghosts but she never saw one.
Text 4

When I was a student nurse, I perfected the art of giving a great back rub, a skill I had first learned as a nurse’s aide. In my second year of nursing school, whenever we worked evening shift, I gathered my equipment and pushed my cart from room to room, checking IVs, changing rumpled sheets, offering fruit juice with crushed ice, and giving every patient a back rub. For most patients, this quiet ritual brought relief from boredom and loneliness. For me, it became a private interlude during which I could listen to patients tell their stories or admit their fears. While my hands massaged their backs, patients could escape, at least temporarily, the hospital’s stark walls and the day’s long hours.

After taking a patient’s blood pressure and checking the pumps and monitors that hummed monotonously in the background, I closed the curtain, raised the bed so I wouldn’t have to bend over too far, and helped the patient turn on her side. Next I folded back the sheet to expose her back with its tense muscles and skin that was red and impressed with lines from the pressure of the bedclothes.

I opened the bottle of lotion and poured a little of it into my palm, warming it before beginning the massage.

TIME: ____________

1. When did the author first learn how to give a back massage?
   A. when she was a child
   B. when she was a nurse’s aide
   C. when she was a nursing student
   D. when she was a licensed nurse

2. What was the patient’s body position during the back massage?
   A. lying down, facing up
   B. sitting up, facing forward
   C. lying down, on the side
   D. lying down, facing down

3. Why did the author put lotion in her hand before starting the massage?
   A. to measure the amount
   B. to check the smell
   C. to soften her hands
   D. to warm the lotion
Strange things happen at the bedsides of dying patients in hospitals and nursing homes. Odd events become as routine to nurses as the daily tasks of monitoring and medicating patients or staying with them as they die. Nurses understand that something beyond our human comprehension occurs when the last breath is expelled, the one that completes the cycle that began when the patient took in that first lungful of air as a newborn.

This last exhalation releases not only the energy accumulated in the chemical processes of living but also the essence of the soul. The body deflates like a balloon with a pinpoint leak. The patient’s eyes (if they remain open) may reflect the moment of death most apparently. The gaze wanders, first fixing on something nearby, perhaps a beloved spouse or child. The life spark, that glimmer in the eyes we are so used to, dims as if a cloud has passed over. Then the eyes, which until then had focused on earthly sights, look away to focus on some other vision, one that those of us still living cannot share.

Sometimes patients can see their fate coming, like the elderly woman with cancer who refused to open her fists because she was afraid of giving in to death.

1. What did the author say about what happens in the final moment of a patient’s life?
   A. It is easily explained by science.
   B. She learned about it in nursing school.
   C. It is impossible for humans to understand it.
   D. She wants to research it more in the future.

2. What images did the author use to explain the last moment of a patient’s life?
   A. a balloon and a cloud
   B. a juice box and a blanket
   C. a tire and an umbrella
   D. a boat and a storm

3. Why did the elderly woman with cancer keep her fists closed?
   A. She was in intense pain.
   B. She wanted the nurse to stay with her.
   C. She didn’t want to die.
   D. She was angry at the nurse.
On another windy Tuesday in January, I walk into the nursing home, greeted by a rush of warm air. At the same time, I recoil from the musty, antiseptic smell. I’m two years into my first job as a nurse practitioner in private practice with a group of physicians, and today is my day to do nursing home rounds. At the end of the afternoon I’ll go back home, but right now I must go from floor to floor, seeking out our patients who linger here. After all my years in nursing, all my years going in and out of hospitals, I can’t get used to this contrast: the world outside, bright and clear with the bustle of winter birds in the courtyard, and the sudden heaviness inside. Out there is sky and low clouds bringing snow. In here are the subtle odor of urine and the muted voices of the elderly patients, a steady low drone occasionally punctuated by a laugh or a cry.

Emma is first on my list today, an angry eighty-year-old woman who amazes me with verbal attacks that pack more foul language than that of a sailor. I find her in bed, a rhinestone brooch pinned haphazardly to her red nylon bathrobe. As I open the robe and listen to her heart, I imagine the blood circling endlessly through the valves.

TIME: ______________

1. What did the author notice when she walked into the nursing home?
   A. silence and the smell of food
   B. the sound of music and the doctors’ voices
   C. warm air and a musty smell
   D. bright lights and children playing

2. What sounds did the author sometimes hear in the nursing home?
   A. screams of pain
   B. laughter and crying
   C. the television news
   D. mothers scolding their children

3. What did Emma have on her bathrobe?
   A. blood
   B. a nametag
   C. spilled food
   D. jewelry
Text 1

The three girls were awakened by a strange voice shouting, ‘Come on, girls, wake up!’ Miss Evans started pulling the blankets off the beds and shaking the girls. When she arrived at Molly’s bed, she stared at the three girls lying there. ‘Ah yes…the new girls. Get yourselves ready and then get some breakfast.’

One of the bigger girls came over and introduced herself to them. ‘I’m Martha Jones,’ she said, ‘I’ve been here over a year now, so I can guide you around.’ Martha was friendly and full of information, so the new girls from Jigalong trusted her immediately.

A loud bell sounded somewhere. ‘That’s the breakfast bell,’ said Martha. ‘We have to hurry, or the food will get cold.’ They went outside, where they could see the boys’ dormitory and children running to another building. There was a little sunlight, but the air was freezing cold. ‘It’s not so terrible here once you get used to it,’ Martha told them.

Molly remained silent while staring at the other dormitory buildings, which had metal bars across every window and big locks on every door. ‘This place is not pleasant at all,’ she thought. ‘It’s basically a prison where they lock you up at nighttime and release you in the morning.’

1. How did Miss Evans wake up the girls?
   A. with a loud bell
   B. by introducing herself
   C. by shaking them
   D. by singing to them

2. Why did Martha say she could guide the new girls around?
   A. because she was from Jigalong
   B. because she had already been there for a long time
   C. because that was her job
   D. because she trusted the girls

3. What did Molly notice about the other dormitories?
   A. They had metal bars on the windows.
   B. All their curtains were closed.
   C. They were made of brick.
   D. They were smaller than hers.
Text 2

It was an autumn day and the sky was full of dark grey clouds. All day I had ridden on my horse through flat and boring countryside, but at last I saw the end of my journey. There, in front of me, was the House of Usher. Immediately I had a strange feeling of gloom and sadness. I looked up at the old house with its high stone walls and narrow windows. I noticed that the stones of the wall had cracks and holes in them. I felt cold and sick, and I could not think of any happy thoughts to chase away my sadness.

I was planning to spend the next few weeks in this house. Its owner, Roderick Usher, had been a good friend when I was a boy. I had not seen him for many years, but recently he had sent me a letter. He had written that he was ill, both physically and mentally, and that he needed to see me. I was his only friend, he said, and the only person who could understand and help him.

I went up to the front of the house, where a servant took my horse and I entered the large front hall. Another servant guided me silently upstairs. On the walls there were many dark pictures which made me feel nervous. I remembered these pictures from my childhood, but I didn’t remember feeling nervous before.

1. What did the author notice about the wall of the house?
   A. The stones in the wall had cracks.
   B. The paint on the wall was coming off.
   C. The wood of the wall was rotting.
   D. The windows in the wall were broken.

2. Why did the owner of the house send the author a letter?
   A. He was lonely.
   B. He was sick.
   C. He had become poor.
   D. He was planning to move.

3. After he entered the house, what made the author feel nervous?
   A. the servants
   B. his horse
   C. some pictures
   D. strange sounds
At his desk, Wyatt began to look at the data from the hurricane. His office was in the U.S. military base on the island of San Fernandez. From his window he could see the blue waters of the ocean, sparkling in the hot sun. Although he was still only twenty-four, Wyatt was already a smart and successful weather scientist. He did not like hurricanes, especially this new one.

When he was done adding up the numbers, he looked at his map carefully and then looked out the window. Then he went to see Officer Schelling, the chief weather scientist at the base. ‘This hurricane is going to be a very bad one,’ Wyatt told him. ‘The winds could be up to 270 kilometers an hour. It’s moving slowly north now, but it could change direction and come this way. I think we should inform the president.’

‘No, if we look at examples of other hurricanes in the past, it will most likely keep going north. I think we can safely tell the president that it won’t hit the island at all.’

‘I disagree!’ Wyatt said angrily. ‘We don’t know enough about hurricanes and how they move. Don’t you remember five years ago when that hurricane in Florida changed direction seven times in ten days and caused so much damage?’

1. How old is Wyatt?
   A. in his twenties
   B. in his thirties
   C. in his forties
   D. in his fifties

2. Why was Wyatt worried about the hurricane?
   A. It might start going faster.
   B. It might start going slower.
   C. It might keep going the same direction.
   D. It might change direction.

3. What did Wyatt say about the Florida hurricane five years ago?
   A. It was very fast.
   B. It caused much damage.
   C. It didn’t hurt anyone.
   D. It hit a large island.
Unfortunately, nurses hardly ever give back rubs these days. The time that was once spent with patients is now spent arguing with insurance companies about why a patient should stay in the hospital a few more days. Now, instead of those ten minutes of human interaction during which a nurse’s hands might become a conduit of healing, we rush in and give the patient a pre-packaged, pre-warmed towel. Those little bottles of almond lotion that used to be in every patient’s bedside table have now disappeared.

A lot has changed since the days when I wore my blue students’ uniform and patients spent weeks recovering in the hospital. Today I wear street clothes and a lab coat, and only the sickest patients qualify for admission—and even they are discharged after a few days. Patients have to follow recovery schedules that are pre-decided by insurance companies rather than allowing their bodies to heal according to their natural rhythms. We nurses, who used to be masters of the evening back rub, complain about all that has been lost from our noble profession. Some of my nurse friends have even left the profession, and these days fewer and fewer young women and men pursue nursing careers.

1. What do nurses argue with insurance companies about?
   A. whether they can give back rubs or not
   B. what kind of clothes they should wear
   C. how long their patients can stay in the hospital
   D. what kind of medication to give the patient

2. What do nurses offer patients instead of a back rub these days?
   A. a foot massage
   B. a warm towel
   C. almond lotion
   D. better pain medication

3. How long did patients usually spend recovering in the hospital when the author was younger?
   A. one day
   B. a few days
   C. more than a week
   D. more than a month
I received a late-night telephone call from my daughter, who was in the emergency room of a small-town hospital with her ten-month-old daughter. She begged me to come immediately, explaining that her daughter was having severe breathing difficulty and that she didn’t trust the young doctor.

It was past midnight as I drove through a terrible thunderstorm and finally arrived. I rushed into the ER room where my granddaughter was lying motionless. A tall doctor in a blue coat was nervously walking around. My daughter whispered, “He doesn’t have much experience, and a nurse informed me that he doesn’t have the training required to insert a breathing tube into an infant’s throat. He telephoned a pediatrician, but the pediatrician has not answered her phone. There is an ambulance coming from the children’s hospital because the helicopter cannot fly in this thunderstorm, but the ambulance won’t arrive for over an hour.” My granddaughter stared at the ceiling as white steam from the nebulizer condensed into drops on her skin and the raspy sound of her breathing mixed with the hissing sound of the oxygen machine. We all stood there helplessly and listened as the child struggled to breathe.

1. At what time of day did this story happen?
   A. morning  
   B. afternoon  
   C. evening  
   D. night

2. How did the doctor appear?
   A. nervous  
   B. tired  
   C. angry  
   D. sad

3. Why couldn’t the author’s granddaughter be transported by helicopter?
   A. All the helicopters were being used for other patients.  
   B. Using the helicopter was too expensive.  
   C. The helicopter could not fly because of bad weather.  
   D. The helicopter pilot was not answering his phone.
Flora was the final patient on that Friday in August, an urgent addition at the end of the day, and the receptionist’s explanation was short: “The patient is twenty-two weeks pregnant and suspects that her water broke recently, but she didn’t go to the hospital.” It had been a busy week and I was exhausted from standing, annoyed by the noisy children in the waiting room, and tired of trying to educate mothers who continued smoking cigarettes, drinking alcohol, doing illegal drugs, and otherwise damaging themselves and their pregnancies.

Flora sat on the examination table, undressed except for a sheet covering her lower body. She was having trouble sitting still, so I wondered whether she was taking drugs. I approached her in my white laboratory coat with an automatic smile that disguised my inner feelings. “Hello Flora,” I started. “I understand you are worried that your water might have broken recently. Would you explain exactly what happened?”

“My water definitely broke,” she said. “It was a flood of liquid—I guess it was a few days ago, but I cannot remember—maybe last weekend?” I continued by asking whether her baby had been moving very much. “Maybe not so much today,” she replied.

1. According to the author, how did many of the mothers harm themselves?
   A. smoking and drinking
   B. eating unhealthy food
   C. lifting heavy objects at work
   D. not taking their nutritional supplements

2. What facial expression did the author have when she started examining Flora?
   A. a serious stare
   B. a fake smile
   C. an annoyed frown
   D. a warm look

3. What did the patient say happened when her water broke?
   A. It was quite painful.
   B. She didn’t understand what was happening.
   C. A lot of liquid came out.
   D. She started crying.
Text 1

Dr. Gully was certainly the most intelligent man I had ever met. He told me interesting things and made me laugh, and he was so much kinder than my husband, Alexander, or even my own father. I was very happy during our holiday together in Germany. We talked and laughed all day, and sang and danced in the evening. For a while, I thought I might be falling in love with him. But of course he was so much older than me. Some people in the hotel thought he was my grandfather. Some people smiled at us, but others looked at us angrily. We slept together in the same bedroom even though we were not married; most people thought that was very wrong. One evening, he suddenly and surprisingly proposed to me. I didn’t know how to respond, so I smiled and touched his face, but then I slowly walked away. I didn’t want to make him unhappy, but I worried about what people would think if I married a man twice my age. ‘It’s complicated, James,’ I said. ‘My parents already refuse to talk to me because of our relationship. They say I should marry a husband nearer my own age, and perhaps they’re right. We are good friends—isn’t that enough?’ He held my hands, and with a sad expression he looked into my eyes.

TIME: ____________
Minty was lying on her bed listening to music when her aunt called her for lunch. She went downstairs, and just as she sat down at the table, the telephone rang. Aunt Mary went to answer it while Minty started eating an egg sandwich. Suddenly her aunt appeared in the doorway, her face white and her eyes shocked. ‘I don’t quite know how to say... Minty, your mother—there’s been an accident, and your mother was taken to the hospital.’ Minty screamed loudly and then started crying, her worried tears pouring out like a waterfall.

It had been a bad crash on a busy road—Minty’s mother had been caught between two large trucks, and now she was in the emergency room with a serious head injury. Minty was in her room feeling cold and sick when the telephone rang again. A moment later, Aunt Mary brought up a cup of hot, sweet tea. ‘That was John Benson, your mother’s coworker. He’s going to the hospital now to check on the situation there.’

‘I want to go, too!’ Minty said. But Aunt Mary told her she could go tomorrow, when her mother would hopefully be awake. Minty wondered whether her mother would really be awake the next day. ‘Can I go outside?’ Minty said. ‘I need to get up and do something.’

1. What happened when Minty sat down for lunch?
   A. Some music started playing.
   B. Her mother left for work.
   C. Someone knocked on the door.
   D. The telephone rang.

2. What was Minty’s reaction when she heard about the accident?
   A. She started crying.
   B. She became very quiet.
   C. She ran to her bedroom.
   D. She had a surprised expression on her face.

3. What did Aunt Mary hope would happen the next day?
   A. Mr. Benson would go to the hospital.
   B. Mr. Benson would come visit Minty.
   C. Minty’s mother would wake up.
   D. Minty’s mother would come home.

TIME: ____________
It was a long night. After a few hours the big guns stopped, but the women continued to hear various noises. They heard the sounds of bombs exploding, the yells of soldiers in the streets, and the cries of injured and dying men. Julie and Mrs. Weedle were hiding quietly in the hotel upstairs, listening to the awful sounds of the battle outside. Julie stared into the darkness and thought about Wyatt. She wondered where he was and what had happened to him. For a while they heard some soldiers downstairs in the hotel bar, drinking and shouting and breaking things.

When morning came, Wyatt had still not returned. By early afternoon, Julie realized that she and Mrs. Weedle must leave the hotel and try to get up into the mountains by themselves. Mrs. Weedle was scared, but they gathered some food and blankets and hurried out to their truck, which was parked a few hundred meters from the hotel. They were not alone. Hundreds of other people were also in the streets, trying to escape from the fighting in the city.

When Wyatt arrived at the hotel, it was late in the evening. He found a small note attached to the bar counter and read it aloud: ‘We’ve gone east, up into the mountains. Love, Julie.’

TIME: ____________

1. Where were the two women hiding?
   A. in a hotel
   B. in a bar
   C. in a truck
   D. in the mountains

2. What did Julie think about during the night?
   A. She wondered why the soldiers were fighting.
   B. She wondered where Wyatt was.
   C. She wondered whether Mrs. Weedle would survive
   D. She wondered how she would get home.

3. What did Wyatt find on the counter?
   A. an injured soldier
   B. a picture of Mrs. Weedle’s husband
   C. a note from Julie
   D. a broken bottle
The driver kept the ambulance motor running while the doctor, a woman with brown hair and tired brown eyes, kept trying to insert the breathing tube into my granddaughter’s throat; she was finally successful on the third try. Immediately they rushed her to the ambulance, the doctor running beside the stretcher while hand-pumping the breathing bag. The nurse held the IV bags in the air and pulled the portable crash cart beside the stretcher, keeping her eyes on the monitor that measured my granddaughter’s heart activity. The ambulance drove off into the stormy night with my daughter and husband following behind. They later told my daughter she was fortunate they had arrived soon enough—if they had been a few minutes later the girl probably wouldn’t have survived.

I don’t remember the doctor’s name or the nurse’s, but I do remember their quick, confident actions. I imagine them sitting beside my granddaughter in the ambulance, pumping the breathing bag all the way to the Children’s hospital despite their aching hands and the dangerous roads. I know they kept my granddaughter alive until the ICU staff could attach her to the respirator that would support her breathing until her body could recover.

1. At the beginning, what was the doctor trying to do?
   A. put an IV in the patient’s arm
   B. put a breathing tube in the patient’s throat
   C. put a thermometer in the patient’s mouth
   D. put medicine in the patient’s mouth

2. What was the author’s relationship to the patient?
   A. The author was the patient’s nurse.
   B. The author was the patient’s doctor.
   C. The author was the patient’s mother.
   D. The author was the patient’s grandmother.

3. What did the author imagine the doctor and nurse doing in the ambulance?
   A. driving
   B. talking to the patient
   C. pumping the breathing bag
   D. giving IV medication to the patient
Three months later, my father died. I had no siblings and my mother had died 10 years before, so I suddenly found myself alone in the world. To make things worse, five months later I had to have emergency surgery for an unexpected medical problem. After recovering from that operation, I had an abnormal mammogram result, so yet again I had to return to the hospital to have a biopsy performed. Although the biopsy was negative for cancer, my increasingly dark mood turned into a long period of depression. I continued my daily routines at work and home, but I felt strangely separated from my family, my coworkers—indeed, from everything in my life. I felt an increasing distance between myself and my patients, like I couldn’t make a real connection with them anymore. Having too many patients and not enough time, communicating with them through translators, and even the latex gloves I had to wear felt like obstacles to real relationships. At night I continued writing stories for the book I was working on, stories about brave patients overcoming all sorts of illness and hardship. So why couldn’t I overcome the problems in my own life? What was missing from my life, and why did I feel so alone?

1. Why did the author need a biopsy?
   A. She had a strange mammogram result.
   B. She had an unexpected problem during surgery.
   C. She felt increasingly depressed.
   D. She had a history of cancer in her family.

2. Which one of these was an obstacle between the author and her patients?
   A. her medical and health problems
   B. communicating through a translator
   C. her depression
   D. her story writing

3. What was the topic of the stories the author was writing at night?
   A. patients overcoming difficult times
   B. patients with strange medical problems
   C. patients from other countries and cultures
   D. patients with depression
The doctor’s notes informed me that Mrs. Cardiff might die at any moment. As I stood by her bed that morning, her chart clenched in my hands and the sunlight illuminating the room, I wondered if she knew that the end was near. I imagined what I would do to save her if her heart suddenly stopped. I would call for help and then pull her from the bed to the floor, place the heel of my hand on her chest, and pump her heart back to life before anyone else could arrive. Or maybe she would stop breathing and I would be the only one there—I’d grab the breathing mask and place it tightly over her mouth and start squeezing the bag, filling her lungs with life-sustaining oxygen. Of course I really wanted Mrs. Cardiff to simply get better and go home, but I have to admit that a small part of me wanted something dramatic to happen so that I could be the hero.

I introduced myself to Mrs. Cardiff as the new student nurse who would be caring for her during the next few weeks. She smiled and sighed—she was probably used to having different aides and interns and students come and go. I began getting things ready for her bath, but she stopped me and told me that she had sponged herself earlier.

TIME: ____________

1. How did the author know that Mrs. Cardiff was near death?
   A. The doctor told her.
   B. The nurse told her.
   C. She read it in the doctor’s notes.
   D. She read it in the nurse’s notes.

2. What did the author imagine?
   A. what she would do if the patient had a medical emergency
   B. what she would do when she became a real nurse
   C. what she would do when she got off work that evening
   D. what she would do if she was the patient

3. Why did the patient stop the author from giving her a bath?
   A. The patient wanted the author to massage her feet instead.
   B. The patient had already washed herself.
   C. The patient was waiting for her nurse to give her a sponge bath later.
   D. The patient didn’t like the author’s attitude.
Text 1

There were many trees and flowers surrounding the school where Molly and Daisy had been sent, and in a different situation, it might even be considered beautiful. But everything was too strange for Molly. Her heart ached for home, for the red earth and dry salt lakes of the Western Desert, for trees and flowers that were familiar. Then Molly noticed a small building with just one high window. It was near the edge of the school grounds, partly hidden in the trees. She asked her friend, Gracie, about it. Gracie explained, ‘That’s the punishment building. They put you in there if you do something bad. Some girls tried to escape one time, but they got caught and then they had to stay in the punishment building for a whole week, with only bread and water.’

‘Oh!’ Molly said, her eyes wide with fear. But then she asked, ‘Has anyone ever escaped successfully? I mean, has anyone ever ran away without getting caught?’ Gracie shook her head sadly, saying, ‘No, some people have tried, but they always get caught and then punished, so no one tries anymore.’ But Molly knew that she couldn’t stay in this terrible place. She knew they must make a plan to escape and return to their home in Jigalong.

TIME: ____________

1. Molly noticed a building. What did it look like?
   A. It was old and dirty, and the paint was coming off.
   B. It had only one window, and it was hidden in the trees.
   C. It looked freshly painted, but there was no door.
   D. It was red, and there were lots of flowers around it.

2. What was Molly’s reaction when she heard about the girls who were punished?
   A. sadness
   B. anger
   C. compassion
   D. fear

3. Which of these questions did Molly ask Gracie?
   A. Has anyone ever escaped without getting caught?
   B. How long does the punishment last?
   C. Who tries to find you if you run away?
   D. What is the best plan for trying to escape?
Holmes and I went back to the police station later that afternoon. Lester, the police detective, would not stop talking and complaining that we were not any closer to catching the murderer than when we had started. Suddenly, Holmes jumped up and walked out into the street. He told me to stay behind because I was tired, but I went with him anyway. ‘Come on, then, Watson,’ said Holmes. ‘We must hurry. The murderer is nearby—I can feel it.’ Holmes walked faster and faster, and I almost had to run to stay with him. He kept looking from side to side while walking, and suddenly he stopped and stared into the darkness. ‘He will kill twice tonight, Watson. We stopped him from killing last night, so he must kill two women tonight.’ Before I could answer, he was moving again. Suddenly, he pulled me into a dark corner and gestured for me to be silent. I could hear someone coming closer. ‘It’s him,’ whispered Holmes. I tried to stay calm, but my body started shaking with fear. ‘I’ll follow him,’ said Holmes, ‘while you go back to the police station and get Lester!’ I cannot explain why I didn’t follow his instructions, but instead of going to the police station, I decided to follow Holmes.

TIME: ____________

1. What was Lester, the police detective, complaining about?
   A. that they were having trouble finding the criminal
   B. that the murderer had killed two more women
   C. that Holmes would not do what he was told to do
   D. that Holmes was always disappearing and keeping secrets

2. Why did Holmes think the murderer was nearby?
   A. He had received information from Watson.
   B. He heard the murderer’s phone call.
   C. He just had a feeling about it.
   D. He heard the murderer’s footsteps.

3. Why did Holmes think the murderer would kill two people that night?
   A. because the murderer knew that Holmes would find him soon
   B. because the murderer had told him earlier in the day
   C. because the murderer didn’t kill someone the day before
   D. because the murderer was getting very tired
As Jinny did her work on the farm, she kept thinking about the strange woman and her son in the little house on the hill. She wanted to go back up there again today.

But during lunch her father said, ‘You’ll have to make the butter today, Jinny, because your mother is making the bread.’ Then, turning to Jinny’s younger brother, he said, ‘you’ll come with me to repair the fence around the south field. And Jinny, if there’s any daylight after you finish the butter, you can start planting the vegetables. They need to be finished this week.’

Jinny was exhausted from the morning’s work, so hearing all this made her feel quite stressed. Later, she talked to her mother about it. ‘Why do I have to spend all my time working on the farm? This is my summer holiday, but I never get any free time like my friends. They all sleep late and go horse riding or shopping whenever they want!’

Her mother explained, as she had many times before, that the family had chosen to leave the busy life of London. They moved to this farm for a healthier and more peaceful life, but living on a farm was hard work, too. Jinny said, ‘But I didn’t decide to do that, only you and Father did! I never get to decide anything!’

1. Where was Jinny planning to go later that day?
   A. to buy bread and butter
   B. to the house on the hill
   C. shopping with her friends
   D. to ride her horse

2. What did Jinny’s father tell her to do after making butter?
   A. Start fixing the fence.
   B. Start making the bread.
   C. Start planting vegetables.
   D. Start making dinner.

3. What did Jinny’s mother say about why they moved to the farm?
   A. It was because their life in London was too boring.
   B. It was because London was not safe for them.
   C. It was because they could make more money on the farm.
   D. It was because life on the farm was slower and quieter.
Text 4

Focusing only on facts and numbers made me forget about the individual: the woman in the yellow bathrobe who just wanted her feet massaged, or the sad boy who forgot his book on the bedside table. Instead, I became fascinated with the diseases and treatments themselves. I started to enjoy the challenge of treating complex diseases and life-and-death emergencies. I stopped allowing myself to become involved with each patient, and over time this caused me to stop feeling real compassion for them. Soon, I realized that I was relying only on the information in the chart to make impersonal judgments based only on what I knew about the patient’s current physical condition. This allowed me to make decisions quickly and efficiently, without making an emotional connection. But it also meant that I ignored important information, especially about a patient’s beliefs, desires, needs, and emotions.

How can we help students and young nurses to develop and keep the empathy that I had lost? How can we encourage them to stay connected to their patients and not become alienated? Maybe one answer is to try to teach them that each patient, each family member, each nurse, and each physician has a unique story.

1. What did the sad boy forget?
   A. his book
   B. his drink
   C. his jacket
   D. his toy

2. What challenge did the author begin to enjoy?
   A. treating complex illnesses
   B. treating as many patients as possible
   C. making a close, personal connection with each patient
   D. keeping in touch with patients’ family members

3. The author thinks that if we teach nurses that each patient has a unique story, what might happen?
   A. Nurses might be able to save more lives.
   B. Nurses might be able to maintain their own emotional health.
   C. Nurses might be more compassionate.
   D. Nurses might be more efficient.
All nurses experience “shock and awe”, or what could be called “the sacred moments of nursing”. Some of us have become patients and been shocked by illness and also amazed at both the vulnerability and the strength of the human body. Our experience with shock and awe enables us to interact humanely with our patients, impacting their lives while simultaneously enriching our own.

However, shock and awe is also a surprisingly heavy burden. How often have we thought that we couldn’t deal with the suffering anymore, or felt like giving up when confronted with our inability to help a patient as much as we had expected to? And what about the joy and satisfaction that are also encountered in our profession? How do we process and fully experience the mixture of emotions: pain, helplessness, hope, joy, loss, failure, and achievement?

A nurse acquaintance of mine recently took up photography. She says looking through an artistic lens refreshes her spirit and comforts her. Another nurse friend enjoys acting, which she says helps her understand the narratives of her patients’ lives from a different perspective. For me, poetry helps me process my feelings and reflect on the meaningful experiences I have had.

1. According to the author, how does the experience of “shock and awe” help nurses impact their patients’ lives?
   A. It helps them provide the best and most evidence-based treatments.
   B. It helps them recommend enriching hobbies to their patients.
   C. It helps them treat their patients compassionately, as humans.
   D. It helps them follow their treatment plans more rigorously.

2. According to the author, why do nurses sometimes feel like giving up?
   A. They cannot help their patient as much as they wanted to.
   B. They do not have any free time to pursue their hobbies.
   C. They have a heavy burden of continual training and education.
   D. They cannot understand their patients’ perspective and emotions.

3. How does acting benefit one of the author’s nurse friends?
   A. It helps her process her feelings of loss and failure.
   B. It helps her reflect on the “sacred moments” she has experienced.
   C. It helps her improve her self-image as a nurse.
   D. It helps her see her patients’ experiences from a different point of view.
Emma, an 80-year-old dementia patient who is constantly yelling angrily at someone about something, is my first examination today. When I press the stethoscope against her chest, she yells, ‘Be careful, that’s cold!’ As I palpate her abdomen, the skin loose and scarred from six pregnancies, she scolds me, ‘Stop pushing so hard, you’ll injure the baby!’ Perhaps her mind has returned to one of her pregnancies and a physician’s carelessly probing hands. These examinations seem rather meaningless, but they are necessary to follow the government’s regulations. Most of my afternoon will be spent completing endless forms and reports, but I understand that this is an essential role of a nursing home nurse. ‘Emma,’ I call her name but she just responds by shouting terrible words at me. I make another attempt: ‘Emma.’ This time she returns to reality for a moment and really connects with me. ‘Hello, where am I?’ she asks with a sincere expression, reaching toward me with her trembling hand. I attempt to explain her situation, but, as expected, soon Emma disappears into her cloudy memories again, arguing with someone from her past. Emma’s angry voice fades behind me as I proceed to my next patient.

1. Why was the skin on Emma’s stomach so loose?
   A. She had many operations.
   B. She gained a lot of weight.
   C. She gave birth to many children.
   D. She didn’t eat enough nutritious food.

2. According to the author, why are all of the seemingly meaningless examinations needed?
   A. They help her provide the best care for her patients.
   B. They will help her get a better job someday.
   C. They prevent many accidents.
   D. They are required by law.

3. What physical movement did Emma make when she “returned to reality” for a moment?
   A. She waved at the nurse.
   B. She reached her hand toward the nurse.
   C. She tried to stand up.
   D. She looked outside the window.
APPENDIX B

ORAL READING RATE/FLUENCY TEXTS AND QUESTIONS

[Form A (Time 1): Simplified, general-topic narrative fiction]

1. My husband, Charles, and I enjoyed many happy years together. Of course, we argued sometimes, but all couples argue sometimes, don’t they? Charles was a good man and a good lawyer, and I loved him very much, and everyone knows he loved me too.

2. The silver airplane flew south, above the small green islands called the Lesser Antilles. The pilot, Harry Hansen, was staring at the blue sky ahead of him. He was an officer in the United States Navy. He had twelve years of flying experience.

3. At two o’clock in the morning, Jenny walked along the quiet, narrow road and up the hill. When she reached the forest, her father suddenly stepped out and touched her arm. “Oh!” she cried in surprise, but then she remembered to be quiet.

1. According to the first text, what do all married couple do sometimes?
   a. go for walks  b. eat dinner together
   c. argue with each other  d. kiss and hug

2. In the second text, what was the airplane flying over?
   a. some islands  b. some farms
   c. a big city  d. a blue river

3. In the third text, what did Jenny’s father do?
   a. He walked up the hill.  b. He cried out in surprise.
   c. He got angry with her.  d. He touched her arm.
[Form A (Time 1): Authentic, nursing-related narrative non-fiction]

1. When I became a nursing student, all of that began to change, gradually but surely. I learned a new language of scientific and clinical terminology, a language that identified me as part of an elite group and separated me from my patients.

2. Linda returned to the break room and announced that she was needed in the OR, and Rick was rushing out to do an ultrasound on a woman who was pregnant with twins. It became clear that the girl with the tattoos would be my responsibility.

3. An hour later, when my granddaughter’s airway had narrowed to less than a quarter-inch, the team from the Children’s Hospital finally arrived. As the team started working, one of the nurses hugged my daughter and explained what was happening.

1. In the first text, what was the result of the “new language” learned by the nursing student?
   a. It helped her work faster.       b. It kept her separate from her patients.
   c. It required a lot of studying.  d. It helped her save more lives.

2. In the second text, what kind of patient was Rick going to do an ultrasound on?
   a. a pregnant woman       b. an elderly cancer patient
   c. a man with abdominal pain d. a little girl getting ready for an operation

3. In the third text, what medical emergency did the granddaughter have?
   a. an injured spinal cord       b. a high fever
   c. breathing trouble           d. bone cancer
[Form B (Time 2): *Simplified, general-topic narrative fiction*]

1. Mrs. Hollins pretended to feel sorry for Becky. ‘That poor girl, she doesn’t have any newspapers or radio or television in her house,’ she explained. ‘Her parents grow their own food and make their own clothes. They never buy anything new!’

2. Alan moved his wheelchair closer to his daughter and held her hands tightly. He told her he was sorry and said, ‘Anna hated everyone, but she doesn’t matter now, Jane. You’re alive, and that’s all that matters. Nothing else is important now.’

3. They entered the church, which was cold and dark after the brightness outside. After looking around, they went out into the sun and saw a high gate. A beautiful garden was on the other side, and Minty suddenly felt like time had stopped.

1. What did Mrs. Hollins say about Becky’s family?
   a. They don’t have a television.  
   b. They don’t have food.  
   c. They don’t have clothes.  
   d. They don’t have a house.

2. What does Alan think is important now?
   a. He has a wheelchair.  
   b. He is sorry.  
   c. Anna hated everyone.  
   d. His daughter is alive.

3. What was on the other side of the gate?
   a. a church  
   b. a garden  
   c. a house  
   d. a road
[Form B (Time 2): Authentic, nursing-related narrative non-fiction]

1. Because I’m unable to guide my patient through her grief and guilt, I contact the women’s center to ask when the special counselor will be available. Maybe she can help my patient deal with the nightmares that have been keeping her awake at night.

2. Twenty years earlier, Shellie had been diagnosed with breast cancer, but she had survived. Now her heart was unstable, and she needed a variety of IV medication to prevent her from going into heart failure, which would likely be fatal.

3. When we first wear our student uniforms and gather around a suffering patient, we feel anxious and nervous. We also feel an intense emotion which is based not only on the patient’s situation but also our own experience with illness and injury.

1. What problem is caused by the patient’s nightmares?
   a. They make the patient sad.
   b. They make it difficult for the patient to work.
   c. They keep the patient awake.
   d. They cause the patient to be sick.

2. Why did the patient need medicine?
   a. Her heart was not stable.
   b. She had breast cancer.
   c. She had diabetes.
   d. Her asthma was getting worse.

3. What causes the strong emotion that new nursing students feel?
   a. wearing their uniforms for the first time
   b. the patient’s experience with nurses
   c. their own experience with sickness
   d. the training they received from their teachers
1. ‘See all those rabbit holes? Maybe we can dig out one of those and make it big enough to sleep in,’ said Mindy. They found an empty hole and started digging and digging, until finally there was enough room for the three tired girls to fit inside.

2. When I was a child, my family had many animals in the house, so I used to spend all day playing with them and caring for them. As I grew older, my love for animals also grew. I found that they were friendlier and more honest than most humans.

3. The two men slept in the narrow cave entrance. Alex was trapped at the back of the cave, helpless and hopeless. He finally fell asleep sometime after midnight. He dreamed that he was in a play, but everyone in the audience was laughing at him.

1. What did the girls want to do in the rabbit hole?  
   a. play  
   b. eat  
   c. sleep  
   d. hide

2. What happened when the narrator (“I”) grew older?  
   a. She didn’t like animals anymore.  
   b. She stopped thinking about animals.  
   c. She liked animals even more than before.  
   d. She began to like humans more than animals.

3. Where did Alex sleep?  
   a. at the back of the cave  
   b. near the cave entrance  
   c. on the stage  
   d. behind the stage
[Form C (Time 3): Authentic, nursing-related narrative non-fiction]

1. When I get sick, I pretend that I am one of my own patients. I consider my medical history and take a physical inventory. I pay attention to whether I have any serious symptoms such as shortness of breath, chest pain, or unrelenting abdominal pain.

2. Nurses are very familiar with shock and amazement; we certainly have an enhanced understanding of how illness and recovery can take people by surprise. Perhaps we’ve been amazed by a patient’s bravery or stunned by the deadly progression of a virus.

3. Caregivers and patients should be free to scream and cry as loudly and for as long as necessary. I believe that grief, when fully expressed, will eventually transform into a stronger and deeper understanding of the fragility and beauty of life.

1. In the first text, what does the narrator (“I”) say she does when she becomes ill?
   a. She continues taking care of her patients.
   b. She asks another nurse to check her symptoms.
   c. She pretends she feels fine.
   d. She imagines she is her own patient.

2. In the second text, which of these did the author say a nurse might be amazed by?
   a. a brave patient  
   b. a patient nurse  
   c. a skilled doctor  
   d. a weak virus

3. What does the author think grief can change into?
   a. bravery  
   b. happiness  
   c. understanding  
   d. intelligence
[Form D (Time 4): Simplified, general-topic narrative fiction]

1. Wyatt sped along the muddy roads while Dawson sat beside him, thinking about the last three days. They had lived through a deadly battle and a terrible storm. Dawson felt he had learned a lot about himself, and he was no longer afraid of death.

2. Janet walked slowly away from the hospital, feeling lonely. She was thinking about her mom, wishing she was still alive and sitting next to her dad in the hospital. ‘I don’t want all these problems,’ she thought. ‘Why does Dad need another woman?’

3. Mary went into the garden and walked down six steps to a path with a row of tall trees on both sides of it. Suddenly, she felt sure that someone was watching her. She looked back at the big house, but there was no one in any of the windows.

1. What had the two men survived in the past few days?
   a. a terrible sickness and cold weather  
   b. a war and a big storm  
   c. a dangerous fire and a robbery  
   d. prison and an earthquake

2. What did Janet wish?
   a. that her mother was not dead  
   b. that her mother had not left  
   c. that her mother was not sick  
   d. that her mother was not so strict

3. What did Mary first do when she entered the garden?
   a. started watching someone  
   b. climbed a tree  
   c. ran to the center  
   d. walked down some stairs
1. “This is an uncomfortable exam,” I explain to the student, “and it’s important to check everything carefully. But remember that the patient can’t see what’s happening, so you should always explain what you’re doing. This will reduce her anxiety.”

2. When I returned to St. Joe’s the next week, I discovered that Martha had died the day after I washed her feet. The nurses and doctors had pounded her chest and forced oxygen into her lungs and used every possible medication, but nothing worked.

3. Since then, I have greater compassion toward women who keep seeking medical attention even after all the test results have come back negative. I carefully assess whether they have any past experiences that caused depression, grief, or guilt.

1. Why should the student always explain what he is doing?
   a. so the patient can learn how to do it herself next time
   b. so the patient will feel less pain
   c. so the patient’s body will not be damaged
   d. so the patient will calm down

2. When did Martha die?
   a. a few hours after having her feet washed
   b. one day after having her feet washed
   c. a few days after having her feet washed
   d. one week after having her feet washed

3. What does the author (“I”) think might be caused by the past experiences of her female patients?
   a. feelings of depression and guilt
   b. injuries and physical pain
   c. infectious diseases and cancer
   d. incorrect results from blood tests
APPENDIX C

READING COMPREHENSION TEXTS AND QUESTIONS

[Form A (Time 1): Simplified, general-topic narrative fiction]

[Text 1]
Jigalong was a long way from anywhere, right out in the desert. In the early years, it was just a small depot for the white men who worked on the rabbit-proof fence. They rode on horses up and down the fence, clearing away branches of trees and dead animals. They also mended holes made by fire or storms or kangaroos.

There was also a government store at Jigalong. It gave out free food, clothing, and blankets to the local people. In the 1930s, the Mardu people decided that Jigalong would be a good place for them to stay. It became a new home, a ‘sitting down place’ for their people.

Molly grew up at Jigalong, among her people, the Mardu. Her mother, Maude, was a young Mardu woman who used to work as a domestic servant for the superintendent at the depot. Her father was Thomas Craig, an Englishman. He worked as an inspector of the rabbit-proof fence. He called the baby Molly, like his sister, and often brought her presents of clothing and dresses. But after a few years his work on the fence finished, and he moved on.

Molly was a pretty child. But while she was still quite young, she already knew she was different. Her skin colour was not as dark as the other Mardu children’s.

1. Why did white people first come to Jigalong?
   A. to hunt animals
   B. to build a railroad
   C. to take care of a fence
   D. to raise horses

2. Why did Thomas Craig leave Jigalong?
   A. His daughter’s skin colour was strange.
   B. He finished his work there.
   C. His wife left him.
   D. He was seriously injured.

3. Why was Molly’s skin colour different?
   A. She was born with a rare disease.
   B. She played outside too much as a child.
   C. She came from a different country.
   D. Her mother and father were of different races.

4. Why do you think the Mardu people decided to stay in Jigalong?
   A. They could get supplies and jobs more easily there.
   B. The climate and weather were perfect for them.
   C. Their original home had been destroyed.
   D. They wanted to be educated by the white people.
‘No school today! Wonderful! Wonderful!’
Alexis woke up suddenly, to find his noisy young brother shouting in his ear.
‘Do you realize what day it is?’ asked Theo.
Alexis sat up quickly. ‘Oh! The Theatre Festival!’
He jumped up and looked out of the window. The sky over Athens had turned from dark blue to grey, and people all over the city were waking up.
Soon all the family were dressed in their best clothes and ready to leave. Alexis, in a new white tunic, went in front with his father, Leon, and his brother. Behind them came his mother and Nico, his older sister. Both wore veils over their faces—on Father’s orders. He thought that a woman’s place was in the home and did not really like them going out into the city.
As they walked through the narrow streets of the city in the early golden sunlight, Alexis thought how lovely Athens looked. Hundreds of people were already going into the theatre—Athenians, foreigners from the western islands, Cretans, Egyptians…People from every country came to Athens for the Theatre Festival. ‘And why not?’ Alexis thought. ‘Athens is the finest city-state in Greece.’
Inside the theatre, Alexis’ family joined the crowd climbing up the terraces to find seats.

5. What did Alexis’s father think about women?
   A. They should walk next to their husbands.
   B. They should work harder than men.
   C. They should usually stay home.
   D. They should not talk much

6. What did Alexis think about Athens?
   A. It was the best city in Greece.
   B. It was the best city in the world.
   C. It was the most boring city in Greece.
   D. There were too many people there.

7. What was the relationship between Theo and Nico?
   A. son and mother
   B. cousins
   C. brother and sister
   D. husband and wife

8. When do you think this story is happening?
   A. a long time ago
   B. a few years ago
   C. in the present day
   D. in the future
‘Look, here she comes now!’ Jane Cole said.

The two Americans looked along the street, which was crowded with people. Soldiers were riding towards them on horseback, and behind them came a golden carriage pulled by six black horses.

‘My father is the coachman—the man driving the horses,’ Jane said.

The American woman said: ‘It must be fantastic to see your father driving the Queen! Quick, Harry, film it with the video camera!’

‘I am filming, but we’re still too far away’ her husband said. ‘Is it possible to get a little nearer, Jane?’

‘We can try if you follow me,’ Jane said. She took them closer to the entrance to Parliament. ‘This is where the coach will stop and the Queen will get out. Then she’ll go upstairs to open Parliament for this year.’

‘Didn’t someone put a bomb under your Parliament once?’ the American man asked. ‘I learned about that at school. Guy…something?’

‘Guy Fawkes,’ Jane said. ‘In 1605. He tried to blow up Parliament, that’s right. But don’t worry, there’s no Guy Fawkes here today.’

She smiled at the Americans. She was a student, and this was her part-time job—to show tourists around London. She felt proud to show them her father, driving the Queen on a wonderful day like this.

9. What is Jane’s father’s job?
   A. protecting the Queen
   B. driving the Queen’s carriage
   C. working in Parliament
   D. guiding tourists

10. Where will the Queen get out of the coach?
    A. in front of Parliament
    B. at the royal castle
    C. in front of the tourists
    D. at the royal church

11. When did someone put a bomb under Parliament?
    A. a few weeks ago
    B. many months ago
    C. a few years ago
    D. many years ago

12. Why do you think the Americans are in London?
    A. They wanted to meet Jane and her father.
    B. They are tourists on vacation.
    C. They are planning to live there permanently.
    D. They are college students.
On a sunny, late February day in 1970, I knelt for the first time ever to wash a patient’s feet. She was a fifty-five-year-old woman in heart failure who would die less than a week later. I was in my first year of nursing school.

I washed Mrs. Cardiff’s feet at Saint Joseph’s Hospital, a community institution commandeered by Sister Mary Margaret, a tiny, middle-aged nun in voluminous black who moved silently through the wards, always appearing just in time to correct a near error in patient care or to catch a sloppily tightened draw sheet. At St. Joe’s—what we students and many of the nurses called the hospital—most patients lingered in wards, four or five beds along one wall, and equal number against the other. Green fabric curtains on squeaky metal rings surrounded the beds, allowing only sounds and smells and the silent vibrations of sorrow to penetrate their borders. When all the curtains were opened, as they were at night and for meals, the two rows of patients faced each other like chess pieces. Upright and fragile in their johnny gowns or lying on their scratchy bleached sheets, every patient knew what was wrong with, and what was happening to, every other patient.

13. What did Mary Margaret often do while walking around the hospital?
   A. She cleaned things and picked up garbage.
   B. She took notes about what she saw.
   C. She fixed the nurses’ mistakes.
   D. She washed patients’ feet.

14. How many beds were in each ward?
   A. 2
   B. 4
   C. 6
   D. 8

15. What did the nurses do before meals?
   A. They cooked the food.
   B. They washed the patients’ feet.
   C. They opened the curtains around the beds.
   D. They played chess with the patients.

16. How does the author probably feel about Mary Margaret?
   A. The author thinks Mary is strict.
   B. The author thinks Mary is sick.
   C. The author thinks Mary is kind.
   D. The author thinks Mary is weak.
When I see a patient waiting inside an examination room, I wonder who he or she is, how we are different, and how we are alike. Working in women’s health, I have a professional advantage, which is that I too live in a woman’s body and so am subject to that body’s strange and wonderful whims. Like many of my patients, I’ve given birth; I’ve offered my breasts, laden with milk; I’ve bled too much and too long; I’ve had lumps in my breast and operations to remove them; I’ve felt the excruciating pelvic pains, like a knife in the vagina, that men will never experience. Because I am physically similar to my patients, I can empathize with them and better understand their stories and their symptoms, and they believe I’m less likely to judge them. When I walk into a room, patients sometimes say, “Oh good, I was hoping I’d get a woman.”

Still, there is always an invisible boundary between patient and caregiver, a boundary constructed and maintained by both, for the safety and well-being of both. Our patients want us to be like them, yet they also want us to be strong when they are vulnerable, healing when they are hurting, kind when they are overwhelmed, and knowledgeable when they are questioning.

17. Why does the author feel that she has an advantage in her field, women’s health?
   A. because she has treated many women
   B. because she has a daughter
   C. because she is female
   D. because she has studied women’s health issues

18. What kind of pain does the author say that men will never experience?
   A. pain in the pelvic area
   B. pain in the breasts
   C. pain in the lower back
   D. pain in the ankles and feet

19. Why do patients feel that the author is unlikely to judge them?
   A. because she is a highly professional nurse
   B. because she shows her emotions easily
   C. because her body is similar to theirs
   D. because her technical skills are excellent

20. How do you think the author feels about the boundary between nurse and patient?
   A. It prevents nurses from providing adequate care.
   B. It ensures that nurses have enough free time.
   C. It benefits both nurses and patients.
   D. It prevents patients from revealing their true feelings.
A young woman, who I will call Susan, had been coming to our hospital clinic for years complaining of a persistent pelvic ache that interfered with her intimate relations with her husband. She had undergone cultures for infection, Paps, ultrasounds, and even an exploratory laparoscopy that had discovered no adhesions, nor cysts, nor anything that would explain this troublesome pain.

This particular day, she sat slouched on the exam table with a markedly melancholy demeanor, and when I sat down she fixed her gaze on me.

“Sorry I was late, but I wasn’t sure whether I should even bother to come” she offered. “The secretary said that you’re still having pelvic pain,” I prompted. “Yes, but everything always comes back negative, so my husband thinks it’s all in my mind.”

I hesitated before asking her, “Do you believe it’s all in your mind? What do you think causes your pain?”

My question released a torrent of emotion, and Susan began to cry, lifting the sheet and pressing it to her eyes. I thought I could guess the reason for her sudden outburst: maybe she’d been raped or abused, and my question had somehow allowed her to acknowledge the pain of those memories. But Susan’s story surprised me.

21. Why was Susan late for her appointment?
   A. She knew the tests would be painful.
   B. The previous tests had negative results.
   C. Her husband abused her.
   D. The nurses usually didn’t try to help her.

22. What did Susan’s husband think was the cause of Susan’s pain?
   A. a cyst
   B. a psychological condition
   C. sexual abuse
   D. no guess

23. What was the cause of Susan’s pain?
   A. a cyst
   B. a psychological condition
   C. sexual abuse
   D. unknown

24. According to the text, which one seems like the least common test for a patient like Susan?
   A. cultures for infection
   B. a Pap smear
   C. an ultrasound
   D. a laparoscopy
I remember the cold morning when I first encountered Florence. There was snow on the nearby mountains, but the sun was shining. Florence’s mother had accompanied her to my hospital. Florence didn’t have a physical illness, but she was very unhappy. She had run away from her husband, Alexander Ricardo, but her father had told her to return to him.

‘I refuse to live there!’ she cried. ‘I’m never returning to Alexander!’ I looked at her carefully and noticed a dark bruise on her face. She was a beautiful woman, twenty-five years old, and she had been married for six years. I asked whether her husband had physically abused her. ‘Yes, he frequently hits me. I hate him!’

‘But you’re married, Florence,’ her mother said. ‘You must live with your husband like a good English wife.’

‘No, Mother, you don’t understand!’ Florence screamed. ‘When he’s drunk, he becomes dangerous and hurts me. I’ll kill myself if you force me to go back!’ She stood up and punched the window with her fist.

‘This young woman is emotionally unstable,’ I told her mother. ‘She needs a quiet, safe place where she can rest and be calm. We have a small house located in the hospital gardens where she can stay until she recovers.’

1. What did Florence’s mother think Florence should do?
   A. run away from Alexander
   B. return to Alexander’s house
   C. live with her parents for a while
   D. stay at the hospital for a while

2. In what situation does Florence’s husband often hit her?
   A. when he is angry at her
   B. when he is stressed from work
   C. when he has had too much alcohol
   D. when she visits her parents

3. When the doctor (narrator) saw the bruise on Florence’s face, what did he think?
   A. that her husband might have hurt her
   B. that her father might have hurt her
   C. that she might have hurt herself
   D. that she might have fallen down

4. What do you think is the biggest reason the doctor began to worry about Florence’s emotional condition?
   A. because her mother and father did not support her enough
   B. because she ran away from her husband
   C. because she didn’t obey her mother
   D. because she hit the window
No school today! Wonderful! Wonderful!

Alexis woke up suddenly, to find his noisy young brother shouting in his ear.
‘Do you realize what day it is?’ asked Theo.
Alexis sat up quickly. ‘Oh! The Theatre Festival!’

He jumped up and looked out of the window. The sky over Athens had turned from dark blue to grey, and people all over the city were waking up.

Soon all the family were dressed in their best clothes and ready to leave. Alexis, in a new white tunic, went in front with his father, Leon, and his brother. Behind them came his mother and Nico, his older sister. Both wore veils over their faces—on Father’s orders. He thought that a woman’s place was in the home and did not really like them going out into the city.

As they walked through the narrow streets of the city in the early golden sunlight, Alexis thought how lovely Athens looked. Hundreds of people were already going into the theatre—Athenians, foreigners from the western islands, Cretans, Egyptians…People from every country came to Athens for the Theatre Festival. ‘And why not?’ Alexis thought. ‘Athens is the finest city-state in Greece.’

Inside the theatre, Alexis’ family joined the crowd climbing up the terraces to find seats.

5. What did Alexis’s father think about women?
A. They should walk next to their husbands.
B. They should work harder than men.
C. They should usually stay home.
D. They should not talk much

6. What did Alexis think about Athens?
A. It was the best city in Greece.
B. It was the best city in the world.
C. It was the most boring city in Greece.
D. There were too many people there.

7. What was the relationship between Theo and Nico?
A. son and mother
B. cousins
C. brother and sister
D. husband and wife

8. When do you think this story is happening?
A. a long time ago
B. a few years ago
C. in the present day
D. in the future
Sherlock Holmes became a detective in 1877. In the beginning, he enjoyed every case, but soon he began to find the work easy. Ten years later he was famous, but he was unhappy and bored.

‘The modern criminal is so painfully slow and stupid,’ he often said. ‘I need an interesting case, Watson, one which will require my full intelligence. Are there no clever thieves or murderers in the world these days?’

One day in 1888 a letter arrived from Scotland Yard. Upon opening it, Holmes laughed and jumped to his feet. ‘Inspector Lestrade wishes to consult with me,’ he exclaimed. ‘The police need my help, Watson. Surely you know that someone is murdering women in Whitechapel?’

‘Of course,’ I replied. ‘The newspapers are reporting on it constantly. Three women are dead, and the police cannot find the killer. But murders are common on the streets of Whitechapel, especially for women of that kind. Why are you interested in their deaths?’

‘Actually this is an unusual case, Watson,’ Holmes cried. ‘I have been investigating it. I knew that the police would need my help. It is mysterious that the women who died were poor, and neither young nor beautiful,’ he said. ‘So they were not killed for money or for love.’

9. Why was Holmes unhappy even though he was famous?
   A. because he had too much work to do
   B. because finding the criminals was not challenging
   C. because the police refused to work with him
   D. because he was not as smart as before

10. What did Holmes think about the reason the women were murdered?
    A. It was because they were beautiful.
    B. It was because they were young.
    C. It was not because they were female.
    D. It was not because they were rich.

11. Why did the police send a letter to Holmes?
    A. They knew Holmes was bored and unhappy.
    B. They knew Holmes lived near Whitechapel.
    C. They thought Holmes knew the murderer.
    D. They had trouble finding the murderer.

12. How do you think Holmes felt when the police requested his help?
    A. happy
    B. bored
    C. annoyed
    D. shy
On a sunny, late February day in 1970, I knelt for the first time ever to wash a patient’s feet. She was a fifty-five-year-old woman in heart failure who would die less than a week later. I was in my first year of nursing school.

I washed Mrs. Cardiff’s feet at Saint Joseph’s Hospital, a community institution commandeered by Sister Mary Margaret, a tiny, middle-aged nun in voluminous black who moved silently through the wards, always appearing just in time to correct a near error in patient care or to catch a sloppily tightened draw sheet. At St. Joe’s—what we students and many of the nurses called the hospital—most patients lingered in wards, four or five beds along one wall, and equal number against the other. Green fabric curtains on squeaky metal rings surrounded the beds, allowing only sounds and smells and the silent vibrations of sorrow to penetrate their borders. When all the curtains were opened, as they were at night and for meals, the two rows of patients faced each other like chess pieces. Upright and fragile in their johnny gowns or lying on their scratchy bleached sheets, every patient knew what was wrong with, and what was happening to, every other patient.

13. What did Mary Margaret often do while walking around the hospital?
   A. She cleaned things and picked up garbage.
   B. She took notes about what she saw.
   C. She fixed the nurses’ mistakes.
   D. She washed patients’ feet.

14. How many beds were in each ward?
   A. 2
   B. 4
   C. 6
   D. 8

15. What did the nurses do before meals?
   A. They cooked the food.
   B. They washed the patients’ feet.
   C. They opened the curtains around the beds.
   D. They played chess with the patients.

16. How does the author probably feel about Mary Margaret?
   A. The author thinks Mary is strict.
   B. The author thinks Mary is sick.
   C. The author thinks Mary is kind.
   D. The author thinks Mary is weak.
I sometimes wonder why the idea of “story” is so important to me. Perhaps it is because we nurses, who often come to our profession filled with compassion and a desire to help people, can gradually stop seeing the individual patient. We sometimes forget the importance of empathy, which is difficult to sustain but essential to providing excellent health care. It isn’t easy to go from bedside to bedside and continually show the emotional response that allows us to best help each patient.

We see patients at unusual and stressful moments in their lives. We often know them only from the visual and physical information they present—symptoms, test results, and chief complaints. But stories of caregiving can help us to sharpen our empathy. They help us to be better, more compassionate, and even more emotionally fulfilled caregivers.

If we can write, talk, or read about our experiences with patients, those stories might transform us, inspire us, and give us the energy we need to reach out to our patients again and again. Patients’ stories are more than just the case histories written in their charts. We need to remember that they have rich, complicated, changing, multi-layered lives.

17. What do nurses sometimes stop doing after being a nurse for a long time?
   A. being careful about giving the correct medicine to each patient
   B. realizing that patients are under stress
   C. making an emotional connection to each patient
   D. discussing their experiences with other nurses

18. How can telling stories about patients help nurses?
   A. It inspires them to be kinder to individual patients.
   B. It helps them form closer relationships with each other.
   C. It teaches them how to recognize patients’ symptoms and complaints.
   D. It helps them do their jobs without becoming sad.

19. Why does the author think it is important to know more than the visual and physical information about a patient?
   A. because nurses show too much empathy for each patient
   B. because patients are under a lot of stress when they are at a hospital
   C. because nurses need to move on to the next patient quickly
   D. because patients have complex and dynamic lives

20. How does this author probably feel about assigning more patients to each nurse during a shift?
   A. She would agree, because she wants to help as many people as possible.
   B. She would agree, because she wants the hospital to profit more.
   C. She would disagree, because nurses need more time off.
   D. She would disagree, because nurses should spend more time with each patient.
When I was one year old, my mother was diagnosed with tuberculosis and my father, who had just returned from the war, was suffering from terrible nightmares—what we would today call post-traumatic stress disorder (PTSD). My parents decided it would be best for me to move away, so they sent me to live with their best friends, a couple with two children of their own, a boy and a girl.

As the months went by, I seemed to forget about my parents at times. But I was also never quite at home with my new family. After I had lived with this family almost as long as I had lived with my own parents, my mother and father came to reclaim me. My mother’s health had improved, and my father seemed to be coping with his nightmares. My parents told me I readjusted quickly and my childhood was mostly happy and constant after that. However, for a long time, I couldn’t tolerate even the slightest separation. If my mother turned the corner in a grocery store, out of sight for only a moment, I would fall to the floor screaming, thinking she was gone forever.

On my father’s advice, I began to write stories and poems as a way of re-experiencing my childhood and trying to make sense of my place in the world.

21. Why was the author’s father having nightmares?
A. He had too much stress at work.
B. He had seen a person killed when he was a child.
C. He had been a soldier in a war.
D. He had lost his job.

22. After living with a different family, why did the author return to her own family?
A. Her parents recovered from their health problems.
B. Her parents didn’t like the other family.
C. She didn’t feel comfortable with the other family.
D. She wanted to live with her own parents.

23. Why did the author scream when her mother turned the corner in a grocery store?
A. because she had nightmares about her mother dying
B. because her father had been having nightmares
C. because she had lived away from her parents for so long
D. because her mother was very sick

24. Why do you think the author needed to write stories to understand her place in the world?
A. because she didn’t know what kind of career to pursue
B. because she had some confusion about her identity during childhood
C. because her parents struggled with physical and mental illness
D. because she was abused by the family she lived with for a while
I began to ask people on the road how to get to the Balfour house. Their reactions worried me. Some seemed surprised, some angry, and some afraid when I spoke the name of Ebenezer Balfour.

I continued on until I met a dark, wild-looking woman coming towards me. When I asked her about the house, she took me to the top of a hill. She pointed to a large building at the bottom of the next valley with no other houses around it. The fields around it were green and the farmland looked good, but the house itself looked unfinished and empty. Part of its roof was missing. There was no road to it, and no smoke coming from any of its chimneys. “There it is—the house of Balfour. Black is the heart of Ebenezer Balfour! You can tell him that I hope to see him die, and his house fall down around him!” The woman turned and ran away, but her words left me shocked and afraid.

When sun went down, I saw a little smoke coming out of the chimney. This made me feel a little more hopeful, so I walked down the hill to the front door of the house. I could hear a man’s voice mumbling somewhere in the dark house, so I bravely lifted my hand to knock on the wooden door. The house was suddenly silent and there was no reply.

1. What made the people on the road get shocked and upset?
   A. the author’s appearance
   B. the wild woman’s words
   C. seeing the Balfour house
   D. hearing the Balfour name

2. When did the author see smoke coming from the Balfour house?
   A. before he saw the house itself
   B. when he first saw the house from the hilltop
   C. when it became nighttime
   D. when he heard a voice inside

3. After hearing the wild woman’s words, where did the author wait?
   A. at the woman’s house
   B. at the top of the hill
   C. at his own house
   D. in the valley

4. Why do you think the man in the house didn’t reply to the knock?
   A. He didn’t want any visitors.
   B. He wanted to prepare his room first.
   C. He couldn’t hear the knock.
   D. He was busy working.
Buck growled at the strange man, but he was surprised when the rope was pulled hard around his neck. He tried to attack, but the man knocked him down hard. For a few moments he was paralyzed, so the man quickly put him in the train. When Buck woke up, the man was watching him carefully, but Buck lunged at him quickly and bit his hand hard; then the rope was pulled again and Buck let go. That evening, the man took Buck to a bar in San Francisco, where the barman looked at the man’s hand and clothes covered in blood. ‘How much are you getting for this?’ he inquired. The man said he would receive a hundred and fifty dollars, and the barman remarked, ‘That’s a good price for a quality dog like him.’ The two men pushed Buck into a wooden box, where spent the night sleepless and confused. His neck ached, and he could not understand what was happening and what these men intended to do with him. The next day Buck was carried to the railway station and put on a train. For two whole days the train travelled north, and Buck was not given anything to eat or drink. The men on the train pushed sticks at him through the holes in the box. Buck became frustrated and started biting at anything that moved.

5. What did the man do when Buck bit his hand?
   A. He hit Buck hard.
   B. He put Buck on a train.
   C. He called the barman for help.
   D. He pulled the rope.

6. What did the barman say when he heard how much the man was being paid?
   A. He thought it was too much money.
   B. He thought it was a reasonable amount.
   C. He thought it was too cheap.
   D. He kept quiet about his opinion.

7. Why was Buck’s neck hurting while he was in the box at the bar?
   A. from having his rope pulled
   B. from being pushed into the small box
   C. from riding on the train for a long time
   D. from not having any water to drink

8. Why do you think Buck is travelling north?
   A. His original owner is there.
   B. The person who paid the first man is there.
   C. He will be set free into the forest there.
   D. There were no trains going in other directions.
Sherlock Holmes became a detective in 1877. In the beginning, he enjoyed every case, but soon he began to find the work easy. Ten years later he was famous, but he was unhappy and bored.

‘The modern criminal is so painfully slow and stupid,’ he often said. ‘I need an interesting case, Watson, one which will require my full intelligence. Are there no clever thieves or murderers in the world these days?’

One day in 1888 a letter arrived from Scotland Yard. Upon opening it, Holmes laughed and jumped to his feet. ‘Inspector Lestrade wishes to consult with me,’ he exclaimed. ‘The police need my help, Watson. Surely you know that someone is murdering women in Whitechapel?’

‘Of course,’ I replied. ‘The newspapers are reporting on it constantly. Three women are dead, and the police cannot find the killer. But murders are common on the streets of Whitechapel, especially for women of that kind. Why are you interested in their deaths?’

‘Actually this is an unusual case, Watson,’ Holmes cried. ‘I have been investigating it. I knew that the police would need my help. It is mysterious that the women who died were poor, and neither young nor beautiful,’ he said. ‘So they were not killed for money or for love.’

9. Why was Holmes unhappy even though he was famous?
   A. because he was not as smart as before
   B. because he had too much work to do
   C. because finding the criminals was not challenging
   D. because the police refused to work with him

10. What did Holmes think about the reason the women were murdered?
    A. It was not because they were female.
    B. It was not because they were rich.
    C. It was because they were beautiful.
    D. It was because they were young.

11. Why did the police send a letter to Holmes?
    A. They knew Holmes lived near Whitechapel.
    B. They thought Holmes knew the murderer.
    C. They had trouble finding the murderer.
    D. They knew Holmes was bored and unhappy.

12. How do you think Holmes felt when the police requested his help?
    A. shy
    B. annoyed
    C. bored
    D. happy
At the hospital where I am employed, I am constantly surrounded by people crying: a woman is mourning the death of her newborn baby, an elderly widower is holding his wife’s belongings, a mother watches helplessly as her badly burned daughter suffers. My response was usually to hurry over and comfort these people; perhaps because I was uncomfortable with their grief, I would attempt to reduce their sadness with my cheery expressions and comforting words. I would embrace the woman and encourage her to try getting pregnant again; I would reassure the widower, reminding him that his wife had enjoyed a long and fruitful life; I would enter the burned child’s room with a smile rather than allowing the mother to weep in my arms. When my mother died, I was confused about whether I was allowed to be the grieving daughter or should continue being the grief-denying professional. It wasn’t until my father’s death many years later that I allowed myself to truly grieve, crying loudly and not caring about who might notice. I have finally come to the realization that it is both necessary and human for us each to grieve in our own individual way, so I avoid comforting others with false hope and cheer.

13. What did the author used to do when she saw people crying?
   A. She told them to be quiet.
   B. She suggested that they move to a different room.
   C. She stayed silent and kept an eye on the situation.
   D. She tried to encourage them and give them hope.

14. What was the author confused about when her mother died?
   A. She didn’t know whether she should cry.
   B. She didn’t know how to explain death to her daughter.
   C. She didn’t know whether she should talk to a counselor or not.
   D. She didn’t know whether she could keep working as a nurse.

15. Why did the author smile when she went into the burned child’s room?
   A. because another nurse told her something funny
   B. because her natural expression was always a smile
   C. because she wanted to help the child’s mother feel happier
   D. because she felt nervous

16. What would the author probably do now if she saw a man in the hospital chapel crying loudly?
   A. try to encourage him by saying everything will get better soon
   B. ask him to be a little more quiet and close the door
   C. tell him an interesting story to shift his attention away from the grief
   D. put a hand on his shoulder and tell him to let all his feelings out
I sometimes wonder why the idea of “story” is so important to me. Perhaps it is because we nurses, who often come to our profession filled with compassion and a desire to help people, can gradually stop seeing the individual patient. We sometimes forget the importance of empathy, which is difficult to sustain but essential to providing excellent health care. It isn’t easy to go from bedside to bedside and continually show the emotional response that allows us to best help each patient.

We see patients at unusual and stressful moments in their lives. We often know them only from the visual and physical information they present—symptoms, test results, and chief complaints. But stories of caregiving can help us to sharpen our empathy. They help us to be better, more compassionate, and even more emotionally fulfilled caregivers.

If we can write, talk, or read about our experiences with patients, those stories might transform us, inspire us, and give us the energy we need to reach out to our patients again and again. Patients’ stories are more than just the case histories written in their charts. We need to remember that they have rich, complicated, changing, multi-layered lives.

17. What do nurses sometimes stop doing after being a nurse for a long time?
   A. discussing their experiences with other nurses
   B. being careful about giving the correct medicine to each patient
   C. realizing that patients are under stress
   D. making an emotional connection to each patient

18. How can telling stories about patients help nurses?
   A. It teaches them how to recognize patients’ symptoms and complaints.
   B. It helps them do their jobs without becoming sad.
   C. It inspires them to be kinder to individual patients.
   D. It helps them form closer relationships with each other.

19. Why does the author think it is important to know more than the visual and physical information about a patient?
   A. because patients are under a lot of stress when they are at a hospital
   B. because nurses need to move on to the next patient quickly
   C. because patients have complex and dynamic lives
   D. because nurses show too much empathy for each patient

20. How does this author probably feel about assigning more patients to each nurse during a shift?
   A. She would disagree, because nurses should spend more time with each patient.
   B. She would disagree, because nurses need more time off.
   C. She would agree, because she wants the hospital to profit more.
   D. She would agree, because she wants to help as many people as possible.
As caregivers, we want the line between the state of being a “patient” and the state of being “well” to be clearly drawn. Patients exist in bodies that are, for some reason, not working properly. We caregivers exist in bodies that are working smoothly, allowing us to continue doing our daily tasks. Knowing the cost and pain of illness, we want to stay forever on the outside looking in, peering into patients’ ears and eyes and throats, listening to their hearts, changing their bandages and IVs. We want to be the ones in control. We want to treat and comfort and then step back, always returning safely home to our normal lives at the end of the day.

But what happens when one day that line disappears? What if *we* become the one changed by illness and suddenly transformed from healthy caregiver to suffering patient? It is likely that being a patient who is also a caregiver, full of knowledge about disease and healing, would affect the way we deal with our own illnesses or injury. Or what if the opposite happens? What if the line between patient and caregiver becomes so thick and dense that we cannot understand our patients’ feelings anymore, causing us to stop showing them true compassion?

21. What does the author say about the line between illness and wellness?
   A. Patients want that line to be clear.
   B. Patients want that line to disappear.
   C. Caregivers want that line to be clear.
   D. Caregivers want that line to disappear.

22. What does the author think would affect how a caregiver deals with his/her own illness?
   A. They know so much about sickness and treatments.
   B. They want to help their patients get well.
   C. They want to go home quickly.
   D. No one will show compassion to them.

23. According to the author, why do caregivers want to be in control?
   A. They don’t want to become ill themselves.
   B. They don’t want to help their patients.
   C. They think they know all the answers.
   D. They think patients are trying to hurt them.

24. Which of these does the author probably think is the bigger problem?
   A. caregivers becoming too involved in their patients’ lives
   B. caregivers not taking enough personal time for themselves
   C. caregivers not receiving good enough care when they become patients
   D. caregivers becoming too distant and careless with their patients

269
[Text 1]

‘These seats are excellent!’ exclaimed Father. They sat down and were getting comfortable when a tall young man approached them. He had long hair and was wearing a bright purple coat and shiny gold jewelry. ‘That seat is mine,’ he said. ‘You’ll have to find a different one.’ Father looked at him coldly and said, ‘I suppose, from your manner of speaking, that you are from Athens, although your hair is as long as a Spartan’s. If you are Athenian, however, you should be aware that these seats are for anyone.’

The young man replied loudly, ‘I was already sitting there! I just went to the restroom. Are you going to move or not?’ Father looked around. ‘No, certainly not. But there is a place for you behind us.’ The young man’s face turned red, and he moved angrily to the empty seat. Nico asked Father whether he knew the man. ‘His name is Hippias,’ said Father. ‘He is from a family with too much money and too much free time!’

For the rest of the morning, Alexis couldn’t stop watching the stage above them. He was completely immersed in the music, the dancing, and the wonderful words of the actors. Alexis enjoyed writing plays himself, too, but he did it secretly to avoid getting yelled at by his Father.

1. How did Alexis’s father guess that the young man was from Athens?
   A. his way of speaking
   B. his fashion sense
   C. his hairstyle
   D. the color of his skin

2. Where did the young man sit in the end?
   A. closer to the stage
   B. farther from the stage
   C. on the other side of the stage
   D. next to Alexis and his family

3. Why did the young man say the seat belonged to him?
   A. He had a ticket.
   B. He left his coat on the seat.
   C. He was sitting there first.
   D. He was richer than them.

4. What can we guess about Alexis’s father?
   A. He is from Sparta.
   B. He is a good friend of Hippias and his family.
   C. He doesn’t want Nico to ask so many questions.
   D. He doesn’t want Alexis to write plays.
Buck growled at the strange man, but he was surprised when the rope was pulled hard around his neck. He tried to attack, but the man knocked him down hard. For a few moments he was paralyzed, so the man quickly put him in the train. When Buck woke up, the man was watching him carefully, but Buck lunged at him quickly and bit his hand hard; then the rope was pulled again and Buck let go. That evening, the man took Buck to a bar in San Francisco, where the barman looked at the man’s hand and clothes covered in blood. ‘How much are you getting for this?’ he inquired. The man said he would receive a hundred and fifty dollars, and the barman remarked, ‘That’s a good price for a quality dog like him.’ The two men pushed Buck into a wooden box, where spent the night sleepless and confused. His neck ached, and he could not understand what was happening and what these men intended to do with him. The next day Buck was carried to the railway station and put on a train. For two whole days the train travelled north, and Buck was not given anything to eat or drink. The men on the train pushed sticks at him through the holes in the box. Buck became frustrated and started biting at anything that moved.

5. What did the man do when Buck bit his hand?
   A. He pulled the rope.
   B. He called the barman for help.
   C. He put Buck on a train.
   D. He hit Buck hard.

6. What did the barman say when he heard how much the man was being paid?
   A. He kept quiet about his opinion.
   B. He thought it was too much money.
   C. He thought it was a reasonable amount.
   D. He thought it was too cheap.

7. Why was Buck’s neck hurting while he was in the box at the bar?
   A. from riding on the train for a long time
   B. from not having any water to drink
   C. from having his rope pulled
   D. from being pushed into the small box

8. Why do you think Buck is travelling north?
   A. The person who paid the first man is there.
   B. He will be set free into the forest there.
   C. There were no trains going in other directions.
   D. His original owner is there.
I was a wild child and often got in trouble. My parents tried to punish me, but they never succeeded in changing me. I always refused to obey them and instead did whatever I wanted.

The first school I can remember was a large, old house in a small, quiet English town. I can still feel the coolness of the gardens outside the house. I can still smell the sweetness of the flowers and hear church bells ringing in the distance. There was a high wall around the gardens, and it had broken glass on top, just like a prison.

We only went outside the walls three times a week. On Saturdays after lunch, we took a walk in some fields nearby, under the watchful eyes of our teachers. And on Sundays we went to the village church in the morning and again in the evening.

I was not bored or unhappy during my life at school. Children can amuse themselves very easily. In my imagination, I lived an exciting life, full of mystery and adventure.

On the playground, I was the kind of boy who liked to give orders. I always wanted to win every game, every fight, and to be first in everything. All the other boys, even those older than me, followed an obeyed me. I suppose you could say that I was the playground bully.

9. How did the narrator’s parents try to change him?
   A. by teaching him
   B. by giving him freedom
   C. by punishing him
   D. by making him work

10. When were the students allowed to leave the school?
    A. only on weekends
    B. only in the evenings
    C. only during holidays
    D. only for church

11. When the boys were playing on the playground, which one of these things could they see?
    A. a wall
    B. a church
    C. a prison
    D. a field

12. Which type of person do you think this child will most likely become?
    A. someone who tries to learn things, like a scientific researcher
    B. someone who tries to help people, like a nurse or doctor
    C. someone who tries to grow things, like a farmer
    D. someone who tries to control people, like a manager or president
Can we ever truly understand another person’s suffering? Reading or writing about patients’ experiences might provoke a temporary emotional response, but does it actually help us become more sensitive and respond more compassionately to our patients? As a nurse, I have watched, recorded, discussed, and written about hundreds of bedside scenarios. As an author, I have written about those same encounters, trying to make them more accessible to future nurses and caregivers so that their hearts and minds will be more fully aware of the experiences of those who suffer. But when I put these words on a page, no matter how good my writing is, I can still only represent the original event—the patient’s reality—from a distance. Students and nurses might understand their patients in a new or deeper way through literature and writing. But in the end we can never truly see anything from the patient’s perspective. The only emotions I can honestly and accurately describe are my own. I can share how I feel when I’m the patient being invaded by instruments or illness, or when my body is being examined by a doctor’s hands, or I can describe what goes through my mind as a nurse bending over a dying patient.

13. Why does the author write stories about her patients?
   A. to entertain people and make them laugh and cry
   B. to help caregivers understand things from a patient’s perspective
   C. to help relieve the suffering of her patients themselves
   D. to prepare herself for when she becomes a patient in the future

14. Which feelings can the author describe most accurately?
   A. her own feelings
   B. the feelings of other nurses and caregivers
   C. her patients’ feelings
   D. the feelings of her own family members

15. Which of these statements best summarizes the author’s main point?
   A. Stories about patients are not useful for helping us understand their feelings.
   B. Stories about patients are the only way we can understand their feelings completely.
   C. Stories about patients can help us understand their feelings a little, but not completely.
   D. Stories about patients are interesting and entertaining, but not educational.

16. Which method would the author probably recommend for researching patients’ perspectives?
   A. evaluating a video recording of patient reactions to different treatments
   B. interviewing nurses about their patients’ experiences
   C. having patients fill out a questionnaire
   D. evaluating an audio recording of patients explaining their thoughts and feelings
When I feel ill, I conduct an objective evaluation of my medical history and symptoms, just like I would do with a patient. After telephoning my manager at the hospital to request the day off, I began my self-diagnosis. My throat was extremely painful, and when I used a flashlight to inspect it, I observed that it was bright red. I inhaled slowly and deeply, checking whether my lungs expanded fully, which definitely was not possible. The virus seemed to be transitioning from my throat down into my lungs; when I took a breath, it felt like I was thirty meters underwater. Upon awakening this morning, I had discovered that I had no energy at all, and I was annoyed by a continuous high-pitched sound in my ears. I had also had an intense coughing episode the previous evening at the hospital. I experienced some dizziness, and I started worrying that I might possibly have a very serious disease. I considered whether I should visit the emergency room. I suddenly panicked and wondered whether I might be dying. Younger people rarely think this way, but when we transition into our fifties and sixties, a serious illness can occasionally bring up frightening thoughts about whether the end might be near.

17. What did the author think about her illness?
   A. It is probably just a seasonal cold.
   B. It might be a severe allergic reaction.
   C. A viral infection could be moving into her lungs.
   D. The dizziness might be caused by her medication.

18. According to the author, which people are more likely to worry about the possibility of death?
   A. medical professionals
   B. people who get sick frequently
   C. people who are getting older
   D. people who are more highly educated

19. Which of these symptoms did the author NOT mention?
   A. a severe headache
   B. difficulty breathing
   C. a sore throat
   D. lack of energy

20. When the narrator called her manager, what do you think the manager said?
   A. “Please stay home today and try to get well.”
   B. “We have so many patients, so we really need you to come to work anyway.”
   C. “You should apply to be a manager soon, because you are quite good at your job.”
   D. “Please be sure to arrive at work earlier tomorrow.”
As caregivers, we want the line between the state of being a “patient” and the state of being “well” to be clearly drawn. Patients exist in bodies that are, for some reason, not working properly. We caregivers exist in bodies that are working smoothly, allowing us to continue doing our daily tasks. Knowing the cost and pain of illness, we want to stay forever on the outside looking in, peering into patients’ ears and eyes and throats, listening to their hearts, changing their bandages and IVs. We want to be the ones in control. We want to treat and comfort and then step back, always returning safely home to our normal lives at the end of the day.

But what happens when one day that line disappears? What if we become the one changed by illness and suddenly transformed from healthy caregiver to suffering patient? It is likely that being a patient who is also a caregiver, full of knowledge about disease and healing, would affect the way we deal with our own illnesses or injury. Or what if the opposite happens? What if the line between patient and caregiver becomes so thick and dense that we cannot understand our patients’ feelings anymore, causing us to stop showing them true compassion?

21. What does the author say about the line between illness and wellness?
   A. Caregivers want that line to be clear.
   B. Caregivers want that line to disappear.
   C. Patients want that line to be clear.
   D. Patients want that line to disappear.

22. What does the author think would affect how a caregiver deals with his/her own illness?
   A. No one will show compassion to them.
   B. They know so much about sickness and treatments.
   C. They want to help their patients get well.
   D. They want to go home quickly.

23. According to the author, why do caregivers want to be in control?
   A. They don’t want to help their patients.
   B. They think they know all the answers.
   C. They think patients are trying to hurt them.
   D. They don’t want to become ill themselves.

24. Which of these does the author probably think is the bigger problem?
   A. caregivers not taking enough personal time for themselves
   B. caregivers becoming too involved in their patients’ lives
   C. caregivers becoming too distant and careless with their patients
   D. caregivers not receiving good enough care when they become patients
APPENDIX D

VOCABULARY SIZE TEST ITEMS

[Form A (Time 1)]

Example:
drive: He drives fast.  
A. swims  
B. learns  
C. throws balls  
D. uses a car

1. period: It was a difficult period.  
A. question  
B. time  
C. thing to do  
D. book

2. see: They saw it.  
A. closed it tightly  
B. waited for it  
C. looked at it  
D. started it up

3. poor: We are poor.  
A. have no money  
B. feel happy  
C. are very interested  
D. do not like to work hard

4. microphone: Please use the microphone.  
A. machine for making food hot  
B. machine that makes sounds louder  
C. machine that makes things look bigger  
D. small telephone that can be carried around

5. time: They have a lot of time.  
A. money  
B. food  
C. hours  
D. friends

6. pub: They went to the pub.  
A. place where people drink and talk  
B. place that looks after money  
C. large building with many shops  
D. building for swimming

7. circle: Make a circle.  
A. rough picture  
B. space with nothing in it  
C. round shape  
D. large hole

8. figure: Is this the right figure?  
A. answer  
B. place  
C. time  
D. number

9. soldier: He is a soldier.  
A. person in a business  
B. person who studies  
C. person who uses metal  
D. person in the army
10. **restore**: It has been **restored**.
   A. said again  
   B. given to a different person  
   C. made like new again  
   D. given a lower price  

11. **dig**: Our dog often **digs**.
    A. solves problems with things  
    B. creates a hole in the ground  
    C. wants to sleep  
    D. enters the water  

12. **pave**: It was **paved**.
    A. prevented from going through  
    B. divided  
    C. given gold edges  
    D. covered with a hard surface  

13. **pro**: He’s a **pro**.
    A. someone who is employed to find out important secrets  
    B. a stupid person  
    C. someone who writes for a newspaper  
    D. someone who is paid for playing a sport  

14. **compound**: They made a new **compound**.
    A. agreement  
    B. thing made of two or more parts  
    C. group of people forming a business  
    D. guess based on past experience  

15. **strap**: He broke the **strap**.
    A. promise  
    B. top cover  
    C. shallow dish for food  
    D. strip of strong material  

16. **allege**: They **alleged** it.
    A. claimed it without proof  
    B. stole the ideas for it from someone else  
    C. provided facts to prove it  
    D. argued against the facts that supported it  

17. **crab**: Do you like **crabs**?
    A. very thin small cakes  
    B. tight, hard collars  
    C. sea creatures that always walk to one side  
    D. large black insects that sing at night  

18. **dinosaur**: The children were pretending to be **dinosaurs**.
    A. robbers who work at sea  
    B. very small creatures with human form but with wings  
    C. large creatures with wings that breathe fire  
    D. animals that lived an extremely long time ago
19. **deficit**: The company had a large **deficit**.
   A. spent a lot more money than it earned
   B. went down a lot in value
   C. had a plan for its spending that used a lot of money
   D. had a lot of money stored in the bank

20. **weep**: He **wept**.
    A. finished his course  B. cried
    C. died  D. worried

21. **haunt**: The house is **haunted**.
    A. full of decorations  B. rented
    C. empty  D. full of ghosts

22. **cube**: I need one more **cube**.
    A. sharp thing used for joining things  B. solid square block
    C. tall cup with no saucer  D. piece of stiff paper folded in half

23. **nun**: We saw a **nun**.
    A. long thin creature that lives in the earth
    B. terrible accident
    C. woman following a strict religious life
    D. unexplained bright light in the sky

24. **strangle**: He **strangled** her.
    A. killed her by pressing her throat  B. gave her all the things she wanted
    C. took her away by force  D. admired her greatly

25. **marrow**: This is the **marrow**.
    A. symbol that brings good luck to a team  B. soft center of a bone
    C. control for guiding a plane  D. increase in salary

26. **jug**: He was holding a **jug**.
    A. a container for pouring liquids  B. an informal discussion
    C. a soft cap  D. a weapon that blows up

27. **cavalier**: He treated her in a **cavalier** manner.
    A. without care  B. with good manners
    C. awkwardly  D. as a brother would

28. **malign**: His **malign** influence is still felt.
    A. good  B. evil
    C. very important  D. secret

29. **olive**: We bought **olives**.
    A. oily fruit  B. scented flowers
    C. men's swimming clothes  D. tools for digging
30. **quilt**: They made a *quilt*.
   A. statement about who should get their property when they die
   B. firm agreement
   C. thick warm cover for a bed
   D. feather pen

31. **shudder**: The boy *shuddered*.
   A. spoke with a low voice  
   B. almost fell
   C. shook
   D. called out loudly

32. **veer**: The car *veered*.
   A. moved shakily
   B. changed course
   C. made a very loud noise
   D. slid without the wheels turning

33. **bristle**: The *bristles* are too hard.
   A. questions
   B. short stiff hairs
   C. folding beds
   D. bottoms of the shoes

34. **yogurt**: The *yogurt* is disgusting.
   A. dark grey mud found at the bottom of rivers
   B. unhealthy, open sore
   C. thick, soured milk, often with sugar flavoring
   D. large purple fruit with soft flesh

35. **candid**: Please be *candid*.
   A. be careful
   B. show sympathy
   C. show fairness to both sides
   D. say what you really think

[Form B (Time 2)]

*Example:*

**circle**: Make a *circle*.
   A. rough picture
   B. space with nothing in it
   C. round shape
   D. large hole

1. **jump**: She tried to *jump*.
   A. lie on top of the water
   B. get off the ground suddenly
   C. stop the car at the edge of the road
   D. move very fast

2. **shoe**: Where is your *shoe*?
   A. the person who looks after you
   B. the thing you keep your money in
   C. the thing you use for writing
   D. the thing you wear on your foot
3. **see**: They **saw** it.
   A. closed it tightly       B. waited for it
   C. looked at it           D. started it up

4. **standard**: Her **standards** are very high.
   A. the bits at the back under her shoes
   B. the levels she reaches in everything
   C. the marks she gets in school
   D. the money she asks for

5. **period**: It was a difficult **period**.
   A. question       B. time
   C. thing to do    D. book

6. **basis**: This was used as the **basis**.
   A. answer       B. place to take a rest
   C. next step    D. main part

7. **stone**: He sat on a **stone**.
   A. hard thing       B. kind of chair
   C. soft thing on the floor   D. part of a tree

8. **maintain**: Can they **maintain** it?
   A. keep it as it is       B. make it larger
   C. get a better one than it   D. get it

9. **microphone**: Please use the **microphone**.
   A. machine for making food hot
   B. machine that makes sounds louder
   C. machine that makes things look bigger
   D. small telephone that can be carried around

10. **pave**: It was **paved**.
    A. prevented from going through       B. divided
    C. given gold edges       D. covered with a hard surface

11. **upset**: I am **upset**.
    A. tired       B. famous
    C. rich       D. unhappy

12. **remedy**: We found a good **remedy**.
    A. place to eat in public        B. way to fix a problem
    C. way to prepare food       D. rule about numbers
13. **strap**: He broke the **strap**.
   A. promise  
   B. top cover  
   C. shallow dish for food  
   D. strip of strong material  

14. **restore**: It has been **restored**.
   A. said again  
   B. given to a different person  
   C. made like new again  
   D. given a lower price  

15. **bacterium**: They didn’t find a single **bacterium**.
   A. small living thing causing disease  
   B. plant with red or orange flowers  
   C. animal that carries water in lumps on its back  
   D. thing that has been stolen and sold to a shop  

16. **rove**: He couldn’t stop **roving**.
   A. getting drunk  
   B. traveling around  
   C. making a musical sound through closed lips  
   D. working hard  

17. **dig**: Our dog often **digs**.
   A. solves problems with things  
   B. creates a hole in the ground  
   C. wants to sleep  
   D. enters the water  

18. **miniature**: It is a **miniature**.
   A. an instrument for looking at very small objects  
   B. a very small thing of its kind  
   C. a very small living creature  
   D. a small line to join letters in handwriting  

19. **peasantry**: He did a lot for the **peasantry**.
   A. local people  
   B. place of worship  
   C. businessmen’s club  
   D. working class people  

20. **dash**: They **dashed** over it.
   A. moved quickly  
   B. moved slowly  
   C. fought  
   D. looked quickly  

21. **peel**: Shall I **peel** it?
   A. let it sit in water for a long time  
   B. take the skin off it  
   C. make it white  
   D. cut it into thin pieces
22. **thesis**: She has completed her *thesis*.
   A. talk given by a judge at the end of a trial
   B. first year of employment after becoming a teacher
   C. long written report of study carried out for a university degree
   D. extended course of hospital treatment

23. **scrub**: He is *scrubbing* it.
   A. cutting shallow lines into it       B. repairing it
   C. washing it energetically            D. drawing simple pictures of it

24. **palette**: He lost his *palette*.
   A. container for carrying fish       B. wish to eat food
   C. young female companion              D. artist’s board for mixing paints

25. **puritan**: He is a *puritan*.
   A. person who likes attention       B. person with strict morals
   C. person with a moving home         D. person who keeps money and hates spending it

26. **jug**: He was holding a *jug*.
   A. a container for pouring liquids   B. an informal discussion
   C. a soft cap                       D. a weapon that blows up

27. **erratic**: He was *erratic*.
   A. without fault                    B. very bad
   C. very respectful                  D. unsteady

28. **quiz**: We made a *quiz*.
   A. thing to hold arrows              B. serious mistake
   C. set of questions                  D. box for birds to make nests in

29. **olive**: We bought *olives*.
   A. oily fruit                       B. scented flowers
   C. men’s swimming clothes           D. tools for digging

30. **bloc**: They have joined this *bloc*.
   A. group of countries with a common purpose
   B. band of criminals
   C. small group of soldiers sent ahead
   D. musical group

31. **vocabulary**: You will need more *vocabulary*.
   A. words                            B. skill
   C. money                            D. guns
32. **veer**: The car **veered**.
   A. moved shakily
   B. changed course
   C. made a very loud noise
   D. slid without the wheels turning

33. **mumble**: He started to **mumble**.
   A. think deeply
   B. speak in an unclear way
   C. stay further behind the others
   D. shake uncontrollably

34. **null**: His influence was **null**.
   A. had good results
   B. did not help much
   C. had no effect
   D. lasted a long time

35. **stealth**: They did it by **stealth**.
   A. spending a large amount of money
   B. hurting someone so much that they agreed to their demands
   C. moving secretly with extreme care and quietness
   D. taking no notice of problems they met

[Form C (Time 3)]

Example:

**circle**: Make a **circle**.
   A. rough picture
   B. space with nothing in it
   C. **round shape**
   D. large hole

1. **joke**: We did not understand his **joke**.
   A. attempt at humor
   B. false statement
   C. way of speaking
   D. way of thinking

2. **shoe**: Where is your **shoe**?
   A. the person who looks after you
   B. the thing you keep your money in
   C. the thing you use for writing
   D. the thing you wear on your foot

3. **fine**: She was **fine**.
   A. laughing
   B. good
   C. late
   D. excited

4. **drawer**: The **drawer** was empty.
   A. sliding box
   B. place where cars are kept
   C. cupboard to keep things cold
   D. animal house
5. **lonesome**: He felt **lonesome**.
   A. ungrateful  B. very tired
   C. without company  D. full of energy

6. **examine**: I need to **examine** it.
   A. trace its former path  B. conceive or picture it
   C. look at it closely  D. gain entry to it

7. **eye**: He looked at her **eye**.
   A. the lowest part of her leg  B. one of her organs of sight
   C. the knife mark on her body  D. the plan for spending money

8. **guarantee**: There is no **guarantee** that you will win.
   A. promise given  B. sign of hope
   C. feeling of sureness  D. acknowledgement or confession

9. **accessory**: They gave us some **accessories**.
   A. papers giving us the right to enter a country  B. official orders
   C. ideas to choose between  D. extra pieces

10. **authority**: They had no **authority**.
    A. were not disciplined  B. did not know who they were
    C. possessed nothing useful  D. lacked any power

11. **upset**: I am **upset**.
    A. tired  B. famous
    C. rich  D. unhappy

12. **remedy**: We found a good **remedy**.
    A. place to eat in public  B. way to fix a problem
    C. way to prepare food  D. rule about numbers

13. **authentic**: It is **authentic**.
    A. real  B. very noisy
    C. old  D. like a desert

14. **compound**: They made a new **compound**.
    A. agreement  B. thing made of two or more parts
    C. group of people forming a business  D. guess based on past experience
15. **input**: We need more **input**.
   A. money given to something  
   B. workers cooperating to do something  
   C. artificial filling for a hole in wood  
   D. information or power contributed to something

16. **compost**: We need some **compost**.
   A. strong support  
   B. help to feel better  
   C. hard stuff made of stones and sand stuck together  
   D. fertilizer made from plant material

17. **system**: They created a **system**.
   A. business  
   B. dangerous thing  
   C. organized way of doing things  
   D. problem that is difficult to solve

18. **awe**: They looked at the mountain in **awe**.
   A. with a worried expression  
   B. with an interested expression  
   C. with a sense of wonder  
   D. with a feeling of respect

19. **latter**: I agree with the **latter**.
   A. man from the church  
   B. reason given  
   C. last one  
   D. answer

20. **threshold**: They raised the **threshold**.
   A. flag  
   B. point or line where something changes  
   C. roof inside a building  
   D. cost of borrowing money

21. **peel**: Shall I **peel** it?
   A. let it sit in water for a long time  
   B. take the skin off it  
   C. make it white  
   D. cut it into thin pieces

22. **thesis**: She has completed her **thesis**.
   A. talk given by a judge at the end of a trial  
   B. first year of employment after becoming a teacher  
   C. long written report of study carried out for a university degree  
   D. extended course of hospital treatment

23. **eclipse**: There was an **eclipse**.
   A. A strong wind blew all day  
   B. I heard something hit the water  
   C. A large number of people were killed  
   D. The sun was hidden by the moon
24. **strangle**: He **strangled** her.
   A. killed her by pressing her throat  B. gave her all the things she wanted
   C. took her away by force           D. admired her greatly

25. **ruble**: He had a lot of **rubes**.
   A. very valuable red stones        B. distant members of his family
   C. Russian money                   D. moral or other difficulties in the mind

26. **multitudinious**: **Multitudinous** birds filled the sky.
   A. meat eating birds               B. birds that fly from one country to another
   C. very large numbers of birds      D. birds with feathers of many different colors

27. **haze**: We looked through the **haze**.
   A. small round window in a ship    B. unclear air
   C. cover for a window made of strips of wood or plastic D. list of names

28. **malign**: His **malign** influence is still felt.
   A. good                            B. evil
   C. very important                  D. secret

29. **demography**: This book is about **demography**.
   A. the study of patterns of land use
   B. the study of the use of pictures to show facts about numbers
   C. the study of the movement of water
   D. the study of population

30. **bloc**: They have joined this **bloc**.
   A. group of countries with a common purpose
   B. band of criminals
   C. small group of soldiers sent ahead
   D. musical group

31. **feasible**: It was not **feasible**.
   A. something we could predict       B. something we could do
   C. something we could remove        D. something we could find again

32. **maxim**: Their **maxim** is “Reduce, Re-use, Recycle.
   A. repeating short song              B. code of conduct
   C. strategy for concealment          D. business strategy

33. **saga**: What’s the latest **saga**?
   A. joke to play on people            B. series of dramatic events
   C. way to rebel                      D. way to fool oneself
34. **egalitarian**: This organization is very egalitarian.
   A. does not provide much information about itself to the public
   B. dislikes changing their policies for anyone
   C. frequently asks a court of law for a judgment
   D. treats everyone who works for it as if they are equal

35. **stealth**: They did it by stealth.
   A. spending a large amount of money
   B. hurting someone so much that they agreed to their demands
   C. moving secretly with extreme care and quietness
   D. taking no notice of problems they met

[Form D (Time 4)]

**Example:**

**shoe**: Where is your shoe?
   A. the person who looks after you  B. the thing you keep your money in
   C. the thing you use for writing  D. the thing you wear on your foot

1. **extra**: Do you have an extra pen?
   A. another pen  B. a special pen
   C. a simple pen  D. a similar pen

2. **visit**: Let’s visit him tomorrow!
   A. go to his place  B. take him to the hospital
   C. strike him hard  D. tell him about our idea

3. **each**: Give one to each person.
   A. some people  B. half the people
   C. every person  D. most people

4. **standard**: Her standards are very high.
   A. the money she asks for  B. the bits at the back under her shoes
   C. the levels she reaches in everything  D. the marks she gets in school

5. **lonesome**: He felt lonesome.
   A. very tired  B. ungrateful
   C. full of energy  D. without company

6. **basis**: This was used as the basis.
   A. main part  B. next step
   C. place to take a rest  D. answer
7. **true**: Is it true?
   A. the colour of the sky   B. real
   C. ready   D. here now

8. **decent**: He is such a decent man.
   A. a kind and respectable man   B. an angry man
   C. a rich man   D. a well known man

9. **enzyme**: You do not have the proper enzymes.
   A. proteins in your stomach and mouth
   B. shiny coatings on pottery
   C. chemical compounds in plastic
   D. medicine for calming people down

10. **myth**: That is a myth.
    A. contest between many people
    B. combination or blend of things
    C. subject of thought and discussion
    D. false but widely believed idea or story

11. **stable**: The desk is not stable.
    A. the right size   B. made of steel
    C. sitting solidly   D. what we want

12. **hypothesis**: I proved my hypothesis.
    A. principle of faith   B. guess based on reason
    C. loyalty to the cause   D. ability to make babies

13. **fracture**: They found a fracture.
    A. break   B. mistake
    C. short coat   D. discount certificate

14. **scandal**: There is a scandal at my workplace.
    A. an old-fashioned source of light
    B. a person from Denmark, Norway, or Sweden
    C. a shocking and embarrassing situation
    D. an illness that makes you feel hot

15. **allege**: They alleged it.
    A. argued against the facts that supported it   B. claimed it without proof
    C. stole the ideas for it from someone else   D. provided facts to prove it
16. **compost**: We need some **compost**.
   A. fertilizer made from plant material
   B. hard stuff made of stones and sand stuck together
   C. help to feel better
   D. strong support

17. **plant**: I’m going to cut this **plant**.
   A. kind of paper
   B. sweet fancy bread
   C. fruit-filled pie
   D. green living thing

18. **nun**: We saw a **nun**.
   A. long thin creature that lives in the earth
   B. terrible accident
   C. woman following a strict religious life
   D. unexplained bright light in the sky

19. **latter**: I agree with the **latter**.
   A. last one man
   B. answer
   C. from the church
   D. reason given

20. **threshold**: They raised the **threshold**.
   A. point or line where something changes
   B. roof inside a building
   C. cost of borrowing money
   D. flag

21. **patience**: He has no **patience**.
   A. has no free time
   B. has no faith
   C. will not wait happily
   D. does not know what is fair

22. **chaos**: I went shopping in the city, and it was **chaotic**!
   A. appealed to my imagination
   B. gave me some good exercise
   C. ended in disaster
   D. felt confusing and disorderly

23. **scrub**: He is **scrubbing** it.
   A. cutting shallow lines into it
   B. repairing it
   C. washing it energetically
   D. drawing simple pictures of it

24. **unravel**: It is going to **unravel**.
   A. come undone
   B. become unblocked
   C. come open
   D. begin to fall over
25. **reptile**: She looked at the **reptile**.
   A. old hand-written book
   B. animal with cold blood
   C. person who sells things by knocking on doors
   D. picture made by joining small pieces of different colours

26. **multitudinous**: **Multitudinous** birds filled the sky.
   A. meat eating birds
   B. very large numbers of birds
   C. birds with feathers of many different colors
   D. birds that fly from one country to another

27. **whim**: He had lots of **whims**.
   A. old gold coins from Europe
   B. young female horses

28. **gall**: Her presence **galled** me.
   A. saddened me
   B. made me feel frightened and concerned
   C. made me feel happy and excited
   D. annoyed and provoked me

29. **diagnosis**: Could you please explain your **diagnosis**?
   A. figure composed of lines
   B. tendency to favour one side
   C. difficult problem
   D. interpretation of disease symptoms

30. **curtail**: Her plans were **curtailed**.
   A. cut short
   B. disregarded
   C. thorough and clever
   D. blocked

31. **asylum**: They need **asylum**.
   A. a light, bendable metal
   B. summaries of the lectures
   C. protection from pursuers
   D. water in any form

32. **avalanche**: The animals died in the **avalanche**.
   A. channel for waste water on a farm
   B. upper part of an old house’s roof
   C. large mass of snow falling down a mountain
   D. storage place for weapons of war

33. **ambush**: It’s an **ambush**.
   A. a surprise attack
   B. an illegal taking of a person
   C. a big and tightly curled hairstyle
   D. a mythical female soldier

34. **estuary**: It’s beside the **estuary**.
   A. home of a religious brotherhood
   B. resting place of dead people
   C. place of safety
   D. mouth of a tidal river
35. **hooker**: There was a news report about **hookers**.
   A. people who sell sexual services
   B. people who violently take over a vehicle
   C. people who transport illegal imports
   D. people who walk in rough country
APPENDIX E

FACETS COMMAND FILE

title = Prosody-1-simp
output measures file = prevla.out; the output file
convergence = 0.1 ; size of largest remaining marginal score residual at convergence
unexpected = 3.0 ; size of smallest standardized residual to report
arrange = M ; arrange output tables in Measure ascending order
facets = 3 ; there are 3 facets in this analysis
noncenter = 2 ; rater facet floats
Inter-rater = 2
positive = 1 ; for groups, greater score => greater measure
usort = 2,3,1 ; sort residuals by 2=raters, 3= criteria, 1= students
Model=? , ?, ?, R4 ; observations are ratings in range 1-4.

Labels=
1, Students
   1-77 ; 77 students
*

2, Rater
   1-4 ; 4 raters
*

3, Criteria
   1=phrasing ; 3 criteria
   2=stress
   3=intonation
*

data=
   1,1,1-3,2,2,1
APPENDIX F

WINSTEPS COMMAND FILE

&INST                     : shows this is a control file (optional)
TITLE='RC-pre-spring'     : Report title
NAME1=1                   : First column of person label in data file
ITEM1=8                   : First column of responses in data file
NI=24                     : Number of items
CODES=abcd                : Valid response codes in the data file
KEY1=cbdacacabadbcdcacccbbdd
CLFILE=                  : LABELS THE OBSERVATIONS
*                  : "*" is the end of a list
DIF=$S$6$W1
CONVERGE=B
LCONV=.001
RCONV=.01
PERSON=STUDENT            : Person title
ITEM=ITEM                 : Item title
&END                    : Item labels for 24 items follow
 S1-1
 S1-2
 S1-3
 S1-4
 S2-1
 S2-2
 S2-3
 S2-4
 S3-1
 S3-2
 S3-3
 S3-4
 A1-1
 A1-2
 A1-3
 A1-4
 A2-1
 A2-2
 A2-3
 A2-4
 A3-1
 A3-2
 A3-3
 A3-4
END NAMES
A-IR-03cbdacacbbadbcacacaccbcd
APPENDIX G

EXAMPLE SILENT READING RATE CHANGE GRAPHS, ER TREATMENT

Gradual Increase Graph, Example 1 (Student S9)
Gradual Increase Graph, Example 2 (Student S59)
Jagged Increase Graph, Example 1 (Student S8)

Words per minute

Measurement

117.14
145.05
113.42
106.68
132.88
Jagged Increase Graph, Example 2 (Student S71)
Note. Only one incidence of this pattern type was identified.
Plateau Increase Graph, Example 1 (Student S13)

![Graph image]

Words per minute
Measurement

60 70 80 90 100 110 120 130 140 150 160 170 180 190 200 210

1 2 3 4 5

126.60 161.76 161.94 162.69 163.68

126.60 161.76 161.94 162.69 163.68
Plateau Increase Graph, Example 2 (Student S74)