THE EFFECTS OF THE USE OF AN ICT-BASED READING INTERVENTION ON STUDENTS’ ACHIEVEMENT IN GRADE TWO

by

DEBORAH ANNE SCHNEIDER-RICHARDSON

Copyright © Deborah Schneider 2015

A Dissertation Submitted to the Faculty of the

DEPARTMENT OF DISABILITY AND PSYCHOEDUCATIONAL STUDIES

In Partial Fulfillment of the Requirements

For the Degree of

DOCTOR OF PHILOSOPHY

WITH A MAJOR IN SPECIAL EDUCATION

In the Graduate College

THE UNIVERSITY OF ARIZONA

2015
As members of the Dissertation Committee, we certify that we have read the dissertation prepared by Deborah Schneider, titled The Effects of the Use of an ICT-Based Reading Intervention on Students' Achievement in Grade Two and recommend that it be accepted as fulfilling the dissertation requirement for the Degree of Doctor of Philosophy.

Nancy Mather  
Date: 13 April 2015

Shirin Antia  
Date: 13 April 2015

Carl Liaupsin  
Date: 13 April 2015

Final approval and acceptance of this dissertation is contingent upon the candidate’s submission of the final copies of the dissertation to the Graduate College.

I hereby certify that I have read this dissertation prepared under my direction and recommend that it be accepted as fulfilling the dissertation requirement.

Dissertation Director: Nancy Mather  
Date: 13 April 2015
STATEMENT BY THE AUTHOR

This dissertation has been submitted in partial fulfillment of the requirements for an advanced degree at the University of Arizona and is deposited in the University Library to be made available to borrowers under rules of the Library.

Brief quotations from this dissertation are allowable without special permission, provided that an accurate acknowledgment of the source is made. Requests for permission for extended quotation from or reproduction of this manuscript in whole or in part may be granted by the copyright holder.

SIGNED: __________________________________________ Date: 13 April 2015
Deborah Schneider
ACKNOWLEDGEMENTS AND DEDICATION

I wish to acknowledge the kind and constant efforts of my singularly brilliant committee chair, Dr. Nancy Mather, without whose support, encouragement, and mentorship (plus no small amount of door opening) none of this would have been possible.

I also extend my sincere gratitude to the other members of my committee. Prof. Dr. Shirin Antia's earned authority, warmth, wisdom, and intelligence are unmatched and provide an ideal model to her students. Prof. Dr. Carl Liaupsin, who leads by quiet example, taught me that simplest is often best: My manuscripts will not be judge by their weight, but by their quality.

My most sincere thanks go to Prof. Dr. Debora Levine and Dr. Lesli Doan, who shared with me their wealth of statistical knowledge and helped me to build a bit of my own.

I would like to acknowledge the extraordinary efforts of my astute, unselfish, and talented fellow students, and Alex Chambers, Retina Bauschatz, and Merdyth Bauer.

I wish also to recognize my husband, Dr. Tobias Schneider, who regularly juggled two toddlers and ten time zones to ensure that I had the time and ability to complete this endeavor. Merci mille fois, mon amour. Je t'aime plus que la vie elle-meme.

Finally, I wish to dedicate this dissertation to my mother, Anne Richardson, whose quick mind and uns lakable thirst for knowledge have served always as my inspiration. Had she been granted a small fraction of the opportunity I have been so fortunate to enjoy, I know there would now be a veritable alphabet of letters trailing her name. Thanks, Mom.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>LIST OF TABLES</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>12</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>14</td>
</tr>
<tr>
<td>ABSTRACT</td>
<td>15</td>
</tr>
<tr>
<td>CHAPTER ONE: INTRODUCTION</td>
<td>16</td>
</tr>
<tr>
<td>Statement of the Problem</td>
<td>19</td>
</tr>
<tr>
<td>Nature of the Study</td>
<td>19</td>
</tr>
<tr>
<td>Conceptual Framework</td>
<td>20</td>
</tr>
<tr>
<td>Objectives of the Present Study</td>
<td>20</td>
</tr>
<tr>
<td>Importance of the Present Study</td>
<td>22</td>
</tr>
<tr>
<td>Research Question</td>
<td>22</td>
</tr>
<tr>
<td>Statement of Hypotheses</td>
<td>22</td>
</tr>
<tr>
<td>Participants</td>
<td>23</td>
</tr>
<tr>
<td>Method</td>
<td>24</td>
</tr>
<tr>
<td>Limitations</td>
<td>24</td>
</tr>
<tr>
<td>Definitions of Terms</td>
<td>26</td>
</tr>
<tr>
<td>CHAPTER TWO: REVIEW OF LITERATURE</td>
<td>30</td>
</tr>
<tr>
<td>Proposed Advantages of ICT-Based Reading Interventions</td>
<td>31</td>
</tr>
<tr>
<td>Multimodal Instruction</td>
<td>31</td>
</tr>
<tr>
<td>Formative Feedback</td>
<td>32</td>
</tr>
<tr>
<td>Interactivity</td>
<td>33</td>
</tr>
<tr>
<td>Mastery Learning</td>
<td>34</td>
</tr>
</tbody>
</table>
Studies that Did Not Meet Quality Standards............................................75
Studies that Met Quality Standards............................................................76
Discussion..................................................................................................80
Summary of Findings.................................................................................80
Implications for Future Research...............................................................81
Implications for Practice...........................................................................82
Limitations..................................................................................................83
CHAPTER THREE: METHOD.......................................................................84
Overview of Research Questions and Design............................................84
Participants and Settings..........................................................................85
Participants...............................................................................................85
Participants Demographics.......................................................................85
Assignment to Groups..............................................................................86
Intervention Settings...............................................................................87
Materials....................................................................................................87
MindPlay Virtual Reading Coach...............................................................87
MVRC Overview......................................................................................87
Measures....................................................................................................90
Dependent Variable Measures.................................................................90
WJ IV ACH...............................................................................................91
TOSWF-2.................................................................................................91
Rationale for Measure Selection...............................................................93
Fidelity Measures......................................................................................93
CHAPTER FOUR: RESULTS

Data Screening and Cleaning

Measurement of Data
CHAPTER FIVE: RESEARCH REPORT

The Effects of the Use of an ICT-Based Reading Intervention on Students’ Achievement in Grade Two

Teacher Preparation and Professional Development

Purpose

Research Question

Method

Participants and Settings

Participants

Participants Demographics

Assignment to Groups
Intervention Settings.................................................................133

Materials..........................................................................................134

MindPlay Virtual Reading Coach................................................134

Measures..........................................................................................134

Dependent Variable Measures.............................................134

Rationale for Measure Selection.............................................135

Fidelity Measures........................................................................136

Product Usage Logs......................................................................136

Behavioral Observations.......................................................136

Intervention Procedure.............................................................137

Routine Classroom Instruction............................................137

Intervention Instruction...........................................................137

Comparison Instruction.............................................................138

Teacher Training.........................................................................138

Assessment Procedure...............................................................139

Procedures to Ensure Accuracy of Administration................139

Procedures to Ensure Accuracy of Scoring..............................139

Administration of Dependent Variable Measures....................139

Data Analysis..................................................................................140

Measurement of Data....................................................................140

Missing Data..................................................................................140

Identification of Covariates..........................................................141

Test Procedure..............................................................................141
Tests of Equality of Variance and Covariance.........................142
Multivariate Tests.....................................................................142
Univariate Tests.......................................................................143
Additional Post-Hoc Tests...................................................145
Fidelity of Implementation.....................................................147
Product Usage Logs...............................................................147
Planned Activity Check..........................................................147
Discussion..............................................................................148
Limitations............................................................................150
Conclusions and Implications for Future Research..................150
APPENDIX A: ANCHORED MATRIX FOR EVALUATING EXPERIMENTAL
RESEARCH (AMEER).................................................................153
APPENDIX B: MINDPLAY VIRTUAL READING COACH COMPONENT
DESCRIPTIONS........................................................................158
APPENDIX C: PLANNED ACTIVITY CHECK..............................171
REFERENCES.........................................................................172
LIST OF TABLES

Table 1. Overview of Studies.............................................................................................66
Table 2. Study Quality........................................................................................................75
Table 3. Tests of Reading Achievement............................................................................92
Table 4. Descriptive Statistics by Condition for Each Dependent Variable and Interval-Level Covariate....................................................................................................108
Table 5. Shapiro-Wilk Test of Normality........................................................................109
Table 6. Pearson’s Product-Moment Correlation Analysis.............................................110
Table 7. Box's Test of Equality of Covariance Matrices.................................................113
Table 8. Levene's Test of Equality of Error Variances.....................................................113
Table 9. Multivariate Tests..............................................................................................114
Table 10. Tests of Between-Subjects Effects.................................................................116
Table 11. Pairwise Comparisons for Univariate Tests....................................................117
Table 12. Between-Subjects Effects for Real Word Reading...........................................118
Table 13. Pairwise Comparisons for Real Word Reading...............................................118
Table 14. Between-Subjects Effects for Real Word Spelling...........................................119
Table 15. Pairwise Comparisons for Real Word Spelling...............................................119
Table 16. Between-Subjects Effects for Non-Word Reading..........................................120
Table 17. Pairwise Comparisons for Non-Word Reading...............................................120
Table 18. Between-Subjects Effects for Non-Word Spelling..........................................121
Table 19. Pairwise Comparisons for Non-Word Spelling...............................................121
Table 20. Between-Subjects Effects for Reading Fluency..............................................122
Table 21. Pairwise Comparisons for Reading Fluency....................................................122
Table 22. Student Engagement Behavior per the Planned Activity Check.........................123
Table 23. Tests of Reading Achievement........................................................................135
Table 24. Multivariate Tests............................................................................................143
Table 25. Tests of Between-Subjects Effects......................................................................144
Table 26. Pairwise Comparisons for Univariate Tests.......................................................145
Table 27. Pairwise Comparisons for Post-Hoc ANCOVAs..............................................146
Table 28. Between-Subjects Effects for Post-Hoc ANCOVAs.........................................147
Table 29. Student Engagement Behavior per the Planned Activity Check.........................148
## LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>MVRC Lesson Sequence and Assessment Checkpoints</td>
<td>90</td>
</tr>
<tr>
<td>2</td>
<td>RAPS 360 Full Diagnostic Assessment Paths</td>
<td>159</td>
</tr>
<tr>
<td>3</td>
<td>Example RAPS360 Vocabulary Assessment Item</td>
<td>160</td>
</tr>
<tr>
<td>4</td>
<td>Example MVRC Phonemic Awareness Mastery Activity</td>
<td>161</td>
</tr>
<tr>
<td>5</td>
<td>Example MVRC Phonics Mastery Activity</td>
<td>163</td>
</tr>
<tr>
<td>6</td>
<td>Example MVRC Grammar Activity</td>
<td>164</td>
</tr>
<tr>
<td>7</td>
<td>Example MVRC Punctuation Activity</td>
<td>165</td>
</tr>
<tr>
<td>8</td>
<td>Example MVRC Vocabulary Activity</td>
<td>166</td>
</tr>
<tr>
<td>9</td>
<td>Example MVRC Eye Tracking Activity</td>
<td>168</td>
</tr>
<tr>
<td>10</td>
<td>Example MVRC Reading Comprehension Activity</td>
<td>169</td>
</tr>
<tr>
<td>11</td>
<td>Example MVRC Reading Fluency Activity</td>
<td>170</td>
</tr>
</tbody>
</table>
ABSTRACT

A quasi-experimental research design was used to evaluate the efficacy of MindPlay Virtual Reading Coach (MVRC), an ICT-based reading intervention, in addition to regular daily language instruction provided by a classroom teacher. After attrition, participants included 170 students enrolled in eight second-grade classrooms (four classrooms in each school) in two public elementary schools in the southwestern United States. Examiners obtained reading achievement data for each participating student. Pre- and post-test measures included tests of the Woodcock-Johnson Tests of Achievement (WJ IV ACH), as well as the Test of Silent Word Reading Fluency (TOSWRF-2). A multivariate analysis of covariance (MANCOVA) was used to determine whether there were significant mean differences in (a) non-word reading, (b) real word reading, (c) non-word spelling, (d) real word spelling, and/or (e) reading fluency post-test achievement scores favoring students assigned to use the MVRC online reading intervention, once the effects of differences in pre-test achievement scores and relevant demographic variables had been accounted for. Analyses revealed a significant main effect ($\lambda = .668, F[5, 161] = 16.014, p < .001, \text{multivariate } \eta^2 = .332$) of the intervention on achievement scores of participants assigned to the treatment condition, a result which was confirmed across three of the study’s dependent variables: real word spelling ($F[1, 165] = 16.341, p < .001, \text{multivariate } \eta^2 = .090$), non-word spelling ($F[1, 165] = 29.212, p < .001, \text{multivariate } \eta^2 = .150$), and reading fluency ($F[1, 165] = 58.348, p < .001, \text{multivariate } \eta^2 = .261$).

KEY WORDS: achievement, computer, Internet, intervention, language arts, quasi-experimental, reading, software
CHAPTER ONE: INTRODUCTION

Literacy and its component skills, the ability to read with fluency and comprehension and write fluently and coherently, are essential to educational attainment across domains: they “[bridge] the gap between learning to read and reading to learn” (Duke, Bennett-Armistead, & Roberts, 2003, p. 226) and provide the key that opens the door to a world of textually-based knowledge. The American system of education, however, has not yet achieved its potential in ensuring that as many Americans as possible enjoy the benefits of literacy. The findings of the National Assessment of Adult Literacy revealed that 43% of adults in the United States scored at basic or below basic levels in prose literacy, or the ability to understand, summarize, make simple inferences, determine cause and effect, and recognize an author’s purpose when presented with texts of moderate density (Kutner, Greenberg, Jin, Boyle, Hsu, & Dunleavy, 2007). Results of the National Assessment of Educational Progress (Grigg, Daane, Jin, & Campbell, 2003) painted an even bleaker picture of American youth. As of 2002, over two-thirds of fourth-grade students scored at basic or below-basic levels of grade-level literacy (Grigg et al., 2003). Research suggests that once children have reached this point in their education, when the focus of instruction has shifted from learning to read to reading to learn (Duke et al., 2003), they are at increased risk for academic failure (Felton & Pepper, 1995; Juel, 1988), often struggling to acquire the content knowledge necessary for academic success.

Reading failure poses a serious threat to a child’s future educational, professional, and social success. This conclusion is well supported in the literature. Kennely and Monrad (2007) identified a statistically significant correlation between low reading
scores and school dropout, and researchers have consistently found that youngsters with reading-related difficulties are disproportionately represented in the juvenile detention system (Rutherford, Bullis, Anderson, & Griller-Clark, 2002; Shelley-Tremblay, O’Brien, & Langhinrichsen-Rohling, 2007), placing them at increased risk for future criminal behavior and social dysfunction.

The vast majority of children at risk for illiteracy can be taught to read with fluency and comprehension, provided they receive developmentally appropriate instruction in the sound-symbol correspondences of spoken and written language (Ehri, Nunes, Willows, Schuster, Yaghoub-Zadeh, & Shanahan, 2001; Snow, Griffin, & Burns, 2005). In particular, sequential instruction in code-based skills, including explicit, systematic phonics has been shown to positively affect the reading and writing abilities of students with reading-related challenges (Ehri et al., 2001; Hatcher, Hulme, & Snowling, 2004; Torgerson, Brooks, & Hall, 2006).

Systematic reading instruction using information and communication technologies (ICT) has been enthusiastically trumpeted as a means by which to promote reading achievement (Savage et al., 2013), and such instruction often requires little or no direct intervention on the part of the classroom teacher (Bishop & Edwards Santoro, 2006). ICT-based instructional programs have been widely adopted in classroom contexts, “generally with an underlying expectation that student learning can improve … through supportive skill instruction with practice” (Cassady & Smith, 2005, p. 362). This sentiment was mirrored in the National Reading Panel’s (NPR) report of 2000, which characterized ICT-based reading instruction as a potentially promising development, allowing students greater opportunity to “interact instructionally with text” than typically
offered by conventional instruction alone (Ch. 6, p. 8).

Some key potential advantages of beginning reading instruction using information and computer technologies are (a) explicit, systematic instruction in the sound-symbol correspondences of spoken and written language (Camilli et al., 2003; Ehri et al., 2001; Torgerson et al., 2006), (b) multimodal instruction to promote recall and retention (Low & Sweller, 2005; Moreno & Mayer, 2007), (c) formative feedback to guide learning and activate prior knowledge (Narciss, 2013), (d) interactivity to promote attention and engagement (Sims, 2000, 2003), and (e) opportunities for mastery learning to enhance achievement (Guskey, 2007, 2012). The question of whether or not ICT-based reading interventions have actually leveraged the potential advantages of the medium, however, remains unresolved, with the authors of large-scale evaluations of software concluding that ICT-based beginning reading programs generally have provided inconsistent and unsystematic instruction (Edwards Santoro & Bishop, 2010; Grant et al., 2012). Furthermore, ICT-based reading instruction remains poorly theorized (Savage et al., 2013) and inadequately researched (Blok, Oostdam, Otter & Overmaat, 2002; Kulik, 2003; Slavin, Lake, Chambers, Cheung, & Davis, 2009; Torgerson & Zhu, 2003), particularly with regard to studies involving participants aged eight years and younger (Lankshear & Knobel, 2003). In the present study a quantitative, quasi-experimental research design is employed to evaluate the efficacy of the MindPlay Virtual Reading Coach (MVRC), a sequential, code-focused online reading intervention, when used to supplement regular reading instruction provided by a classroom teacher.
Statement of the Problem

With some notable exceptions (e.g., Macaruso, Hook, & McCabe, 2006; McMurray, 2013; Savage et al., 2013; Savage, Abrami, Hipps, & Deault, 2009; Savage, Erten, Abrami, Hipps, Comaskey, & van Lierop, 2010), relatively little high quality experimental or quasi-experimental research has been published examining the effects of ICT-based reading interventions on beginning reading achievement (Blok et al., 2002; Kulik, 2003; Slavin et al., 2009; Torgerson & Zhu, 2003). Prominent voices in the field have suggested that teachers and education authorities remain wary of adopting any ICT-based reading program until it has a consistent base of high quality evidentiary support (Slavin et al., 2009; Torgerson, 2007; Torgerson & Zhu, 2003). Through the present study, the author wishes to fill a gap in the existing ICT-based beginning reading intervention literature, while addressing issues of research design and intervention quality that have been inadequately explored in previous intervention research.

Nature of the Study

In the present study, the author relates the results of quantitative research designed to evaluate the efficacy of MindPlay Virtual Reading Coach (MVRC), a sequential, code-focused online reading intervention, when used to supplement regular reading instruction provided by a classroom teacher. The study employed a quasi-experimental design with multivariate data analyses and statistical controls for differences in pre-test achievement scores and relevant demographic variables.

---

1 Among the recommendations of Torgerson (2007) and Torgerson and Zhu (2003) was the conduct of randomized controlled trials (RCTs). Owing to the small sample size in the present study, it was not
Conceptual Framework

The conceptual framework for the present study is based on five research-supported premises. First, failure to develop strong reading skills in early elementary school has pervasive impacts on student achievement in every aspect of education, and those impacts tend to intensify as children progress through school (Felton & Pepper, 1995; Juel, 1988). Second, the vast majority of children can be taught to read with fluency and comprehension when provided high quality, research supported interventions designed to meet individual needs (Ehri et al., 2001; Hatcher, Hulme, & Snowling, 2004; Torgerson et al., 2006). Third, systematic instruction in code-based skills has been shown to positively affect the reading ability of both typically developing students and those with reading-related challenges (Ehri et al., 2001; Snow et al., 2005). Fourth, sequential ICT-based reading interventions of sufficient duration and intensity can and do improve beginning reading achievement (Macaruso et al., 2006; Savage et al., 2009; Savage et al., 2010; Savage et al., 2013), even among students with reading-related challenges (McMurray, 2013). Fifth, ICT-based reading interventions whose content and delivery are consistent with theoretically informed models of reading development and instructional design are likely to be of greatest benefit to students (Savage et al., 2013).

Objectives of the Present Study

The primary objective of the author of the present study was to contribute to the research base in ICT-based beginning reading interventions by reporting the results of quasi-experimental research designed to evaluate the efficacy of MindPlay Virtual Reading Coach (MVRC) (MindPlay Educational Software for Reading, 2015), an ICT-based reading intervention, when used to supplement regular language instruction.
provided by a classroom teacher. MindPlay Virtual Reading Coach offers highly individualized and developmentally sequenced instruction in phonics, fluency, comprehension, vocabulary, and grammar consistent with the recommendations of the NRP (2000). The MVRC software provides multisensory learning, engaging students visually and auditorily, in order to strengthen associations between learned content (Kast, Meyer, Vögeli, Gross, & Jäncke, 2007) and reduce memory demands on individual cognitive systems (Low & Sweller, 2005). Immediate formative feedback is provided to students while they interact with program content, rather than simple corrective feedback, as formative feedback has been shown to increase retention and decrease demands on cognitive processing (Moreno, 2004). Furthermore, MVRC requires mastery (typically to 90%) of initial concepts and skills before new concepts and skills are introduced, ensuring that students do not have gaps in foundational knowledge and promoting sustained retention of learned content (Guskey, 2010).

The secondary objective of the author was to ensure the conformity of the present study with the highest standards for design, analyses, and reporting in educational research. Therefore, the author elected to align the present study with the indicators for high quality experimental and quasi-experimental research identified by Gersten et al. (2005) and published by the Council for Exceptional Children. The present study, as designed, meets all of the relevant\(^2\) essential quality indicators and five of the desirable quality indicators, thus satisfying the criteria for high quality research.

\(^2\) One essential quality indicator was not relevant to the present study, as it applied only to studies involving populations presenting with disabilities or learning difficulties.
Importance of the Present Study

A critical issue of national significance is ensuring that all students have optimal opportunities to develop the reading skills necessary to succeed in school and in life. The present study employs high quality design and data analytic techniques, and an intervention grounded in evidence-based theory and best practices for the promotion of literacy development among beginning readers. It will yield important findings regarding the efficacy of MVRC’s highly interactive, individualized, and developmentally-sequenced mastery model of instruction, and it will contribute to the knowledge base regarding ICT-based beginning reading interventions and the systematic instruction of phonics, fluency, comprehension, vocabulary, and grammar. Furthermore, it will provide educators and administrators with information critical to the selection of effective ICT-based reading interventions for use in the classroom.

Research Question

In the present study, the following research question was addressed: Are there significant mean differences in (a) non-word reading, (b) real word reading, (c) non-word spelling, (d) real word spelling, and/or (e) reading fluency post-test achievement scores of students assigned to use the MVRC online reading intervention in addition to regular classroom reading instruction and those of students from a business-as-usual comparison condition, once the effects of differences in pre-test achievement scores and relevant demographic variables have been accounted for?

Statement of Hypotheses

The author of the present study has identified the following alternative hypothesis ($H_A$) related to the overall effects of the intervention: Statistically significant main effects
of the intervention favoring the treatment group will be detected in overall reading and spelling achievement. The author has identified the following alternative hypotheses ($H_A$) related to the effects of the intervention on individual dependent variables:

- Statistically significant effects of the intervention favoring the treatment group will be detected in non-word reading.
- Statistically significant effects of the intervention favoring the treatment group will be detected in real word reading.
- Statistically significant effects of the intervention favoring the treatment group will be detected in non-word spelling.
- Statistically significant effects of the intervention favoring the treatment group will be detected in real word spelling.
- Statistically significant effects of the intervention favoring the treatment group will be detected in reading fluency.

**Participants**

Participants included 209 students enrolled in eight, second grade classrooms in two public elementary schools in the southwestern United States. Of those, 107 were assigned to the treatment condition, and 102 were assigned to a business-as-usual comparison condition. Owing to attrition and illness, 39 participants had incomplete data sets following post-testing. Ultimately, the treatment condition comprised 89 complete cases, and the comparison condition comprised 81, representing 170 complete cases, or 81.34% of original cases. While overall data loss was below 20%, data loss did not impact each condition equally, as the treatment group retained 83.12% of original cases, while the comparison group retained 79.41% of original cases, producing a differential
attrition rate of 3.71%.

Method

Participating students assigned to the treatment group received MVRC online reading instruction in addition to a scripted program of reading and language arts instruction in the classroom setting. Participants assigned to the treatment group used the software for 30 minutes each day, Monday through Thursday, for a total of two hours per week throughout the regular school year (mid-September through mid-April), with the exception of holidays, school functions, and mandatory state testing days. Participating students assigned to the business-as-usual comparison condition received the same scripted program of reading and language arts instruction but did not receive the MVRC intervention. During the time allotted for the MVRC intervention, students assigned to the business-as-usual comparison condition received additional language arts instruction from their classroom teachers consistent with the scripted curriculum and routine classroom practice. Examiners obtained reading achievement data from each participating student. Pre- and post-test measures included the Test of Silent Word Reading Fluency, Second Edition (TOSWRF-2; Mather, Hammill, Allen & Roberts, 2014) and four tests of the Woodcock Johnson Tests of Achievement, Fourth Edition (WJ IVACH; Schrank, Mather, & McGrew, 2014).

Limitations

The author of the present study wishes to acknowledge several important limitations. Chief among these is the lack of equivalence across demographic factors between the treatment and comparison groups. While the author employed statistical controls to correct for the differences, further research with random assignment to groups
or a non-\textit{post-hoc} matched design should be performed. Also absent was observational or documentary data related to teachers’ instructional practices. Because the number of participating classrooms was modest, this information would have been helpful in detecting potential confounds. Finally, outcome measures included only narrowly focused tasks of reading fluency, word spelling, and word reading. Each of these tasks measured aspects of decoding or encoding, but none assessed the synthesis of these skills in either reading comprehension or sentence or passage writing. Further research should include broader measures, in particular measures of reading comprehension. While decoding ability is essential to fluent reading, the ultimate goal of reading is the comprehension of connected text.

\textbf{Definitions of Terms}
To facilitate reading of the present study, definitions of relevant terms follow:

*Alphabetic language*: a written language in which the symbols used in writing correspond to the sounds of spoken words

*Alphabetic Principle*: understanding of the systematic and predictable relationships between the sounds of speech and the letters of written words in an alphabetic language

*Blend*: a combination of two or more letters used to represent closely but distinctly articulated consonant phonemes within the same syllable

*Code-focused*: instruction emphasizing the sound-symbol relationships of spoken and written language (See *phonics*.)

*Comprehension*: the ability to read a text and understand its meaning

*Computer-aided instruction (CAI)*: instruction delivered or supplemented by the use of computers, tablets, or laptops

*Computer-adaptive*: instruction or testing delivered by a computer that adapts to the performance or actions of the user

*Consolidated alphabetic phase*: the stage of alphabet learning in which children have knowledge of complex phoneme-grapheme relationships, common orthographic patterns, and morphological units

*Decoding*: the ability to use sound-symbol relationships, knowledge of spelling patterns, and knowledge of whole words to read words

*Diphthong*: a sound created by the articulation of two vowel phonemes in a single
sylable

**Digraph**: a combination of two letters representing a single phoneme

**Encoding**: the ability to use sound-symbol relationships, knowledge of spelling patterns, and knowledge of whole words to spell words

**Explicit instruction**: instruction in which content and concepts are presented in a clear and unambiguous manner

**Fluency**: the rate, accuracy, and prosody of reading

**Full alphabetic phase**: the stage of alphabet learning in which children have complete knowledge of common phoneme-grapheme relationships

**Graphemes**: the symbols with which the sounds of spoken language are represented in written language

**Hyperlink**: a virtual link from one file or document to another

**Hypermedia**: typically Internet media; non-linear media including hyperlinks

**Information and computer technologies (ICT)**: multimedia technologies, including computer software and hypermedia

**Initial sound**: a sound falling at the beginning of a word or syllable

**Interactivity**: the way in which content is driven by and adaptive to user activity

**Literacy**: the ability to read with fluency and comprehension

**Medial sound**: a sound falling in the middle of a word or syllable

**Monophthong**: a sound created by the articulation of one vowel phoneme in a
single syllable

*Morphological unit (morpheme):* the smallest grammatically meaningful unit of language

*Non-word:* a nonsense or made-up word, typically presented to determine knowledge of phonetic encoding or decoding

*Orthography:* the conventional spelling system of written language

*Partial alphabetic phase:* the stage of alphabet learning in which children have incomplete knowledge of phoneme-grapheme relationships

*Phoneme:* the smallest perceptually distinct unit of sound in spoken language

*Phonemic awareness:* the ability to identify and manipulate individual phonemes in spoken language

*Phonemic segmentation:* segmenting a spoken or written word by phonemes

*Phonics:* an instructional approach that emphasizes phoneme-grapheme

*Phonological awareness:* the ability to identify and manipulate the sounds of spoken language

*Pre-alphabetic phase:* the stage of alphabet learning in which children lack knowledge of phoneme-grapheme relationships

*Reading failure:* failure to acquire basic reading and spelling skills by grade three

*Scaffolded instruction:* instruction and instructional supports that are decreased as the learner develops knowledge or skill

*Sequential reading instruction:* reading instruction that progresses systematically
through increasingly difficult blocks of content

*Supplementary reading instruction*: reading instruction designed to supplement routine classroom instruction

*Syllable*: a unit of pronunciation containing one vowel sound (i.e., monophthong, diphthong, or triphthong), with or without surrounding consonant sounds

*Syllabification*: the segmentation of a spoken or written word by syllables

*Systematicity*: the degree to which instruction comprises cycles that progress hierarchically through increasingly difficult blocks of content

*Terminal sound*: a sound falling at the end of a word or syllable

*Triphthong*: a sound created by the articulation of three vowel phonemes in a single syllable
CHAPTER TWO: REVIEW OF LITERATURE

This chapter has two main sections: a review of literature and a critical synthesis of experimental and quasi-experimental research. In the first section, the author presents a brief review of literature related to the design and delivery of ICT-based beginning reading interventions. Proposed advantages of ICT-based models of instructional delivery are discussed, as are models for the acquisition and development of phonetic and alphabetic skills. Theoretically informed taxonomies for the evaluation of ICT-based reading interventions are presented, as well as critiques of existing software packages. This review is followed by a critical synthesis of experimental and quasi-experimental research related to ICT-based beginning reading interventions.

The term information and computer technologies (ICT) is used to refer to multimedia technologies, including computer software and online hypermedia. In the present study, the author focuses on a specific form of instructional ICT, sequential reading interventions, or computer-delivered reading instruction that “progress[es] hierarchically through increasingly difficult blocks of content” (Bishop & Edwards Santoro, 2006, p. 62). Such interventions may be used to supplement reading instruction delivered by a teacher, or they may be used to provide targeted practice or tutorial to students in need of remediation. Well-designed ICT-based reading interventions have been shown to promote reading achievement in typically developing beginning readers (Savage et al., 2009; Savage et al., 2010; Savage et al., 2013), low performers (Cassady & Smith, 2005; Macaruso et al., 2006), and those with learning difficulties (McMurray, 2013).
Proposed Advantages of ICT-Based Reading Interventions

Among the proposed advantages of ICT-based beginning reading interventions are (a) multimodal instruction using purposefully linked visual and auditory content, (b) formative feedback to guide learning and activate prior knowledge, (c) interactivity to promote engagement and meaningful learning, and (d) opportunities for mastery learning to promote retention and achievement. An overview of each of these concepts and relevant research follows.

Multimodal instruction. Multimodal instruction is characterized by the presentation of content in more than one sensory mode. The term is used to refer to instruction that incorporates various sensory modalities (e.g., visual, auditory, tactile), but for the purpose of the present study it will refer only to visual and auditory modalities, consistent with research in cognitive psychology (Clark & Paivio, 1991; Fadel, 2008). The multimodal (i.e., visual and auditory; Clark & Paivio, 1991) instruction provided by computer-based media can serve to significantly improve students’ learning (Fadel, 2008). It is hypothesized that instruction using both visual and auditory modalities reduces cognitive load by splitting processing tasks between visual and auditory processing mechanisms (Clark & Paivio, 1991; Low & Sweller, 2005). When visual and auditory information are purposefully and cohesively linked in instruction, demands on cognitive processing are reduced, and retention is improved (Moreno & Mayer, 2007). Research performed by Thompson and Paivio (1994) illustrates this phenomenon. In three separate trials, the authors found that recall was significantly better for picture-sound pairs than for single modality items, leading the authors to conclude that linking visual and auditory stimuli had an additive effect on memory.
Multimodal instruction can also serve to clarify content: if a learner does not understand a concept when presented in one modality (e.g., text [visual]), an additional representation of that concept in another modality (e.g., speech [auditory]) might improve comprehension (Sankey, Birch, & Gardiner, 2012). Furthermore, the presentation of instructional content using multiple modalities accommodates individual perceptual preferences, thus increasing satisfaction with learning and instruction (Sankey, Birch, & Gardiner, 2012). Multi-modal ICT-based instruction has shown promise in improving reading ability, even among students with dyslexia. As an example, Kast, Meyer, Vögeli, Gross, and Jäncke (2007) provided three months of training using a computer-based program designed to “use a multitude of meaningful visual-auditory associations in order to build up the memory strength of graphemes and phonemes” (p. 357) to Swiss children with developmental dyslexia, as well as typically developing peers. At post-test both participants with developmental dyslexia and typically developing participants experienced significant gains in writing skills with comparison to matched controls; however, gains were largest for participants with developmental dyslexia.

**Formative feedback.** Formative feedback constitutes information that is provided in response to learner inputs or interactions that is designed to modify the learner’s knowledge or behavior and improve learning (Shute, 2008). The major aim of formative feedback is to increase the knowledge and skill of the learner by directing and facilitating instructional interactions (Shute, 2008). Computer-based instruction can provide immediate feedback that goes beyond simple correction to provide strategically useful information designed to guide learning, correct misconceptions, and build on existing knowledge (Narciss, 2013). It is hypothesized that formative feedback reduces cognitive
load by activating existing knowledge and directing learners to effective strategies, thus minimizing the amount of cognitive processing directed to ineffective problem solving strategies or misconceptions (Moreno, 2004). Evidence indicates that formative feedback can improve task motivation (Narciss, 2004), retention (Moreno, 2004), and achievement (Moreno, 2004; Narciss, 2004). Formative feedback is especially beneficial for novice or struggling learners, who are more likely to become cognitively overwhelmed during learning (Shute, 2008). As an example, Mioduser, Tur-Kaspa and Leitner (2001) found that Israeli preschool students deemed at risk for learning disabilities who received an ICT-based reading intervention that delivered detailed, individualized formative feedback significantly outperformed matched peers assigned to a conventional teacher-delivered intervention, as well as those assigned to a no-treatment control condition. Significant differences in achievement favoring the ICT-based reading intervention were detected in measures of phonological awareness, word recognition, and letter naming skills.

**Interactivity.** Despite its near ubiquity in the educational literature, interactivity is a loosely defined term (Rose, 1999) that can be used to characterize both conventional and computer-based models of instruction. For the purpose of the present study, however, interactivity will refer simply “to those functions and/or operations made available to the learner to enable them to work with content material presented in a computer-based environment” (Sims, 2000, p. 46). A commonly cited advantage of interactivity is increased engagement: interactive instruction works to focus attention and minimize boredom because the learner must stay alert and on-task in order to interact effectively with the software (Sims, 2003). Furthermore, interactive instruction is responsive to learner inputs, improving instructional efficiency by adapting to individual
needs and strengths (Sims, 2003). Interactive instruction can be particularly beneficial to novice and struggling learners (Sims, 2003). As an example, Segers and Verhoeven (2005) reported the results of highly interactive computer-based pre-reading intervention provided to Dutch students in their second year of kindergarten. Treatment group students used game-based reading software designed to teach rhyming, oral segmentation, and oral blending. While the intervention was provided, on average, only 15 minutes per week during the intervention period, treatment group students enjoyed significant improvements in grapheme knowledge and aspects of phonological awareness immediately following the intervention and upon follow-up testing in the first grade. Likewise, the authors of a Finnish study of a computer-assisted remedial reading intervention offered to grade one students (N = 166) found statistically significant gains in reading achievement among students who completed a total of only five hours of highly interactive ICT-based reading instruction in phonemic awareness, phoneme-grapheme correspondences, reading skills, and fluency (Saine, Lerkkanen, Ahonen, Tolvanen, & Lyytinen, 2011).

**Mastery learning.** Mastery learning constitutes an approach to instruction in which students are required to demonstrate mastery of initial instructional content to a set criterion before more challenging content is presented (Guskey, 2012). Mastery learning is a form of systematic learning, as it presents content in a structured manner (Crews, 2003). Although mastery instruction has demonstrated advantages, including improved achievement and attitudes toward learning (Guskey, 2007), it can be difficult to fully implement in the classroom setting, as it requires a high degree of individualization to be used successfully (Crews, 2003). Well designed computer-based instructional models,
however, are able to deliver highly individualized learning that adapts to students needs (Mooij, 2007), providing opportunities for mastery of foundational concepts and preventing gaps in knowledge likely to undermine future achievement (Guskey, 2007). The benefits of mastery learning appear to be greatest for struggling learners; however, a large body of research demonstrates positive effects on the achievement of students across grades and across content (Kulik, Kulik, & Bangert-Drowns, 1990). An example of the beneficial effects of mastery learning is provided by Vaughn, Serido, and Wilhelm (2006).³ Vaughn et al. (2006) reported the results of a randomized controlled trial (N = 524; participants in grades 1-4) of a sequential ICT-based reading intervention designed using a mastery model of instruction. Treatment group students received two-to-three hours of instruction of ICT-based reading instruction each week, while students assigned to the comparison group received supplementary literacy activities consistent with routine classroom instruction. The authors employed a pre-post design and performed analyses both on post-test scores and gain scores. While statistically significant differences in vocabulary, comprehension, and overall reading ability favoring students assigned to the treatment condition were detected, gains were greatest among English language learners (ELLs) and students designated struggling readers. Measures of effect size were not provided; however, the percentage of treatment group students reading below grade level fell by nearly half from pre-test to post test (from 64.7% to 33.1%).

**Research Related to ICT-Based Reading Interventions**

The question of whether or not the majority of ICT-based beginning reading interventions actually leverage the potential advantages of the medium remains

³ This study (Vaughn et al., 2006) was not included in the formal synthesis that follows this review of literature, as it was not subjected to peer review.
unresolved (Edwards Santoro & Bishop, 2010; Grant et al., 2012), and ICT-based reading instruction remains poorly theorized and inadequately researched (Savage et al., 2013), particularly concerning studies involving participants aged eight years and younger (Lankshear & Knobel, 2003). This dearth of quality research is exemplified by the results of recent syntheses of ICT-based reading intervention literature. In a meta-analysis of over forty studies designed to evaluate the efficacy of ICT-based reading interventions in promoting early literacy achievement, Blok et al. (2002) found only three studies (Barker and Torgersen [1995] and two sub-studies contained in Foster, Erickson, Foster, Brinkman, and Torgesen [1994]) investigating a sequential, code-based English-language reading intervention. While the authors (Barker & Torgesen, 1995; Foster et al., 1994) documented statistically significant effects of their respective interventions favoring participants in the treatment group, the software used (iterations of DaisyQuest) is now obsolete. All of the other studies included in the meta-analysis performed by Blok et al. (2002) were either conducted in a language other than English or involved narrowly-focused computer-assisted interventions, such as the visual and auditory presentation of text, speech feedback, or virtual flashcards. Though Blok et al. (2002) reported a modest corrected overall effect size estimate of +0.19 across studies, they cautioned that this result should be interpreted with care, as many of the studies evaluated were poorly designed, and several compared the effects of multiple software applications, rather than comparing ICT-based interventions to non-computer mediated approaches.

The caution urged by Blok et al. (2002) appears well founded in the light of the results of other syntheses, whose authors applied more stringent inclusion criteria (e.g., requiring experimental designs or matched quasi-experimental designs with both pre- and
post-test measures). In a 2003 synthesis of the randomized controlled trial (RCT) literature in the field, Torgerson and Zhu identified a dozen studies meeting inclusion criteria. Of those, only one (Mitchell & Fox, 2001) included a sequential, code-based ICT-based reading intervention delivered to participants in the early primary grades, and the authors of that study reported no statistically significant effect of the intervention. These results were paralleled by a synthesis of literature commissioned by the National Science Foundation and performed by Kulik in 2003. While the author (Kulik, 2003) identified nine studies of computer-based reading interventions meeting inclusion criteria, only three of those showed statistically significant positive effects favoring treatment group participants, and all three were published in the early 1990s and evaluated the effects of now obsolete software packages.

In a more recent synthesis, Slavin, Lake, Chambers, Cheung, and Davis (2009) identified nineteen studies of computer-assisted reading programs meeting selection criteria; however, only eight were peer-reviewed studies including a now extant sequential ICT-based reading intervention delivered to students in kindergarten through grade three. Of those eight, four were embedded in a larger study commissioned by the Institute of Education Sciences (IES; Campuzano, Dynarski, Agodini, Rall, & Pendleton, 2009). Of the four interventions evaluated by Campuzano et al. (2009), none were found to have statistically significant positive effects on participants’ reading achievement. Unlike Campuzano et al. (2009), the authors of three of the four remaining studies reviewed by Slavin et al. (2009) reported statistically significant positive effects.

---

4 While Mitchell and Fox (2001) found no significant differences favoring the ICT-based reading intervention group when compared to comparable teacher-delivered intervention group, they did find significant differences favoring the ICT-based reading intervention group when compared to a no-treatment control.
of their respective interventions, with effect sizes ranging from +0.17 to +1.05. While otherwise well-designed, the quasi-experimental study in which significant differences were not found favoring the treatment group in measures of word reading (Paterson, Jacobs Henry, O'Quin, Ceprano, & Blue, 2003), lacked a pre-test measure for that construct, putting its findings concerning reading achievement into question. Notably, both of the still extant software packages for which the authors provided statistically significant evidentiary support, Waterford Early Learning Program (Cassady & Smith, 2005; Tracey & Young, 2005) and Lexia Learning Systems software (Macaruso, Hook, & McCabe, 2006), involved a strong emphasis on systematic phonics instruction and the sequential presentation of code-based skills.

**Considerations Related to Research Design**

One explanation for differences in outcomes among ICT-based reading intervention studies might be variables associated with the design and execution of the research. In their 2002 synthesis, Blok et al. noted that design-related variables accounted for the majority of variance across the studies they evaluated and observed that many of the studies they reviewed suffered from serious methodological weaknesses. Even in research involving ICT-based reading interventions purported by their publishers to be of the highest quality (e.g., Waterford Early Learning Program), findings have been inconsistent, with the conclusions of large-scale, government-funded research (e.g., Campuzano et al., 2009) contradicting the findings of smaller studies (e.g., Cassady & Smith, 2005; Tracey and Young, 2007).

While RCT designs have been hailed as the gold standard for educational research (Slavin et al., 2011; Torgerson & Zhu, 2003; Torgerson, 2007), an RCT design is not a
guarantee of methodological strength. As an example, the large-scale RCT performed by Campuzano et al. (2009),\(^5\) which included a tepid evaluation of the Waterford Early Learning Program, had several methodological weaknesses, including a lack of direct experimental control over product use, inconsistent evaluation methods and instruments, unavailability of fidelity data, and an absence of measures closely aligned to the study’s interventions. While the largely favorable research on the same product conducted by Cassady and Smith (2005) and Tracey and Young (2007) was by no means perfect, suffering from flaws in data collection and design,\(^6\) it is nevertheless difficult to characterize the research performed by Campuzano et al. (2009) as methodologically superior.

The preponderance of studies with methodological weaknesses aside, examples of high quality design also exist in the ICT intervention literature. In their generally positive evaluations of Lexia Reading software, Macaruso et al. (2006) and McMurray (2013) employed validated instruments that were well-aligned to their respective interventions. Similarly, both Macaruso et al. (2006) and McMurray (2013) provided and analyzed data related to fidelity of implementation on the relationship between the duration of program usage and students’ achievement. The findings of Macaruso et al. (2006) and McMurray (2013) were mirrored by those of several carefully designed and executed studies evaluating the efficacy of ABRACADABRA, a code-focused sequential online reading program explicitly grounded in developmental models and theory derived

\(^5\) This study (Campuzano et al., 2009) contained data from the second cohort of a larger study performed by Dynarski et al. (2009).
\(^6\) In the case of Cassady and Smith (2005), the authors failed to take into account possible cumulative effects of a teacher professional development intervention being offered simultaneously, and in the case of Tracey and Young (2007), the authors found modest statistically significant between groups at pre-test, which they failed to account for in their analyses.
from contemporary reading research (Savage et al., 2009; Savage et al., 2010; Savage et al., 2013).

**Content and Design of ICT-Based Reading Instruction**

Factors related to research design constitute an important consideration when evaluating the evidence base supporting a given ICT-based reading intervention. Nevertheless, the quality of content and the appropriateness of instructional design remain the primary factors underpinning an intervention’s efficacy or lack thereof. As to the issue of content, the findings of the National Reading Panel (NRP; 2000) identified five key elements of literacy instruction for which there was substantial evidentiary support: (1) phonemic awareness, a subtype of phonological awareness, or understanding of the sound system of spoken language; (2) phonics, knowledge of the correspondence between speech sounds and the orthographic (spelling) patterns in written language; (3) fluency, the speed and accuracy of reading; (4) vocabulary, knowledge of the body of words used in a particular language; and (5) comprehension, the ability to understand the message and meaning of written text. Of these five elements, two are particularly essential to beginning reading instruction: phonemic awareness and phonics.

Understanding of the sound system of spoken language (phonological awareness) and how speech sounds (phonemes) are denoted in written text form the foundation for both decoding (reading) and encoding (spelling) in alphabetic languages (Ehri, 1998; National Reading Panel, 2000). Both phonological awareness, the umbrella skill under which phonemic awareness is found (Badian, 2001; Pufpaff, 2009), and phonics (Adams,

---

7 While *phonemic awareness* and *phonological awareness* are sometimes used interchangeably, *phonological awareness* denotes the understanding of the sound system of spoken language more generally; *phonemic awareness* refers to the ability to detect and manipulate the smallest perceptually distinct units of sound (e.g., the ability to discriminate between /p/ and /b/ in the words *pat* and *bat*) in a spoken language (Pufpaff, 2009).
1990; Camilli et al., 2003; Ehri et al., 2001; Torgerson et al., 2006) comprise a complex set of skills, the importance of which to future reading achievement is difficult to overstate. The National Early Literacy Panel (NELP; 2008), in a comprehensive review of high quality research in early literacy, found that phonological awareness and knowledge of the alphabetic principle (the systematic, predictable relationships between speech sounds [phonemes] and written letters [graphemes]) were the two best predictors of early reading achievement.

A convergence of evidence demonstrates the importance of systematic, explicit instruction in the sound-symbol correspondences of spoken and written language to reading achievement. In a synthesis of 38 experimental and quasi-experimental studies designed to evaluate the efficacy of systematic phonics interventions, Ehri et al. (2001) concluded that explicit instruction in “[s]ystematic phonics helped children learn better than all forms of control instruction” (p. 393), and that it was effective not only in providing beginning reading instruction, but also in preventing and remediating reading difficulties. These findings were largely mirrored in a replication (of the study performed by Ehri et al. [2001]) conducted by Camilli et al. (2003) two years later. While Camilli et al. (2003) found that even non-systematic phonics instruction benefited learners, the authors noted that “the advantage of systematic phonics instruction over some phonics instruction [was] significant.” Similarly, in a 2006 meta-analysis of the RCT intervention literature, Torgerson et al. found that “[s]ystematic phonics instruction within a broad literacy curriculum…[had a] statistically significant positive effect on reading accuracy” (p. 8), which extended both to typically developing learners and those at risk for reading difficulties.
For code-based skills to be taught effectively, however, they must be introduced in a developmentally appropriate manner. The general trajectory of the acquisition of phonological and alphabetic skills has been well defined in the literature, yet there remains a “need for progression within ICT activities that reflects these qualitative changes in development” (Savage et al., 2013, p. 312). A brief overview of two research-based models for the development of phonological and alphabetic skills follows.

**Research-based developmental models.** In a 2009 synthesis of the literature evaluating the development of phonological sensitivity, Pufpaff laid out a developmental sequence for the acquisition of phonological skills derived from an extensive review of research in early literacy. These skills progress from the detection and production of rhymes, to the segmentation of sentences by words, to the syllabification of compound words (e.g., <cow·boy>), and then the syllabification of multisyllabic words (e.g., <mi·tten>). After these phonological skills have been acquired, development progresses to a subset of phonological awareness characterized as phonemic awareness, or the ability to isolate and manipulate individual speech sounds. The developmental sequence of phonemic awareness skills progresses from blending of sounds (e.g., /k/ /æ/ /t/ to cat), to sound-to-word matching (e.g., Does cat end with the /t/ sound?), to word-to-word matching (e.g., Does cat begin with the same sound as cap?), to phoneme isolation (e.g., What is the beginning sound in cat?), to phoneme segmentation (e.g., Say cat one sound at a time.), to phoneme deletion (e.g., Say cat without the /t/ sound.), to phoneme substitution (e.g., Say cat with the /p/ sound instead of the /t/ sound.), and finally to phoneme reversal (e.g., Say cat backwards, so the beginning sound is at the end and the ending sound is at the beginning [i.e., tack]). It is important to note, however, that there
is individual variation in the acquisition and development of phonological skills, and the research suggests that children do not develop skills in a rigid sequence, but in often overlapping stages (Invernizzi & Tortorelli, 2013); therefore, individual skills need not be taught in a lockstep fashion, but in a responsive manner that is adaptive to the needs of the individual learner.

In an alphabetic language, another major instructional consideration is the introduction of the symbol system of the language, a modified Latin alphabet, in the case of English. Phonological awareness develops in tandem with alphabet knowledge, and children use insights derived from their knowledge of the alphabet to identify the sounds in spoken and written words (Invernizzi & Tortorelli, 2013). In her seminal work on the acquisition of word learning, Ehri (2005) proposed a four-phase model of the development of alphabetic skills based on her own research and that of others in the field. In the first of the four phases, the pre-alphabetic phase, learners have not yet begun to form cognitive connections between letters and the sounds that represent them. By the second phase, characterized as partial alphabetic, however, children have begun to form connections between phonemes and graphemes. Knowledge of letter names, as an example, may be used to infer letter sounds (e.g., T begins with the /t/ sound, or the initial sound in its letter name). Invernizzi and Tortorelli (2013) noted that in this phase, letters whose initial sounds correspond to the phonemes they most typically produce (e.g., B, /b/; D, /d/; K, /k/) should be taught first, followed by letters whose terminal sounds correspond to the phonemes they most typically produce (e.g., F, /f/; M, /m/; N, /n/), and finally, letters whose names do not provide consistent phonetic cues (e.g., C, W). Likewise, letters whose graphical representations are easily distinguished should be
taught before letters whose graphical representations are easily confused, and letter sounds that occur most frequently should be taught before those that occur least frequently (Invernizzi & Tortorelli, 2013).

As children’s facility with simple phoneme-grapheme relationships improves, they develop from *partial alphabetic* to *full alphabetic* knowledge, or the ability to recognize and segment all common sounds within a one-syllable word (Ehri, 2005). This phase is characterized by knowledge of all major phoneme-grapheme relationships, including vowels and digraphs (i.e., a pair of characters used to write a single phoneme [e.g., *ch*]). Full alphabetic knowledge can then be built upon to develop facility with more complex, chunked units of written text, including blends (two consonant phonemes articulated in close succession within the same syllable [e.g., *st*]), onsets (the initial phonological unit in a syllable) and rimes (the terminal phonological unit of a syllable, including the vowel and any consonants that follow it), morphological units (grammatically meaningful units of language [e.g., *-ing*]), and other common orthographic patterns, until a learner has achieved *consolidated alphabetic* knowledge. Consolidated alphabetic knowledge reduces demands on memory (i.e., phonological units are chunked, rather than individually decoded, and many whole words are represented in memory) and forms the basis for fluent and accurate reading, as well as improved reading comprehension and acquisition of novel vocabulary (Ehri, 2005).

**Presentation and delivery of content.** While Ehri’s (2005) model provides a research-supported outline of the acquisition and development of decoding ability, reading instruction must nevertheless be presented and delivered in a manner responsive to individual needs (Invernizzi & Tortorelli, 2013). The presentation and delivery of
content are therefore matters of particular concern when instruction is provided by computer software, which responds to student interactions based upon programmed algorithms that lack the benefit of human insight or experience.

Drawing on theoretical models of reading skills development and input from experts in the field, Grant et al. (2012) created a developmentally sequenced taxonomy of reading skills with which to evaluate ICT-based preschool and early primary reading software. The taxonomy was unusual in its comprehensiveness, covering skills and sub-skills related to (a) concepts about print, (b) alphabetic knowledge, (c) phonological awareness, (d) phoneme-grapheme relationships, (e) phonics, (f) syntactic awareness, (g) decoding, (h) fluency, and (i) comprehension. Using their novel taxonomy, Grant et al. (2012) evaluated commercially available software packages designed to teach beginning reading skills. Of those, only five included instruction in synthetic phonics, and none included instruction in phonemic segmentation. The authors reported that among the interventions evaluated, “the presentation of skills was by no means systematic or consistent across the software levels or in congruence with the reading taxonomy expectations,” and “fewer skills than expected were being taught through the software programs” (Grant et al., 2012, p. 333). Typical weaknesses in the software included a lack of developmentally appropriate and sequential presentation of instruction, an absence of scaffolded learning or computer-adaptivity in response to student performance, insufficient feedback in response to errors, and inconsistency in the quality of instruction from skill to skill (Grant et al., 2012).

These results were paralleled by those of a review of software performed by Edwards Santoro and Bishop (2010). In 2006, Bishop and Edwards Santoro created a
framework for evaluating beginning reading software that took into consideration both content and delivery, providing research-supported criteria for evaluating three distinct aspects of interface design and four distinct aspects of instructional design, as well as the comprehensiveness of phonological and alphabetic content. Concerning interface design, the model proposed by Bishop and Edwards Santoro (2006) focused on the quality of (a) aesthetics, the way texts, graphics, animations, and sounds are designed to enhance user experience; (b) operational support, the way the program provides direct and indirect support to users, allowing them to use the program with minimal help from teachers or other adults; and (c) interactions, the way in which content is driven by and adaptive to user activity. As to instructional design, Bishop and Edwards Santoro (2006) identified (a) systematicity, the degree to which instruction comprises “cycles that progress hierarchically through increasingly difficult blocks of content” (p. 62); (b) instructional support, the appropriateness and timeliness of the provision of content to enhance learning; (c) assessment, the degree to which user performance is tracked, evaluated, and used to inform and adapt goals and content; and (d) motivation, the degree to which the program is relevant, stimulating, appropriately challenging, and rewarding, as key factors in high quality design.

Using their framework, Edwards Santoro and Bishop (2010) evaluated 21 commercially-available reading software packages, concluding “… as a whole, [their] reviewed sample did not meet research-based criteria for interface, instructional design,

---

8 Category names have been altered slightly (e.g., Systematic to Systematicity) to preserve grammatical parallelism.

9 Because the model proposed by Bishop and Edwards Santoro (2006) focused exclusively on beginning reading interventions for at-risk readers, criteria for evaluating program content were limited, including only (a) phonological skills, including phonemic manipulation and segmentation, and (b) alphabetic skills, including phoneme-grapheme relationships and decoding. For a more comprehensive taxonomy of reading skills as they apply to ICT-based reading interventions, see Grant et al. (2012).
and beginning reading content …” (p. 114). Among their findings was a negative
correlation between quality of interface design and quality of content, as well as a lack of
developmentally appropriate instruction and structured feedback. Only three of the 21
programs Edwards Santoro and Bishop reviewed offered any degree of explicit
instruction in code-based skills, and none “scored above ‘neutral’ for the effective
practice activities indicator, ‘The program uses informative and instantaneous feedback
messages to support content learning’” (2010, p. 112), suggesting considerable room for
improvement among ICT-based beginning reading programs.

Though the results of the reviews of software performed by Grant et al. (2012) and
Edwards Santoro and Bishop (2010) paint an unflattering picture of ICT-based beginning
reading interventions generally, neither Grant et al. (2012) nor Edwards Santoro and
Bishop (2010) evaluated all of the interventions commonly used in the public schools.
Examples of widely used programs that were not included in either review are
HeadSprout, iReady, MVRC, Plato Focus, and Waterford Early Learning Program. It
should also be noted that both Grant et al. (2012) and Edwards Santoro and Bishop
(2010) found considerable variation in quality among the programs they reviewed, and
their appraisals of ICT-beginning reading software were not universally negative. Grant
et al. (2012), for example, praised several programs for providing students with
systematic and explicit instruction in developmentally appropriate content, scaffolded
learning, multiple opportunities for practice, and identified achievement goals. Similarly,
Edwards Santoro and Bishop (2010) identified a modest number of programs providing
both high quality content and instructional design, noting that the best among them
“[supplied] adequate opportunities for learners to practice new skills, [presented] the
same requirements in embedded activities as presented in the instructional sequence, [required] mastery of skills before moving onto new skills, and [moved] systematically through the word-reading skill hierarchy” (p. 115).

While reviews of software of the types performed by Grant et al. (2012) and Edwards Santoro and Bishop (2010) serve to identify software consistent with theoretically informed models for the content and design of beginning reading instruction, it is intervention research that provides the best evidence of an intervention’s efficacy. In the subsequent section, the author presents a critical synthesis of the experimental and quasi-experimental research on information and computer technology-based interventions designed to promote literacy achievement among students in kindergarten through grade three. The purpose of this synthesis was to determine whether sufficient evidence existed to support any information and computer technology-based intervention as a promising practice to promote improved reading achievement among students in the early primary grades.
ICT-Based Reading Interventions: A Critical Synthesis of the Literature

In the information age, strong reading skills are key to unlocking opportunities for educational and professional advancement (Eisenberg, Lowe, & Spitzer, 2004). As the economy moves further from a model based on the manufacture of physical goods toward a model based on the production and dissemination of knowledge-based goods (World Bank, 2004), it is critical that all students have optimal opportunities to develop the reading skills necessary to succeed. Unfortunately, the American system of education has not yet achieved its potential in ensuring that as many Americans as possible enjoy the benefits of literacy. According to the National Assessment of Adult Literacy (NAAL), a representative measure of the English language literacy skills of American adults sponsored by the National Center for Education Statistics, ninety-three million adults in the United States scored at basic or below basic levels in prose literacy, or the ability to read and comprehend continuous texts, at the time of sampling in 2003 (Kutner, Greenberg, Jin, Boyle, Hsu, & Dunleavy, 2007). This group comprised over forty percent of the total adult population in the United States -- four in ten Americans who were unable to understand, summarize, make simple inferences, determine cause and effect, or recognize an author’s purpose when presented with texts of moderate density (Kutner, Greenberg, Jin, Boyle, Hsu, & Dunleavy, 2007).

While the negative impacts of illiteracy are perhaps most severe for older youth and young adults (Kennelly & Monrad, 2007; Rutherford, Bullis, Anderson, & Griller-Clark, 2002; Shelley-Tremblay, O’Brien, & Langhinrichsen-Rohling, 2007), the roots of illiteracy are found early in childhood (McCardle, Scarborough, & Catts, 2001). Children who struggle in the beginning stages of reading development are likely to face increasing
obstacles to the development of literacy and its component skills as they grow older (Spira, Bracken, & Fischel, 2005). Those who have not mastered basic reading skills by grade three, when the focus of instruction usually shifts from learning to read to reading to learn (Duke, Bennett-Armistead, & Roberts, 2003), are beset by academic problems that tend to intensify in the later grades. Because they lack the ability to read fluently and with comprehension, such students often fail to gain conceptual knowledge from written texts and consequently fall further behind their literate peers with each passing year (Felton & Pepper, 1995; Juel, 1988; Stanovich, 1986).

Code-based reading instruction using information and communication technologies (ICT) has been widely advanced as a means by which to promote reading achievement (Savage et al., 2013). However, little has been done to “tie computer-mediated reading instruction to contemporary theoretical models of reading acquisition and to coherent pedagogical models for technology” (Savage et al., 2013, p. 310), and the field remains under-researched, especially with regard to interventions delivered to children in the early primary grades (Lankshear & Knobel, 2003). There is also a lack of evaluative research examining the evidence base supporting the use of ICT-based beginning reading programs. A handful of systematic reviews have been performed, but none has focused exclusively on sequential ICT-based reading instruction in the early primary grades. In the present synthesis, the author presents an up-to-date evaluative synthesis of the literature in sequential ICT-based reading instruction in the early primary grades. The primary objective of the author is to identify high quality ICT-based beginning reading interventions that can be used at scale in typical school settings. The secondary objective of the author is to determine which features, if any, are common to high quality ICT-
based programs of beginning reading instruction.

**Purpose**

The purpose of this synthesis is threefold: (a) to build upon previous research by including articles that have been published in the past several years; (b) to narrow the scope of the research by selecting for inclusion only those studies focusing on sequential, primary reading instruction that can be used at scale, as opposed to interventions of limited utility in typical educational settings; and (c) to increase the level of scrutiny to which the articles are subjected by evaluating them using the Anchored Matrix for Evaluating Experimental Research (AMEER), a novel instrument for the assessment of experimental and quasi-experimental research based on the quality indicators identified by Gersten et al. (2005).

**Prior Research**

In a large-scale meta-analysis of computer-assisted reading instruction published between 1990 and 2002, Blok, Oostdam, Otter and Overmaat (2002) identified a total of 42 studies meeting inclusion criteria. Of those, only two (Barker & Torgesen, 1995; Foster, Erickson, Foster, Brinkman, and Torgesen, 1994) involved a sequential, code-focused English-language reading intervention, and that intervention (DaisyQuest) is now obsolete. While Blok et al. (2002) calculated a modest overall effect size of +0.19 across the studies they analyzed, they cautioned that many of the studies were poorly designed, and design-related variables, such as the time of pretesting and the language of instruction, accounted for the majority of variance. In another meta-analysis, Soe, Koki, and Chang (2000) found 17 studies performed between 1982 and 1999 that met their broad inclusion criteria, but only two of those included participants in the early primary
grades, and none included a sequential ICT-based reading intervention offering systematic instruction in code-based skills. Furthermore, the mean effect size across the studies analyzed was negligible (+0.13). Kulik (2003), in a comprehensive synthesis financed by the National Science Foundation, found a total of nine controlled studies of reading software programs; however, only three of those showed significant effects of their respective interventions, and the median effect of the “nine studies was to raise students reading scores by 0.06 standard deviations, a trivial increment” (p. v). Similarly, Torgerson and Zhu (2003) identified very few (12) ICT-based intervention studies meeting inclusion criteria. Of those studies, only one included a condition in which sequential, code-focused instruction was provided (as opposed to virtual storybooks or narrowly-focused fluency or comprehension activities), and its authors (Mitchell & Fox, 2001) reported that the intervention (iterations of DaisyQuest) was not superior to comparable teacher-delivered instruction.10

In a more recent synthesis of research involving reading interventions delivered to struggling readers in kindergarten through fifth grade, Slavin, Lake, Davis, and Madden (2011) found a total of 14 qualifying studies examining ICT-based reading interventions; however, only five of those studies had been published within the prior decade, and all nine of the others investigated now obsolete software programs. Of the more recently published studies, only two focused on students in the early primary grades (i.e., kindergarten through grade three), and the authors of only one of those (Macaruso, Hook, & McCabe, 2006) reported positive effects of the intervention.11 In another recent

---

10 It is worth noting that Mitchell and Fox (2001) found significant differences favoring the ICT-based reading intervention group when compared to a no-treatment control.

11 The study failing to report positive effects was the large-scale RCT performed by Campuzano, Dynarski, Agodini, and Rall (2009), which is discussed in greater depth in the subsequent paragraph.
synthesis of the reading intervention literature, Slavin, Lake, Chambers, Cheung, and Davis (2009) found 19 ICT-based beginning reading intervention studies meeting selection criteria; however, six are now well over fifteen years old, and three were not peer reviewed. Of the remaining ten, four were embedded in a larger study commissioned by the Institute of Education Sciences (IES; Campuzano, Dynarski, Agodini, & Rall, 2009). The study included both a first grade and a fourth grade cohort, subsets of each of which were subjected to one of four different reading interventions. While the study was a large randomized controlled trial (RCT), it had important methodological limitations: None of the four software programs studied by Campuzano et al. (2009) were evaluated individually until the second year of the project because “in the first year, the study operated under the guideline that individual product effects would not be reported” (p. 37); however, “classroom observations were not conducted [by Campuzano et al. (2009)] in the second year” (p. 37), and implementation data were not reported, making it virtually impossible to evaluate the circumstances of programs’ use or examine their relative merits.

Unlike Campuzano et al. (2009), the authors of five of the six remaining studies reviewed by Slavin et al. (2009) reported positive effects of their respective interventions, with effect sizes ranging from +0.17 to +0.71. Notably, all five were published after 2000. Two focused on web-delivered multimedia lessons embedded in teachers’ regular whole-class instruction (Chambers et al., 2006; Chambers et al., 2008), rather than individual ICT-based reading instruction; however, three others (Cassady & Smith, 2005;
Macaruso et al., 2006; Tracey & Young, 2007) involved code-based primary reading instruction delivered to students using personal computers.12

Similarly, the authors of a few studies published within the past ten years have reported positive effects of narrowly focused multimedia language arts interventions. As an example, Englert, Zhao, Collings, and Romig (2005) examined the effects of an online cloze procedure on participants’ acquisition of sight words compared to those of an identical paper-and-pencil procedure. The authors detected a statistically significant positive effect of the intervention on sight-word recognition; however, the number of participants was small (N = 31), and the study did not have a true control in either of its phases. In another more robustly designed study, Chambers et al. (2011) found statistically significant positive effects of computer-facilitated small group tutoring among struggling readers in grade one, but not among students in grade two. Two other studies of the same intervention also documented promising results. In a study by Chambers et al. (2008a), low achieving first graders who participated in the computer-facilitated small group tutoring at desired levels outperformed participants in a comparison group on three of four measures of reading achievement. Similarly, first graders who received ICT-based small group tutoring and other technological enhancements to their reading curriculum outperformed control group participants on five measures of reading achievement (Chambers et al., 2008b). The intervention upon which both studies focused, however, was as a targeted program in code-focused oral and written language skills, allowing a single trained tutor to work with several students at

12 The study whose authors (Paterson, Jacobs Henry, O'Quin, Ceprano, & Blue, 2003) did not report significant effects of the intervention was a mixed-methods quasi-experimental design that lacked a pre-test measure of decoding. Therefore, any statistical inferences derived from reported post-test decoding scores are of limited validity.
once. Chambers et al. explained “[r]ather than replacing the tutor, the program was designed to increase program fidelity by assisting tutors and students in each of the three phases of tutoring -- planning, instruction, and assessment” (2008b). It contained a number of components requiring direct monitoring and instructional intervention from a trained tutor, and it was designed to be used within the context of a larger teacher-delivered scripted reading curriculum, limiting its value to teachers in schools that have not adopted the scripted curriculum.\footnote{It should be noted that the curriculum, Success for All, in which the tutoring program was embedded has significant evidentiary support. (See Slavin & Madden [2012] for a good overview of the program and its evidentiary base.) Therefore, this observation is not intended to diminish the program or its utility in any way.}

The authors of several recent studies have also found statistically significant positive effects of sequential ICT-based reading interventions on very young children’s acquisition of phonological skills and phonemic awareness.\footnote{While generally well-designed, these studies were not eligible for inclusion in the present synthesis due to a lack of measures of decoding administered pre- and post-test, or other disqualifying design-related factors.} In a matched quasi-experimental study of the effects of a sequential ICT-based reading intervention (Lexia Reading) on pre-literacy skills among kindergarten students, Macaruso and Walker (2008) found statistically significant differences in phonological awareness skills favoring the treatment group, particularly among students with the lowest pretest scores. In results of a 2011 follow-up study using the same intervention, Macaruso and Rodman reported that preschoolers in both treatment and control classes showed improvements in pre-reading achievement, but preschoolers in the treatment group experienced larger gains than those in the control group, particularly in measures of phonological awareness, in which the differences achieved statistical significance. Similarly, Mitchell and Fox (2001) found that struggling kindergarten and first-grade students who received five
hours of training on an ICT-based reading intervention (DaisyQuest) providing instruction and practice in rhyme identification, phonological segmentation, and phonological blending performed significantly better on measures of phonological awareness than did students assigned to non-instructional computer activities; however, they did not outperform students whose teachers delivered comparable instruction.

**Rationale**

While the results of previous syntheses of research in ICT-based reading instruction have been mixed, their findings point to improvements in the quality of reading instructional software over time. Of the syntheses of literature in which the authors examined only studies published ten or more years ago, the authors determined that the evidence in support of ICT-based reading instruction was unpromising (Blok et al. 2002), negligible (Kulik et al., 2003; Soe et al., 2000), or insufficient (Torgerson & Zhu, 2003). The authors of a more recent synthesis (Slavin et al., 2009), however, found several studies whose authors reported positive effects of ICT-based reading interventions delivered to students in the early primary grades. The results of this synthesis are paralleled by those of a handful of narrowly-focused multimedia intervention studies (Chambers et al., 2008a; Chambers et al., 2008b; Chambers et al., 2011), as well as research on ICT-based reading interventions delivered to preschool and kindergarten students (Macaruso & Walker 2008; Macaruso & Rodman, 2011), whose authors reported statistically significant positive effects of their respective interventions.

Notably, the authors of the most recent syntheses of ICT-based reading intervention literature (Slavin et al., 2011; Slavin et al., 2009) did not formally evaluate the quality of the research that they examined. While their selection criteria excluded
designs without a control or comparison condition, as well as designs whose measures or procedures were likely to be biased or lack validity, some of the studies that Slavin et al. (2009) reviewed for their synthesis (e.g., those embedded in the larger study performed by Campuzano et al., 2009) had clear methodological limitations. According to the guidelines suggested by Gersten et al. (2005) and published by the Council for Exceptional Children, the careful evaluation of study quality is essential to determining whether or not a practice or intervention can be considered research-supported.

Previous syntheses provided valuable overviews of the research in the area of ICT-based reading interventions, but none were designed to critically evaluate the state of the evidence in support of individual interventions using an established framework (e.g., the guidelines suggested by the What Works Clearinghouse [What Works Clearinghouse, 2008], CONSORT [Shultz, Altman, & Moher, 2010] or the Council for Exceptional Children [Gersten et al., 2005]). Furthermore, the majority of these syntheses (Blok et al., 2002, Kulik et al., 2003; Soe et al., 2000; Torgerson & Zhu, D., 2003) are now over a decade old. Given the rapid pace of innovation in the field and the positive trajectory of the findings of ICT-based reading intervention research, an up-to-date, evaluative synthesis is warranted.

**Research Questions**

In support of this purpose, the author sought to answer the following research questions:

1. Is there sufficient evidence to designate any ICT-based program of beginning reading instruction as a research-supported practice?
(2) What features, if any, are common to high quality ICT-based programs of beginning reading instruction?

Method

Inclusion Criteria

To be included in this synthesis, articles must have satisfied the following criteria:

(1) Publication Criteria
   i. Published in a nationally-circulated academic journal
   ii. Published in English
   iii. Published between 01 January 2000 and 01 December 2014

(2) Design Criteria
   i. Conducted using an experimental or quasi-experimental design with random assignment or statistical adjustment for pre-test differences between groups (e.g., analysis of covariance or repeated measures)
   ii. Included at least one control or comparison condition whose participants did not receive any ICT-based reading intervention
   iii. If quasi-experimental or a randomized controlled trial with fewer than 30 units at the level of randomization, included both pre- and post-test data for all key measures, with at least one standardized measure of decoding
   iv. If a randomized controlled trial with at least 30 units at the level of randomization, included post-test data for all key measures for all participants, with at least one standardized measure of decoding
   v. Included inferential statistical analyses of data (e.g., analysis of variance, multivariate analysis of variance)
(3) Participant Criteria

i. A minimum of 30 participants across conditions, to ensure adequate sample size for statistical inference

ii. Participants in academic kindergarten settings through grade three; studies with participants in other grades qualify, provided they include separate analyses for participants in academic kindergarten settings through grade three

iii. Participants in general education classes, regardless of disability status (not to include foreign language, language learning, language immersion, or self-contained special education classes)

iv. Studies where participants self-selected into interventions (e.g., voluntary after-school programs) or were selected by others based on some criterion (e.g., disability status) were excluded, unless assignment to groups was performed after selection. Likewise, studies in which participants were selected or excluded based on levels of use of an existing intervention were excluded (e.g., Macaruso & Rodman, 2011), as factors associated with differential use might also impact study outcomes.

(4) Intervention Criteria

i. Commercially available or distributed through a government, educational, or research institution

ii. Designed to be delivered individually to students via personal computer, laptop, or tablet (not to include synchronous or asynchronous remote teaching)
iii. Designed to deliver English-language reading instruction

iv. Designed to deliver sequential reading instruction, as opposed to narrowly-focused programs primarily intended to supplement existing curricula (e.g., programs including only assessments, videos, flashcards, spoken text, or virtual books)

v. Designed, at least in part, to teach code-based reading skills (phonics), as opposed to interventions exclusively focused on fluency, comprehension, vocabulary, phonemic awareness, or oral language

vi. Replicable in realistic school settings (i.e., Programs that could not be used without extensive, direct instructional assistance from a teacher or tutor were excluded.)

A decision was made to examine only articles published in nationally-circulated, refereed academic journals because such articles had been subjected to peer review and were therefore likely to be of higher quality than articles published in non-refereed outlets. While this decision surely introduced a degree of publication bias into the analysis, the author felt that the peer review process was a necessary check on the potential for experimenter bias inherent in a field rife with commercially commissioned and independently published research. Owing to the rapid pace of change in instructional technology, the author chose to examine only articles published since January 01, 2000. This decision was bolstered by a comparison of the results of older syntheses (Blok et al., 2002, Kulik et al., 2003; Soe et al., 2000), whose authors found few, if any, studies demonstrating positive effects of ICT-based reading interventions delivered to students in
the early elementary grades, to those of a more recent synthesis (Slavin et al., 2009), whose authors found several.

The author chose to include only articles relating the results of experimental or quasi-experimental research with a control, comparison, or baseline condition. While this criterion falls short of the *gold standard* of randomized controlled trials suggested by the What Works Clearinghouse (Coalition for Evidence-Based Policy, 2006), it is a more realistic standard for research in elementary education, where true randomized assignment to groups is often difficult, if not unethical or impossible (Odom et al., 2005). The author excluded studies that included only students in language learning settings, as such studies’ results were likely to be impacted by the participants’ linguistic status. The authors included only studies involving an English language reading intervention because orthographic depth varies from language to language, and the depth of a language’s orthography has been shown to impact the ease and speed with which children acquire literacy (Brunswick, McDougall, & de Mornay Davies, 2010).

As the early mastery of basic reading skills is essential to later school success (Spira, Bracken, & Fischel, 2005), a decision was made to examine only studies including interventions administered to participants in kindergarten through grade three. Likewise, the author chose to evaluate only interventions that included sequential instruction in code-based skills, as systematic, code-based reading instruction has proven highly effective in promoting reading achievement among students in the early elementary grades (National Reading Panel, 2000). The author also elected to include studies whose interventions are currently accessible to educators, either commercially or through other distribution channels, and would not require a degree of personnel
commitment likely to create barriers to implementation in realistic school settings (i.e., programs that could not be used without extensive, direct instructional assistance from a teacher or tutor).

Because the aim of the author was to identify ICT-based reading interventions that could be used in school settings with minimal direct intervention on the part of classroom teachers, certain of the interventions included in previous syntheses did not qualify for analysis in the present study. Furthermore, the author also chose to exclude the four studies embedded in the IES-commissioned research performed by Campuzano et al. (2009). This research performed by Campuzano et al. (2009) had several important methodological limitations, including a lack of direct experimental control over product use, inconsistent evaluation methods and instruments, unavailability of fidelity data, and the absence of measures closely aligned to the study’s interventions.

**Selection Process**

Articles meeting selection criteria were identified through an online database search process and reference chasing. Preliminary searches were conducted using the Education Resources Information Center (ERIC; EBSCO Industries, 2014), Academic Search Complete (EBSCO Publishing, 2014), and PsychArticles (American Psychological Association, 2014). Search terms included reading AND *experiment AND computer; reading AND *experiment AND Internet; reading AND *experiment AND online; reading AND *experiment AND software; reading AND intervention AND computer; reading AND intervention AND Internet; reading AND intervention AND online; and reading AND intervention AND software. Search results were narrowed by grade level, design, and publication type. Following database searches, the author
reviewed the abstracts of the articles referenced in the reviews of literature and meta-analyses examined for this synthesis, as well as articles referenced in studies selected via online search. Articles whose abstracts were consistent with selection criteria were added to a virtual clipboard. They were then subjected to further examination and, if appropriate, review. This process yielded a total of seven articles consistent with selection criteria. An overview of each of these articles is presented in Table 1.

**Instrument**

To examine selected research, the author employed the quality indicators identified by Gersten et al. (2005) and published by the Council for Exceptional Children. Gersten’s model was chosen because it was designed with the particular constraints and demands of education research in mind. It has been widely adopted in the field and has been cited in numerous other syntheses (e.g., Browder, Wakeman, Spooner, Ahlgrim-Delzell, & Algozzine, 2006; Chard, Ketterlin-Geller, Baker, Doabler, & Apichatabutra, 2009; Justice, Mashburn, Hamre, & Pianta, 2008).

In their 2005 article, Gersten and co-authors outlined a total of 19 quality indicators, 11 of which they deemed to be essential to research and eight of which they deemed desirable. According to Gersten et al. (2005), for a quasi-experimental or experimental study to be considered of acceptable quality, it must meet all but one of the essential quality indicators and at least one of the desirable quality indicators. For a quasi-experimental or experimental study to be considered high quality, it must meet all but one of the essential quality indicators and at least four of the desirable quality indicators. For a practice to be considered promising, there must be at least four acceptable quality studies, or two high quality studies, that support the practice.
While Gersten et al. (2005) identified essential and desirable quality indicators for experimental and quasi-experimental research, they did not provide an instrument with anchored intervals by which to assess research for the presence or absence of quality indicators. For the purpose of this synthesis, therefore, a rubric with anchored intervals was created. This rubric, characterized as the Anchored Matrix for Evaluating Experimental Research (AMEER) as far as possible, mirrored the language used by Gersten et al. in *Quality Indicators for Group Experimental and Quasi-Experimental Research in Special Education* (2005). The rubric was designed such that articles could be assessed for the degree to which they demonstrated the presence of both essential and desirable quality indicators. For each of the essential quality indicators, an anchored scale was created, with values ranging from 0 to 3. A value of 0 corresponds to the absence of the specified quality indicator, whereas a value of 3 corresponds to article content that exceeds the specified quality indicator. Article content that meets but does not exceed a specified quality indicator corresponds to a value of 2, and article content that approaches but does not meet a specified quality indicator corresponds to a value of 1. A scale without an intermediate point was chosen to eliminate the possibility of a no-choice choice. The AMEER is presented in Appendix A.

**Rating Procedure**

The author carefully evaluated each of the articles included in this synthesis using the rubric presented in Appendix A. After rating, the results of each study’s evaluation were reviewed to determine whether the study could be considered of acceptable or high quality per the guidelines established by Gersten et al. (2005). Studies achieving a score of two or higher for any criterion were considered to have met the specified quality
indicator. Studies meeting at least 10 of the essential quality indicators and at least four of the desirable quality indicators were deemed of high quality; studies meeting at least 10 of the essential quality indicators and at least one of the desirable quality indicators were deemed of acceptable quality; and studies meeting fewer than ten of the essential quality indicators, or studies meeting at least ten of the essential quality indicators but none of the desirable quality indicators were deemed not to have met the quality indicators for acceptable or high quality research. In cases in which a particular quality indicator was inapplicable to a study (e.g., information confirming disability status for research in which individuals with disabilities did not participate), it was excluded from analysis, and scales were adjusted accordingly.

**Results**

**Articles Meeting Selection Criteria**

Following a careful search using the selection procedures outlined in the Method section of this article, a total of seven articles meeting selection criteria were identified. Three of those were accounted for in the syntheses performed by Slavin et al. (2009) and Slavin et al. (2011). The other four were either not included in the syntheses performed by Slavin et al. (2009) and Slavin et al. (2011) or published in 2011 or thereafter. Overviews of each of the seven studies are presented in Table 1.
### Table 1

#### Overview of Studies

<table>
<thead>
<tr>
<th>Study</th>
<th>Design</th>
<th>N</th>
<th>Intervention Grade</th>
<th>Measures</th>
<th>Duration</th>
</tr>
</thead>
<tbody>
<tr>
<td>Macaruso et al. (2006)</td>
<td>Quasi-Experimental</td>
<td>179</td>
<td>Lexia Learning</td>
<td>GMRT</td>
<td>64 sessions of unspecified duration</td>
</tr>
<tr>
<td>McMurray (2013)</td>
<td>Quasi-experimental</td>
<td>106</td>
<td>Lexia Learning</td>
<td>GRT II</td>
<td>Twenty weeks of unspecified duration</td>
</tr>
<tr>
<td>Savage et al. (2009)</td>
<td>Experimental</td>
<td>144</td>
<td>ABRA</td>
<td>GRADE; CTOPP, tests of the WJ III ACH</td>
<td>An average of 13 hours over 12 weeks</td>
</tr>
<tr>
<td>Savage et al. (2010)</td>
<td>Quasi-experimental</td>
<td>60</td>
<td>ABRA</td>
<td>GRADE; CTOPP, tests of the WJ III ACH</td>
<td>An average of 16 hours over 13 weeks</td>
</tr>
<tr>
<td>Savage et al. (2013)</td>
<td>Experimental</td>
<td>1,067</td>
<td>ABRA</td>
<td>GRADE; CTOPP, DIBELS</td>
<td>Approximately 20 hours</td>
</tr>
<tr>
<td>Tracey &amp; Young (2007)</td>
<td>Quasi-experimental</td>
<td>265</td>
<td>Waterford Early Learning Program</td>
<td>TERA-2 (half of participants)</td>
<td>Two semesters of sessions of unspecified duration</td>
</tr>
</tbody>
</table>

*Note.* Abbreviations, in order of appearance, are as follows: GMRT to the Gates-MacGinitie Reading Tests (MacGinitie, Kamons, & Kowalski, 2000); GRMT to the Group Reading Test, Second Edition, (Nelson, 1998); GRADE to the Group Reading Assessment and Diagnostic Evaluation (Williams, 2001); CTOPP to the Comprehensive Test of Phonological Processing (Wagner, Torgesen, & Rashotte, 1999), WJ III ACH to the Woodcock–Johnson III: Tests of Achievement (Woodcock, McGrew, & Mather, 2001); DIBELS to the Dynamic Indicators of Basic Early Literacy Skills (University of Oregon Center on Teaching and Learning, 2009); and TERA-2 to the Test of Early Reading Ability, Second Edition (Reid, Hresko, Hammil, 1989).
Evaluation for Essential Quality Indicators

In the following section, a brief overview of the essential quality indicators identified by Gersten et al. (2005) is provided, along with a discussion of the degree to which each of the articles satisfy those criteria. Gersten et al. (2005) organized their essential quality indicators into four categories: description of participants, implementation of the intervention and description of the comparison condition(s), outcome measures, and data analysis. Those same categories are used to organize the following section.

Description of participants. Because student populations are often heterogeneous over a variety of demographic factors, it is essential for researchers to describe the participants in detail sufficient to aware readers of demographic variables that might impact generalization or replication. This is particularly important in the case of research that involves students with learning or functional disabilities, students from minority linguistic backgrounds, and students of minoritized ethnic or racial backgrounds. Also of importance are robust procedures designed to ensure the comparability of participants across conditions, as well as procedures to ensure that intervention providers and examiners are comparable across conditions. Such rigor is essential to minimizing confounds, experimenter effects, and bias.

Description of participants and intervention providers was generally strong across studies. Of the seven articles reviewed for this synthesis, six (Cassady & Smith, 2005; Macaruso et al., 2006; McMurray, 2013; Savage et al., 2013; Savage, Abrami, Hipps, & Deault, 2009; Savage, Erten, Abrami, Hipps, Comaskey, & van Lierop, 2010) met all of the quality indicators for the description of participants with a score of adequate or
ample. In no case did the authors exclusively focus on students with learning or functional disabilities; however, McMurray (2013) selected as participants only students’ whose pretest scores on standardized measures were indicative of poor reading achievement or dyslexia. Similarly, in two of the studies (Cassady et al., 2005; Macaruso et al., 2006), the authors focused at least some of their analyses on students performing at the bottom of the distribution. In the case of Macaruso et al. (2006), the data from students identified at risk based on Title I status were analyzed separately, and Cassady and Smith provided separate analyses by quartile. In each of the cases where the authors examined the data by quartile or risk status (Cassady & Smith, 2005; Macaruso et al., 2006), a rationale was provided and procedures were clearly delineated.

In the research performed by Cassady and Smith (2005), Macaruso et al. (2006), Savage et al. (2009), Savage et al. (2010), and Savage et al. (2013) relevant demographic data were reported, and comparisons were performed to ensure that no significant differences were present among or between groups. McMurray’s design prioritized students with the most significant deficits for assignment to the treatment group; nevertheless, analyses revealed no significant differences between groups in eleven baseline variables. Modest significant differences between groups were detected only in one measure (of two) of a single variable, picture vocabulary. To account for this difference, the author performed regression analyses to determine the proportion of the variance in outcomes for which it was responsible. In the research performed by Tracey and Young (2007), however, significant pretest differences favoring the nonintervention group were found on the Test of Early Reading Ability, Second Edition (TERA-2;

---

15 It should be noted that Macaruso et al. (2006) excluded data derived from students with documented learning disabilities, owing to the students’ uneven distribution across groups.
Wheeler, 1999), but these differences were not addressed or controlled for in subsequent data analyses.

Because each of the studies reviewed for this synthesis included interventions delivered primarily by computer, with little active participation from teachers or other interventionists, the impact of interventionists on participant performance was likely modest. Therefore, the author of the present study chose to eliminate this criterion from her analyses. Nevertheless, the authors of all but one of the studies reviewed for this synthesis (McMurray, 2013) provided information documenting teacher preparation and training to implement their respective interventions, and the authors of all of the studies documented the qualifications of examiners.

**Implementation of the intervention and description of the comparison condition(s).** As with student populations, there is considerable variability among educational interventions and curricula, as well as the conditions and contexts in which they are implemented. It is therefore essential that interventions be described with replicable precision, as are features of the learning environments and contexts in which those interventions take place. Also of importance are measures designed to assess fidelity of implementation and efforts to determine whether the nature of instruction or services provided to control participants might introduce confounds into study data or analyses. If relevant details of the intervention and control conditions are omitted from the research report, the ability of researchers, practitioners, and policymakers to generalize from the research is compromised.

Documentation of the implementation of the intervention and description of the comparison condition was an area of weakness across the majority of studies. Only one
of the studies (Savage et al., 2013) analyzed for this synthesis satisfied all of the quality indicators for the implementation of the intervention and description of the comparison condition(s) identified by Gersten et al. (2005). The authors of all of the studies (Cassady & Smith, 2005; Macaruso et al., 2006; McMurray, 2013; Savage et al., 2009; Savage et al., 2010, Savage et al., 2013, Tracey & Young, 2007) described their respective interventions in sufficient detail; however, the majority of the studies’ authors (Cassady & Smith, 2005; Macaruso et al., 2006; McMurray, 2013; Savage et al., 2009; Savage et al., 2010, Tracey & Young, 2007) failed to adequately document the nature of services provided to participants in the comparison condition. All of the authors described to some degree the nature of typical classroom reading instruction across conditions, but only one (Savage et al., 2013) observed and specifically documented the instruction provided to control group participants.

The authors of all of the studies (Cassady & Smith, 2005; Macaruso et al., 2006; McMurray, 2013; Savage et al., 2009; Savage et al., 2010, Savage et al., 2013, Tracey & Young, 2007) provided at least some documentation of the duration and intensity of their respective interventions. Both Macaruso et al. (2006) and McMurray (2013) implied that the reading software was used unevenly across participants. McMurray (2013) specifically addressed and evaluated the proportion of variance in participant achievement attributable to this factor; however, she did not provide documentation of the mean use of the intervention or the range in intervention use, only stating that it was used for 20 weeks between pre- and post-testing. Macaruso et al. (2006), by contrast, provided detailed usage data, including mean use of the intervention and the range in intervention use. Furthermore, the authors (Macaruso et al., 2006) performed analyses to
determine the relationship between gain scores and number of skill units completed by Title I students in the treatment group.16

**Outcome measures.** Multiple high-quality measures, including both instruments well aligned to the intervention and instruments sufficient to demonstrate that generalizable skills have been taught, are essential to high quality research. Measures with established validity and reliability are preferable to those of uncertain reliability and validity, and multiple measures are necessary to ensure that sufficient data exist to support the researchers’ conclusions concerning intervention effects. Also of importance is information concerning the timing and frequency of the administration of measures. For randomized controlled trials that provide sufficient evidence of comparability of groups, post-test only designs may be adequate. For quasi-experimental designs, at minimum, pre- and post-test measures must be performed within a temporal proximity sufficiently close to the intervention to demonstrate that any significant differences are likely attributable to the intervention.

Studies varied as to the quality of outcome measures. Four of the seven articles (Macaruso et al., 2006; Savage et al., 2009; Savage et al., 2010; Savage et al., 2013) satisfied all the criteria for outcome measures outlined by Gersten et al. (2005). In the case of Savage et al. (2009), Savage et al. (2010), and Savage et al. (2013), the authors used several nationally normed measures of reading achievement, to provide a good balance between broad and narrow measures of reading achievement. While Macaruso et al. (2006) and McMurray (2013) both used only a single instrument to measure reading achievement, each was nationally normed and well aligned to its respective intervention.

---

16 It is worth noting that this analysis might be misleading, as McMurray (2013) reported that the lowest performers in her sample sometimes became stuck within a particular skill unit, despite high levels of program usage.
In the case of Macaruso et al. (2006), individual test data were reported, giving measures of both word decoding and knowledge of phoneme-grapheme correspondences. Cassady and Smith (2005), however, used only state standardized testing data to measure reading achievement; no measures closely aligned to the intervention were performed. The analyses performed by Tracey and Young (2007) were unusual in that only half of participants were given a nationally normed measure of reading achievement. The other half were given a test of auditory processing and a test of reading abilities created by the publisher of the intervention. While the standardized test (TERA-2; Wheeler, 1999) offered to half the participants contained both broad and narrow measures, only composite scores were reported.

All of the studies (Cassady & Smith, 2005; Macaruso et al., 2006; McMurray, 2013; Savage et al., 2009; Savage et al., 2010, Savage et al., 2013, Tracey & Young, 2007) employed pre-post designs; in only one study (Savage et al., 2009), however, did the authors perform follow-up testing to capture any enduring effects of the intervention. While follow-up testing is desirable, it is not essential to high quality research per the quality criteria outlined by Gersten et al. (2005).

**Data analysis.** The quality of data analysis is key to accurately identifying any effects of an intervention, as well as their significance and magnitude. It is therefore essential that analyses be carefully linked to research question(s) and that the unit of analysis is consistent with the unit of group assignment. Research must demonstrate that the assumptions of statistical tests have been met or otherwise addressed. Furthermore, researchers should report not only inferential statistics but also measures of statistical
power and the magnitude of effects. These data are important to determining not only the statistical significance of any treatment effects, but also their practical significance.

The quality of data analysis varied considerably from study to study. Three of the studies (Cassady & Smith, 2005; McMurray, 2013; Tracey & Young, 2007) failed to satisfy at least one of the quality indicators for data analysis outlined by Gersten et al. (2005). The authors of the other four articles (Macaruso et al., 2006; Savage et al., 2009; Savage et al., 2010; Savage et al., 2013) performed sound analyses and provided measures of effect size for all key findings. In the case of Savage et al. (2010) and Macaruso et al. (2006), univariate analyses of covariance were performed on post-test data using pre-test scores as covariates. Macaruso (2006) provided inferential analyses only for composite scores,17 while Savage et al. (2010) evaluated group differences on several tests of reading achievement, applying Bonferroni’s correction for multiple comparisons. Savage et al. (2009) used a similar procedure, following which adjusted scores were used for planned comparisons across the three groups. As for Savage et al. (2013), achievement test data were analyzed using hierarchical linear modeling (HLM) appropriate to the study’s nested design.

While McMurray (2013) employed sound data analysis (univariate analysis of variance on gains for a single measure with post-hoc regression analyses to determine the proportion of variance in reading scores accounted for by various factors), effect sizes were not reported. In the case of Cassady and Smith (2005), effect sizes were provided for key measures; however, the authors failed to account for an important potential confound in their research design. Rather than a traditional randomized controlled trial or matched quasi-experimental design, the authors elected to perform a longitudinal cohort

---

17 Macaruso et al. (2006) also provided descriptive statistics for test performance.
analysis, evaluating the relative achievement of two consecutive cohorts of first-grade students. The first cohort of students did not receive any ICT-based reading intervention, but the second was subjected to an ICT-based reading intervention as a supplement to regular classroom instruction. At the same time, participating teachers were involved in an ongoing professional development program specifically targeted at improving reading instruction. As such, it is impossible to determine with confidence whether or not any benefits accrued to the second cohort were a result of the ICT-based reading intervention, or a function of improved instruction delivered by teachers. As for the research performed by Tracey and Young (2007), significant differences were found between groups on pre-test achievement scores; however, those differences were not controlled for in the univariate analysis of variance on gains, and measures of effect size were not reported. Results of evaluations for the presence of the quality indicators identified by Gersten et al. (2005) are presented in Table 2.
Three studies analyzed for this synthesis (Cassady & Smith, 2005; McMurray, 2013; Tracy & Young, 2007) did not meet the standards for high or acceptable quality research outlined by Gersten et al. (2005). While Cassady and Smith (2005), Tracey and Young (2007), and McMurray (2013) reported statistically significant positive effects of their respective interventions, elements in their research design, analysis, or reporting precluded a designation of acceptable or high quality. Nevertheless, the research performed by Cassady and Smith (2005) is notable in an important regard: Cassady and Smith (2005), like Macaruso et al. (2006), found that the lowest performing participants at pre-test showed the greatest gains at post-tests. In none of the ABRACADABRA intervention studies analyzed for this synthesis (Savage et al., 2009; Savage et al., 2010;
Savage et al., 2013) were student gains evaluated by quartile or disability status; however, such analyses might have yielded valuable findings concerning factors affecting differential response.

The study performed by McMurray (2013) approaches the standards for high quality research outlined by Gersten et al. Were a more detailed follow-up report published by the author (McMurray, 2013), the cumulative weight of the studies performed by McMurray (2013) and Macaruso (2006) might be sufficient to designate Lexia a promising practice per the guidelines established by Gersten et al. (2005).

**Studies that Met Quality Standards**

Four of the seven studies (Macaruso et al., 2006; Savage et al., 2009; Savage et al., 2010; Savage et al., 2013) analyzed for this synthesis met the criteria for high quality research identified by Gersten et al. (2005). Notably, all but one (Macaruso et al., 2006) of the high quality studies were conducted by the same research team and involved the same ICT-based reading software. Though three of the high quality studies (Macaruso et al., 2006; Savage et al., 2009; Savage et al., 2010) failed to satisfy one of the essential criteria, they satisfied a large number of the desirable criteria and thus met the standards for high quality research (Gersten et al., 2005), minor weaknesses notwithstanding. It is worth noting that in the study that satisfied all of the quality criteria (Savage et al., 2013), the authors explicitly stated their intention to perform high quality research and referenced the Consolidated Standards of Reporting Trials (CONSORT) (Shultz, Altman, & Moher, 2010) standards as the framework for their study design, analyses, and reporting.
In one high quality study (Macaruso et al., 2006), instruction using Lexia Phonics Based Reading and Lexia Reading SOS (Strategies for Older Students) software was provided to treatment group participants. Both programs offered intensive, computer adaptive, and sequential instruction in code-based skills. The authors (Macaruso et al., 2006) employed a quasi-experimental design involving 179 first-grade students enrolled in five schools in an urban school district. While the authors failed to find statistically significant differences favoring the treatment group after one school year, they did detect statistically significant differences in both measures of phoneme-grapheme correspondences and measures of word reading favoring students eligible for Title I services. In fact, at post-test, the achievement of Title I students assigned to the treatment group was comparable to that of students not eligible for Title I services.

In three other high quality studies (Savage et al., 2009; Savage et al., 2010; Savage et al., 2013) ABRACADABRA, a free, interactive web-based literacy program distributed by Canada’s Centre for the Study of Learning and Performance (CSLP) (n.d. a) was provided to participants in the treatment condition. While the intervention was the same across three of the high quality studies (Savage et al., 2009; Savage et al., 2010; Savage et al., 2013), the designs employed by the authors varied somewhat from study to study. In the oldest of the studies (Savage et al., 2009), the authors compared student achievement across three different conditions: two treatment conditions involving different aspects of the ABRACADABRA web-based instructional software and one control condition involving teacher-delivered balanced literacy instruction. In both treatment conditions, participating students completed ABRACADABRA activities for word analysis, text comprehension, and fluency; however, the participants in one
treatment condition completed phoneme-based synthetic phonics activities using ABRACADABRA, while participants in the other treatment condition completed rime-based analytic phonics activities using ABRACADABRA (Savage et al., 2009). Students in both treatment conditions experienced significant gains relative to students in the control condition in word blending, elision, word attack, and reading comprehension at the immediate posttest (Savage et al., 2009). At the delayed posttest, students in the analytic phonics treatment condition experienced significant gains relative to the control condition in word blending, elision, rapid object naming, word attack, and reading comprehension, while students in the synthetic phonics treatment condition experienced significant gains relative to the control condition in word blending, elision, and reading fluency (Savage et al., 2009).

In the ABRACADABRA intervention study performed by Savage et al. in 2010, the authors compared gains in student achievement across three ABRACADABRA treatment conditions and one business-as-usual control condition. Treatment conditions were identified based on participating teachers’ self-reported stage of technology use and integration: entry, in which teachers resisted or failed to integrate instructional technology purposefully into classroom instruction; adoption, in which teachers applied instructional technology purposefully and systematically, but did little to integrate that technology into other forms of instruction; and adaptation, in which teachers used instructional technology systemically and purposefully, while adapting and integrating other classroom instruction to support the skills learned through the use of instructional technology (Savage et al., 2010). Though the trial was short in duration and involved only a small number of participants (N = 60), the authors nevertheless found statistically
significant positive effects of the intervention across measures (Savage et al., 2010). Notably, however, these effects were present only for participants in the adoption and adaptation conditions. Participants whose teachers were in the entry condition underperformed those in the control condition, suggesting that teachers’ comfort and facility with instructional technologies may be an important factor mediating student achievement outcomes.

In the most recent study performed by Savage et al. (2013), the authors conducted a carefully designed, large-scale (N = 1,067) randomized controlled trial of the ABRACADABRA intervention in 74 kindergarten and grade one classrooms across eastern Canada. At post-test, participants assigned to the ABRACADABRA intervention treatment condition demonstrated significant gains in phonological blending ability, letter-sound knowledge, and phoneme segmentation fluency relative to controls (Savage et al., 2013). Among participants whose teachers implemented the ABRACADABRA intervention with a high degree of fidelity, significant gains were detected in phonological blending, phoneme segmentation fluency, sight word reading, and letter-sound knowledge (Savage et al., 2013).

Across the interventions performed by the ABRACADABRA research team, the CPSL reports weighted effect sizes of +0.396 for alphabetic knowledge, +0.187 for reading fluency, and +0.340 for reading comprehension (Centre for the Study of Learning and Performance, n.d. b). These are non-trivial effects, suggesting that the ABRACADABRA intervention and similar ICT-based reading interventions have the potential to markedly improve student achievement in reading and related skills. Furthermore, the studies performed by the ABRACADABRA research team (Savage et
al., 2009; Savage et al., 2010; Savage et al., 2013) satisfy the criteria for the identification of a promising educational practice suggested by Gersten et al. (2005). It is therefore concluded that the ABRACADABRA web-based reading intervention is a promising, research-supported practice.

**Discussion**

**Summary of Findings**

The author of this synthesis sought to answer the following research questions: (a) Is there sufficient evidence to designate any ICT-based program of reading instruction as a research-supported practice? and (b) What features, if any, are common to high quality ICT-based programs of reading instructional interventions?

According to Gersten et al. (2005), for a practice to be considered promising, there must be at least four acceptable quality studies, or two high quality studies, that support the practice. The results of the author’s analyses suggest that while there is an overall dearth of high quality research in ICT-based beginning reading instruction, sufficient evidence exists to designate the ABRACADABRA web-based reading intervention a promising research supported practice for the promotion of reading achievement among students in the early primary grades.

As for the features of the ABRACADABRA intervention, according to CSLP (n.d. a), ABRACADABRA is not designed to be used as a stand-alone product, but as a supplement to classroom instruction for approximately 15 minutes each school day. It provides resources in five different areas: instruction, professional development, communication, assessment, and parent tools (Centre for the Study of Learning and Performance, n.d. a). The instruction component offers analytic and synthetic phonics,
word-level practice, letter-sound recognition, word segmenting, modeled reading and expression, sentence writing, and story sequencing, among other activities. Teachers are charged with selecting developmentally appropriate starting points for their students; however, instruction builds sequentially based on student performance (Centre for the Study of Learning and Performance, n.d. a). The professional development component of the ABRACADABRA software includes lesson plans, video clips, and animated demonstrations to assist teachers in understanding and integrating ABRACADABRA into their own classroom instruction (Centre for the Study of Learning and Performance, n.d. a). Teachers can also share their questions, anecdotes, remarks and lesson plans using the communication component of the ABRACADABRA software (Centre for the Study of Learning and Performance, n.d. a). ABRACADABRA’s assessment component provides teachers with activity statistics and error reports to facilitate planning and remediation (Centre for the Study of Learning and Performance, n.d. a). Finally, the parent component of the ABRACADABRA software provides instructional videos, resources for using the ABRACADABRA’s instruction component at home, printable activities, and information to help parents engage their children in reading (Centre for the Study of Learning and Performance, n.d. a).

**Implications for Future Research**

While the results of this synthesis point to a promising practice in the area of ICT-based reading interventions, the base of high quality research in the field nevertheless remains thin. Consistent with previous syntheses of ICT-based reading intervention research, the author found very few studies meeting selection criteria. Clearly, therefore, further research is warranted. High-quality intervention studies, including replication
research, should be performed to evaluate the ICT-based reading interventions most commonly used in American and Commonwealth public schools. This research should be geared not only toward product evaluation, but also to the identification of optimal instructional practices. Particular attention should be paid to the duration and intensity of instruction, level of instructional integration, and level of support provided to teachers and students. Furthermore, researchers should endeavor to identify factors affecting differential levels of responsiveness among students and differential levels of instructional integration and treatment fidelity among teachers. Finally, research should be performed in accordance with an established framework for high quality research and reporting (e.g., the guidelines suggested by the What Works Clearinghouse [What Works Clearinghouse, 2008], CONSORT [Shultz et al., 2010] or the Council for Exceptional Children [Gersten et al., 2005]).

Implications for Practice

The results of this synthesis provide only modest implications for practice. At this time, the ABRACADABRA web-based reading intervention is available free of charge to teachers in Canada (Centre for the Study of Learning and Performance, n.d. a); however, the author of this synthesis was able to find no comparably-researched program of instruction available to teachers elsewhere in the Commonwealth or in the United States. Notable in the design of the ABRACADABRA web-based reading intervention is the degree of coordination between software-based instructional activities, classroom instruction, and at-home learning and reinforcement. Also notable is the fact that the ABRACADABRA web-based reading intervention contains systematic, developmentally
appropriate instruction in phonics and other code-based skills, for which there is strong
evidentiary support in other research contexts.

**Limitations**

This synthesis has several limitations. The search procedures employed by the author were by no means exhaustive, and articles meeting selection criteria may have been overlooked. Furthermore, the selection criteria used by the author were narrow in scope, and only articles published in refereed journals were considered for inclusion. Therefore, unpublished works, such as dissertations and articles in submission, were excluded from analysis. Moreover, the research evaluated for this synthesis was rated by only the author, so inter-rater reliability data are therefore unavailable.

It should also be noted that for the purpose of this synthesis, qualitative designs and mixed-methods or quasi-experimental designs that did not include both pre- and post-test measures of decoding, as well as controls for differences between, groups were excluded. Nevertheless, many of these studies offer important insights about not only the efficacy of interventions, but also their usability, acceptability, and social validity. Furthermore, they often serve to elucidate the conditions and contingencies most likely to promote a program’s success.
CHAPTER THREE: METHOD

In this chapter, the method by which the present study was conducted is discussed. The composition of the sample and methods for controlling for differences between groups are outlined and explained. The content and delivery of the intervention are described in detail, and representative screen shots are provided. The conduct of study measures and the qualification of examiners are described. Instruments are discussed, including the validity, reliability, and suitability to the present study. Finally, the research design and approach to data analysis are detailed and defended.

The purpose of the present study was to determine whether there were significant mean differences in (a) non-word reading, (b) real word reading, (c) non-word spelling, (d) real word spelling, and/or (e) reading fluency post-test achievement scores of students assigned to use the MVRC online reading intervention in addition to regular classroom reading instruction and those of students from a business-as-usual comparison condition, once the effects of differences in pre-test achievement scores and relevant demographic variables had been accounted for.

The present research was conducted using a quantitative, quasi-experimental experimental design with multivariate analysis of covariance (MANCOVA) as the primary means of data analysis. All data collection activities were supervised and coordinated by a third-year doctoral student in special education. The resulting data set was analyzed after having undergone de-identification, including the removal of all individually identifying data from case records.
Participants and Settings

Participants. Participants included 209 students enrolled in eight, second grade classrooms in two public elementary schools in the southwestern United States. Of those, 107 were assigned to the treatment condition, and 102 were assigned to a business-as-usual comparison condition. Owing to attrition and illness, 39 participants had incomplete data sets following post-testing. Ultimately, the treatment condition comprised 89 complete cases, and the comparison condition comprised 81, representing 170 complete cases, or 81.34% of original cases. While overall data loss was below 20%, data loss did not impact each condition equally, as the treatment group retained 83.12% of original cases, while the comparison group retained 79.41% of original cases, producing a differential attrition rate of 3.71%.

Participant demographics. All participants attended schools in the same district and within the same postal code. At the time of testing, 81.6% of students in the district were identified as Hispanic, 12.1% as white, 4.9% as Native America, 2.7% as African American, and 0.6% as Asian or Pacific Islander. Over nine-in-ten (93.2%) students were eligible to receive free or reduced price school meals, and 6.2% were homeless. District-wide, nine percent of students were classified as English language learners (ELL), and an additional 5.3% had been reclassified as fluent English speakers.\textsuperscript{18} Additionally, 11.8% of students within the district had an individualized education program (IEP), indicating a documented disability.\textsuperscript{19}

The percentage of students in the sample who had a documented disability tracked

\textsuperscript{18} This reclassification process occurred when students achieved a satisfactory level of English-language proficiency per standardized testing.

\textsuperscript{19} Data were derived from district publications. A reference to the source of this data has not been included, so as to protect the identity of participating schools.
very closely with district demographics, at 12.3%. Sample demographics, however, varied markedly from district demographics in ELL status, as 37.4% of sample group members were classified as ELLs, and an additional 8.8% were reclassified as fluent English speakers. Data concerning ethnicity and markers of socioeconomic status were not made available at the individual case level in order to protect the privacy of participants.

**Assignment to groups.** Because of the potential for disruption to classroom routines, random assignment at the student level was impossible. Assignment to groups was therefore performed at the school level. One school, including each of its two participating classrooms, was assigned to the treatment condition, and the other school, including each of its two participating classrooms, was assigned to the comparison condition. The school whose third grade students (the youngest grade tested) performed more poorly (39% pass rate in reading) relative to the other (53% pass rate in reading) on the state standardized test of in the 2012-2013 school year was assigned to the treatment condition, and the school that performed better (53% pass rate in reading) relative to the other (39% pass rate in reading) on the state standardized test of reading in the 2012-2013 school year was assigned to the comparison condition. Group assignment was performed in this manner in order to ensure that any demographic advantages would favor the comparison group.

**Intervention settings.** During treatment, treatment group students received MVRC online reading instruction in their schools’ computer labs during a fixed period each day, Monday through Thursday. No other students were permitted in the lab during treatment,

---

20 Data were derived from district publications. A reference to the source of this data has not been included, so as to protect the identity of participating schools.
and the environment was made as reasonably free of distractions as possible.
Participating students were provided with individual computers, monitors, and
headphones. Prior to the beginning of treatment, MindPlay Educational Software for
Reading personnel visited the school sites and confirmed the adequacy of the schools’
computer hardware and Internet connectivity.

Materials

**Mindplay Virtual Reading Coach (MVRC).** The MVRC (MindPlay Educational
Software for Reading, 2015) intervention is a multi-component reading and language arts
curriculum delivered via the Internet. It contains explicit instruction in (1) phonemic
awareness, (2) phonics, (3) fluency, (4) vocabulary, and (5) comprehension in alignment
with the recommendations of the National Reading Panel (NRP) (2000), as well as
instruction in grammar and concepts about print. It provides sequenced instruction
consistent with the models of alphabetic and phonological development proposed by Ehri
(2005) and Pufpaff (2009) and satisfies the large majority of the content criteria outlined
by Grant et al. (2012), as well as all of the instructional and interface design criteria
identified by Bishop and Edwards Santoro (2006).

**MVRC overview.** To access the MVRC program, participants log into an online
account, in which their interactions with the software and progress toward reaching
achievement targets are logged and retained (MindPlay Educational Software for
Reading, 2012). When students first access the program, they are immediately directed
to a diagnostic assessment, which is used for tracking and placement. The full diagnostic
assessment is repeated three times each academic year, and comprehensive progress
monitoring assessments are delivered every 14 days. All assessments are computer
adaptive: items delivered to a student increase or decrease in difficulty in response to that student’s performance.

The first level of instruction to which students are directed is characterized as mastery. Mastery level activities include phonemic awareness, phonics, and grammar and meaning. Placement within mastery activities is dependent upon diagnostic assessment scores, and content within each lesson is adaptive in response to student performance. Instruction and remediation within the activities are delivered via video clips featuring the student’s assigned reading coach. Reading coaches include speech and language pathologists (SLPs) and other specially trained individuals of diverse age and ethnicity, and scripts and other instructional content were developed with the assistances of certified SLPs. The presentation of instructional video clips varies in response to student performance. As an example, an advanced student who submits the response “fone” when asked by the reading coach to type the word “phone” will see a video clip in which the reading coach advises the student that the /f/ sound can have different spellings. If the student again fails to spell the word correctly, a video clip in which the reading coach reminds the student that in words of Greek origin the /f/ sound is spelled using the digraph ph will appear.

The phonemic awareness component of MVRC contains three lesson sets of increasing difficulty, while the phonics and grammar and meaning components contain five lesson sets of increasing difficulty. Each lesson set contains up to 20 different presentations of the lesson content. If, based on the results of internal assessments, a student fails to master at least 90% of the content assigned for a lesson, the lesson will be represented in a different format until all of the lesson’s alternative presentations are
exhausted. The MVRC intervention allocates instructional time in relation to student performance such that skills and activities in which a student demonstrates the greatest proficiency receive the least instructional time, while skills and activities in which a student demonstrates the least proficiency receive the greatest instructional time.

Once a student has demonstrated 90% proficiency or better on required mastery level activities, (s)he begins proficiency level activities, including fluency and vocabulary lessons. Embedded within these lessons are structured spelling and comprehension activities, in addition to lessons in the core skills associated with each lesson. The time devoted to proficiency level activities varies in relation to the student’s performance on mastery level activities. If a student demonstrates a high degree of proficiency in mastery level activities, a greater proportion of instructional time is devoted to proficiency level activities. Conversely, if a student struggles to achieve mastery level performance objectives, less time is devoted to proficiency level activities. Figure 1, used with permission, outlines MVRC’s lesson sequence and assessment checkpoints. Note that all phonemic awareness activities must be successfully completed, and the student must achieve mastery in required phonics activities, before the introduction of grammar and meaning, fluency, and vocabulary activities. For a comprehensive description of each of MVRC’s components, refer to Appendix B.
Dependent variable measures. Dependent variable measures included the Test of Silent Word Reading Fluency, Second Edition (TOSWRF-2; Mather, Hammill, Allen & Roberts, 2014) and tests of the Woodcock Johnson Tests of Achievement, Fourth Edition (WJ IVACH; Schrank, Mather, & McGrew, 2014). These measures were chosen because they are considered highly valid measures of their respective constructs, and they have high levels of test-retest reliability, reducing the likelihood of regression to the mean. An overview of each dependent variable measure, including information
concerning the reliability and validity of instruments, is provided in the subsequent sections.

**WJ IV ACH.** The WJ IV ACH (Schrank et al., 2014) is a standardized achievement instrument that is administered in accordance with standardized administration protocols. The WJ IV ACH (Schrank et al., 2014) contains tests designed to measure major areas of academic achievement, including reading, writing, and mathematics. The Woodcock-Johnson IV (WJ IV) (Schrank et al., 2014) comprises the Tests of Cognitive Abilities (COG) and Tests of Achievement (ACH) batteries, as well as the Tests of Oral Language. All three batteries were co-normed on a stratified random sample of 7,416 participants ranging in age from 2 to >90 years, over 3,891 of whom were grade school students at the time of administration (Schrank et al., 2014). Norming and technical analyses were performed in accordance with the *Standards for Educational and Psychological Testing* (American Educational Research Association, American Psychological Association, and the National Council on Measurement in Education, 2014). Mean alternate form and test-retest reliabilities across all age groups for Basic Reading Skills, Basic Writing Skills, and Phoneme-Grapheme Knowledge ranged from 0.94 to 0.95.

**TOSWRF-2.** The TOSWRF-2 is a timed standardized test in which examinees are required to place a line or slash between words that have no spaces between them. It is a pencil-and-paper test that can be administered individually or to groups, the object of which is to draw lines between as many word boundaries as possible within the allotted time. The test employs a graded word list, the vocabulary and complexity of which
progressively increase in difficulty (Mather et al., 2014). The test, therefore, measures not only sight word recognition and reading rate but also vocabulary knowledge.

The TOSWRF-2 (Mather et al., 2014) was normed on a nation-wide sample of over 2,429 participants, ranging in age from six years and three months to 24 years and 11 months. These participants were derived from 35 states, and participant demographics closely resembled national demographics reported in the 2011 census (Mather et al., 2014). Test re-test and alternate form reliabilities ranged from 0.84 to 0.91 (Allen, Morey, & Hammill, 2013). Table 3 provides an overview of the description of each dependent variable measure performed.

Table 3

<table>
<thead>
<tr>
<th>Tests of Reading Achievement</th>
<th>DV</th>
<th>Description of Measure</th>
<th>Administration</th>
</tr>
</thead>
<tbody>
<tr>
<td>Letter-Word Identification</td>
<td>Real Word Reading</td>
<td>Reading aloud letters and real words from a list</td>
<td>Individual</td>
</tr>
<tr>
<td>Word Attack</td>
<td>Non-Word Reading</td>
<td>Reading aloud nonsense words from a list</td>
<td>Individual</td>
</tr>
<tr>
<td>Spelling</td>
<td>Real Word Spelling</td>
<td>Spelling of orally administered real words</td>
<td>Group</td>
</tr>
<tr>
<td>Spelling of Sounds</td>
<td>Non-Word Spelling</td>
<td>Spelling of orally administered nonsense words</td>
<td>Group</td>
</tr>
<tr>
<td>Test of Silent Word Reading Fluency, Second Edition (TOSWRF-2; Mather et al., 2014)</td>
<td>Reading Fluency</td>
<td>Rapidly drawing lines to indicate the boundaries between connected real words</td>
<td>Group</td>
</tr>
</tbody>
</table>

Note. DV = dependent variable. The Letter-Word Identification, Word Attack, Spelling, and Spelling of Sounds instruments are tests of the fourth edition of the Woodcock-Johnson Tests of Achievement (WJ IV ACH; Schrank et al., 2014).
**Rationale for measure selection.** The dependent variable measures selected for this intervention were chosen both because of their high reliabilities and because they were designed to measure code-based skills foundational to proficient reading. Tests of real word and non-word reading and spelling were selected to ensure that the study included measures of the identification and production of both phonetically regular and phonetically irregular words. Identification and production of the former are indicative of graphophonemic knowledge, while identification and production of the latter are indicative of sight word knowledge and familiarity with the word patterns of English. A test of reading fluency was selected both as a broad measure of decoding ability and because reading rate and accuracy are essential to reading comprehension, and there is a strong correlation between reading fluency and comprehension (Pikulski & Chard, 2005), which is the ultimate goal of reading.\(^{21}\)

**Fidelity measures.** To measure fidelity of implementation, both product usage logs and behavioral observation measures were used. Two measures of fidelity of implementation were selected to evaluate not only whether the participants were using the intervention for the duration specified, but also whether they were actively engaged while using the intervention. The active engagement of participants, it was reasoned, would be a more nuanced measure of the acceptability of the intervention among participants than would simple duration of use, thus serving as a metric for an important facet of social validity.

**Product usage logs.** The frequency and duration of individual student's use of the MVRC online reading intervention were measured using software usage logs generated by the MVRC product. These logs included data concerning the duration of use of

\(^{21}\) Because of time constraints, comprehension measures were not conducted.
product components in minutes, the type and number of activities successfully completed, and individual students’ progress toward achievement targets. Only active use of the MVRC product was reported in the software usage logs, as the MVRC online intervention automatically logs users out of the program after three minutes of inactivity.

**Behavioral observations.** A 10-data point planned activity check, using partial-interval recording method, was employed to measure student engagement behavior biweekly throughout the study. This instrument was developed by a doctoral candidate in Special Education at the University of Arizona. Student engagement behavior was operationalized as physical position (head up, eyes facing screen), headphone use, keyboard use, and screen activity. To administer this measure, two observers simultaneously scanned each classroom at two-minute intervals for 20 minutes and documented the number of participants who appeared to be engaged, as well as the number of participants who did not appear to be engaged, per the criteria previously outlined. Data collected with this instrument were used to document the percentage of treatment group participants who displayed engagement behavior at each observed interval. Observers varied their observation schedules each week to reduce the likelihood of participants' modifying their behavior in anticipation of observation. A copy of this instrument is provided in Appendix C.

**Intervention Procedure**

**Routine classroom instruction.** All participants across both groups received Success for All language arts instruction in the regular classroom setting. Over 75% of teachers in the participating schools voted to implement the Success for All program in their schools, and school administrators expressed a high degree of support for the
program to examiners who visited their schools.

Success for All is a comprehensive, code-based, standards-aligned reading curriculum that includes structured lessons, cooperative group activities, and regular assessments (Slavin & Madden, 2012). Success for All classroom activities included intensive daily reading blocks of 90 minutes duration, as well as systematic phonics instruction, guided practice, and ongoing formative evaluation. Success for All activities are scripted and highly structured, so there is little variability in instruction among classrooms implementing the program with a high degree of fidelity. A typical Success for All lesson plan for a single day’s instructional activities exceeds ten pages and includes detailed scripts for every teacher-student interaction. An example of a typical phonics activity embedded in a Success for All daily lesson plan follows:

**Sound words**—Say each of the words below, and have the students repeat them.

turn burn curl

• Ask: **What sound can you hear in those words?** [/ur/.] Say each word again, and have the students repeat each one. Stretch the /ur/ sound in each word.

**Key picture**—Show the key card for “ur.” **Let’s look at this picture card to learn more about our sound for the day.** **This is a picture of a nurse with a purse.** **Nurse with a purse.** Let’s say that phrase together. [Nurse with a purse.] Say each word in the phrase, and ask the students to repeat it. Stretch the /ur/ sound in each word. (Success for All Foundation, 2010)

Success for All early reading instruction is conducted in small groups of children with comparable levels of reading achievement. These groups can include children in
grades one through three. Texts used in the early grades are decodable, or phonetically regular, and aligned to the student’s actual level of achievement. Early primary students also read shared stories, in which a portion of the text is decodable and intended to be read by the students and another portion is more complex and intended to be read by the teacher. As the student’s achievement improves, the portion of the text read by the student increases, while the portion of the text read by the teacher decreases. Other reading activities include partner reading, in which a student and a partner read alternate pages of a text and then discuss what they have read.

Success for All phonics activities begin with oral language and then move into sound-symbol correspondences. As a student's achievement increases, individual letter sounds are contextualized in words, words in sentences, sentences in passages, and, finally, passages in stories. A variety of highly structured metacognitive activities are used to increase students’ retention and integration of the concepts learned, and interactive activities are conducted to promote and reinforce the application of skills. These activities include skits, participatory videos, and puppet shows.

Specific components of the Success for All curriculum comprise (a) FastTrack Phonics, in which activities include guided instruction, chants, games, and puppet shows designed to promote and reinforce knowledge of sound-symbol correspondences; (b) Shared Stories, in which reading, vocabulary, fluency, and comprehension are promoted through partner reading and teacher-student reading; (c) Story Telling and Retelling, in which students are taught to clarify, question, summarize, and visualize texts through story preview, reading, review, retell, and critique activities; (d) Language Links, in which a teacher models and guides higher order discussions involving specific reading-
related skills, such as the ability to identify story elements; and (e) Adventures in Writing, in which students perform process-based writing, following a cycle of prewriting, planning, drafting, checking, and polishing.

**Intervention instruction.** Participating students assigned to the treatment group received MVRC online reading instruction in addition to Success for All reading and language arts instruction. Participants assigned to the treatment group used the software for 30 minutes each day, Monday through Thursday, for a total of two hours per week throughout the regular school year (mid-September through mid-April), with the exception of holidays, school functions, and mandatory state testing days. Classroom teachers brought students to the computer lab at the assigned time and assisted with student log in when required; however, they did not assist students in any other way.

**Comparison instruction.** Participating students assigned to the business-as-usual comparison condition received Success for All reading and language arts instruction but did not receive the MVRC intervention. In addition to regular language arts instruction, participants in the comparison group also received two hours of supplementary reading instruction weekly from their classroom teachers. This instruction employed materials and instructional techniques consistent with the Success for All curriculum and routine classroom practice.

**Teacher training.** Teachers of treatment group classrooms underwent training prior to the implementation of the intervention. Training included a one-hour webinar and a four-hour in-person seminar delivered by MindPlay personnel. The MVRC online reading intervention requires very little participation on the part of classroom teachers, as the content and difficulty of the intervention are adaptive in response to individual
student performance. Ultimate achievement targets are aligned with Common Core standards; however, the content and instruction delivered to individual students are determined solely by embedded diagnostic assessments (i.e., they cannot be modified by teachers). Therefore, teacher training focused on the underlying pedagogical and theoretical structure of the program, the structure and content of the embedded lessons and activities, facilitating and supporting student access and use of the intervention, and accessing and understanding the data contained in student and classroom reports generated by the software. Teachers were also provided with a teacher’s guide to accompany the program.

Assessment Procedure

Procedures to ensure accuracy of administration. Examiners included a graduate student in special education, who had over five years of experience working with children in schools and two years of experience administering standardized tests of achievement and cognitive abilities, as well as several retired special educators, who had extensive experience working with children in educational settings and administering standardized tests of achievement and cognitive abilities. Prior to the administration of reading achievement dependent variable measures, examiners met to review administration. In addition, all examiners received supplemental training in the administration of the WJ IV ACH from one of the tests’ co-authors.

Procedures to ensure accuracy of scoring. Following data collection, test protocols were re-scored by a second examiner to ensure accuracy of scoring. While formal data concerning scoring disagreements and agreements were not collected or retained, examiners reported very few disagreements when rescoring protocols.
**Administration of dependent variable measures.** Pre-test dependent variable data collection was performed over a one-week period during the first 30 days of the academic year, and post-test dependent variable data collection was performed over a one-week period during the last 30 days of the academic year. Prior to dependent variable data collection, examiners endeavored to establish rapport with participating students. Individual data collection was performed in quiet, low-distraction rooms in the participants’ school; group data collection was performed in the participants’ classrooms. Data collection was staggered over several days to avoid fatigue-related effects, and measures were counter-balanced to avoid order effects.

The *Spelling and Spelling of Sounds* tests of the WJ IV ACH (Schrank et al., 2014), as well as the TOSWRF-2 (Mather et al., 2014), were administered to participating students in whole-class groups, whereas the *Word Attack* and *Letter-Word Identification* tests of the WJ IV ACH were administered individually and in a manner consistent with standardized administration protocols. Group-administered tests were administered beginning with the first item in each assessment and ending with items likely to be well beyond the performance of any participant. Consequently, basal (start point) requirements were satisfied for all participants tested, and no participant failed to reach a ceiling (end point) in their performance on any assessment.

**Administration of measures of fidelity of implementation.** The frequency and duration of students’ use of the MVRC online reading intervention were measured using software usage logs generated by the MVRC product. Only active use of the MVRC product was reported in the software usage logs, as the MVRC intervention automatically logged users out of the program after three minutes of inactivity.
A novel planned activity check was employed to measure student engagement behavior, including physical position (i.e., head up, eyes facing screen), headphone use, keyboard use, and screen activity, during treatment. Participating students assigned to the treatment group were observed twice each month for 20 minutes throughout the treatment period, and their engagement behaviors were recorded using the planned activity check instrument. All behavioral observations were performed simultaneously by two qualified examiners.

**Operational Definitions of Variables**

**Independent variable.** The study’s sole independent variable was treatment with MVRC online reading instruction in accordance with intervention protocols. This variable was measured at the nominal level with values of 0 (not subjected to treatment) and 1 (subjected to treatment).

**Dependent variables.** To operationalize non-word reading, scores from the Word Attack test of the WJ IV ACH (Schrank et al., 2014) were used. To operationalize real word reading, scores from the *Letter-Word Identification* test of the WJ IV ACH (Schrank et al., 2014) were used. To operationalize non-word spelling, scores from the *Spelling of Sounds* test of the WJ IV ACH (Schrank et al., 2014) were used. To operationalize real word spelling, scores from the *Spelling* test of the WJ IV ACH (Schrank et al., 2014) were used. To operationalize reading fluency, scores from the TOSWRF-2 (Mather et al., 2014) were used. All dependent variable data were reported at the interval level.

**Demographic variables.** Demographic variable data, including English-language learner (ELL) status and special education status per an individualized educational
program (IEP) were recorded and reported as nominal data.

**Data Analysis**

**Rationale.** In selecting a method of data analysis, the author considered several strategies. The first of these was univariate analysis of variance, or ANOVA, using participant gain scores (calculated by subtracting pre-test scores from post-test scores), which is among the most commonly used techniques for comparing pre-test and post-test data (Dimitrov & Rumrill, 2003). Univariate analyses of variance compare group means for a single dependent variable (Tabatchnick & Fidell, 2013), and the use of gain scores, as opposed to post-test scores alone, for example, serves as an informal means by which to control for group differences (Dimitrov & Rumrill, 2003). The author rejected this strategy for three reasons. First, there is evidence that gain scores can be an unreliable means by which to compare group differences (Cronbach & Furby, 1970). Second, preliminary univariate tests revealed significant differences between groups on pre-test scores for several dependent variable measures, indicating that participant potential for gains might differ between groups. Univariate ANOVA, however, does not control for such differences. Finally, the author wished to avoid the inflation in error created by the conduct of multiple univariate tests (Tabatchnick & Fidell, 2013). Owing to all these factors, therefore, the author determined that ANOVA was inadequate to her purpose.

The second data analytic strategy considered by the author was repeated measures ANOVA using pre- and post-test scores, which would account for the pre-test differences between groups (Dimitrov & Rumrill, 2003). While repeated measures ANOVA is a

---

22 While more recent research indicates that the degree to which gain scores are unreliable varies in relation to the strength of correlation between pre- and post-test scores (Dimitrov & Rumrill, 2003; Zimmerman & Williams, 1996), the author nevertheless desired to minimize the degree of unreliability of her analyses.

23 Described in greater depth in subsequent sections.
popular means by which to handle data in a pre-test post-test design, research has
demonstrated that the results of such analyses are often misleadingly conservative
(Dimitrov & Rumrill, 2003). Furthermore, the conduct of multiple repeated measures
ANOVAs would create the same inflation in error as the conduct of multiple ANOVAs
on gain scores. Moreover, data screening performed by the author indicated that the data
set would violate the assumption of sphericity required to perform the test.

The third data analytic strategy considered by the author was univariate analysis of
covariance, or ANCOVA. Univariate analyses of covariance compare group means
while controlling for the effects of one or more covariates (Tabachnick & Fidell, 2013).
This statistical control increases the power and precision of analyses by reducing error
variance accounted for by the covariate(s) (Dimitrov & Rumrill, 2003). The choice of
ANCOVA, notably, would preclude the analysis of gain scores, as an important
covariate, pre-test achievement scores, would overlap with the dependent variable, gain
scores. Therefore, the use of ANCOVA would require the analysis of post-test scores
with pre-test scores as covariates. While this strategy would account for the differences
between groups, it would nevertheless leave the analyses subject to the inflation in error
created by the conduct of multiple univariate tests (Tabachnick & Fidell, 2013).

The fourth method of data analysis considered by the author was repeated measures
MANCOVA on pre- and post-test scores with relevant demographic covariates. While
this multivariate method of data analysis would control for multiple comparisons,
preliminary analyses performed by the author indicated that the data would violate the
assumption of sphericity required to perform the test. Therefore, this method was also
rejected.

---

24 Gain scores are calculated by subtracting pre-test scores from post-test scores.
The fifth and final data analytic strategy considered by the author was multivariate analysis of covariance, or MANCOVA, on post-test scores with pre-test scores and relevant demographic variables as covariates. Multivariate analysis of covariance can be used to identify statistically significant differences between groups on multiple dependent variables after the effects of relevant covariates have been removed from the analysis (Tabachnick & Fidell, 2013). There are several important advantages to the use of a MANCOVA design over the use of multiple separate analyses of variance (ANOVA), or even analyses of covariance (ANCOVA). First, as with ANCOVA, MANCOVA can be used to control for pre-test differences between groups. This is particularly important in quasi-experimental designs, as pre-test differences between groups can introduce systematic bias, potentially compromising the results (Dimitrov & Rumrill, 2003). Second, MANCOVA accounts for the intercorrelations among the dependent variables in the analysis, leading to a more precise overall picture of change (Tabachnick & Fidell, 2013). Finally, the use of MANCOVA allows the researcher to more accurately identify variables that have changed in relation to treatment, substantially reducing the likelihood of a Type I error (Tabachnick & Fidell, 2013).

**Test procedure.** Owing to these factors, the author elected to perform a multivariate analysis of covariance (MANCOVA) to detect mean differences in (a) non-word reading, (b) real word reading, (c) non-word spelling, (d) real word spelling, and/or (e) reading fluency post-test achievement scores of students assigned to use the MVRC online reading intervention and those of students assigned to a business-as-usual comparison condition (regular classroom reading instruction alone), once the effect of differences in pre-test achievement scores (averaged across tests) and relevant
demographic variables had been accounted for.

**Confirmatory measures.** As a confirmatory measure, the author chose to perform univariate analyses of covariance (ANCOVA) post-hoc to the full MANCOVA to investigate group differences by dependent variable while controlling for pre-test achievement scores (averaged across tests) and relevant demographic variables. The author also elected to perform a second set of post-hoc ANCOVAs to isolate the relationship between each dependent variable, its corresponding pre-test covariate, and the grouping variable (condition). By controlling for pre-test achievement individually for each dependent variable, rather than using the mean pre-test score covariate employed in the full MANCOVA, the author sought to lend more precision to the analyses and ensure that the specific effects of pre-test achievement on each dependent variable were adequately accounted for.
CHAPTER FOUR: RESULTS

In this chapter, the author relates the results of data analyses, including data screening and cleaning procedures, omnibus testing, univariate testing, and post-hoc testing for all dependent variable measures. Measures of fidelity implementation are also reported, as are results of behavioral observations.

Data Screening and Cleaning

Measurement of data. All dependent variable and pretest covariate data collected for this study were measured as obtained, raw scores using discrete, quantitative scales. Obtained, raw scores for each post-test achievement measure comprised the study’s dependent variables. A pre-test covariate was created by taking the mean of sample-referenced z-scores across pre-tests. The conversion of scores was undertaken prior to averaging to ensure that all data reflected equivalent scales of measurement. Such a procedure was necessary because the number of items varied across tests, as did item difficulty, underlying constructs, and scoring procedures. As an example, raw scores on the TOSWRF-2 (Mather et al., 2014) were calculated by subtracting items incorrect from items correct, while raw scores for all other measures were calculated by adding only correct items.

Missing data. Owing to attrition and illness, 39 participants had incomplete data sets following post-testing. Ultimately, the treatment condition comprised 89 complete cases, and the comparison condition comprised 81, representing 170 complete cases, or 81.34% of original cases. Because more than 10% of the original cases were missing

---

25 The author chose to create a single pre-test achievement covariate, rather than add the pre-test achievement scores for each dependent variable a separate covariate to the MANCOVA model. This decision was taken to reduce the number of variables added to the MANCOVA model, thereby increasing statistical power.
data, the authors chose to exclude cases listwise from analyses. Imputation of missing data was considered and rejected consistent with the recommendations of Tabachnick and Fidell (2013), who suggested imputation only when fewer than 10% of cases were missing data and data were missing at random.

**Data screening.** Prior to analysis, dependent variable data were evaluated for data entry errors using Microsoft Excel and for normality, linearity, skewness, and kurtosis using IBM SPSS EXPLORE. Univariate outliers with respect to the dependent variables (defined as values more than three standard deviations above the mean for the relevant variable) were detected using IBM SPSS FREQUENCIES, and multivariate outliers were detected using IBM SPSS REGRESSION to calculate Mahalanobis' distances.

Careful evaluation of the data revealed no values suggestive of data entry errors; however, a total of three cases with univariate outliers were detected. Each case had only a single outlying variable, and all three involved values between three and four standard deviations above the mean. In order to preserve the maximum number of cases possible, outlying values within a case were treated as missing and replaced using the highest or lowest non-outlying value possible (i.e., a value exactly three standard deviations above or below the mean), consistent with the recommendations of Tabatchnick and Fidell (2013). In addition to the univariate outliers, a single multivariate outlier was detected. The value was not extreme, however, so the author elected to retain the case, but include and review the Box's $M$ statistic in the subsequent MANCOVA.

Following the transformation of outlying data, the author split the data by condition and generated scatterplot matrices to evaluate the linearity of data. Upon visual inspection, the author determined that the data were sufficiently linear to proceed with
further analyses. The author then generated histograms, stem-and leaf plots, and Q-Q plots for each of the study’s dependent variables. A visual inspection of the plots and histograms for each of the dependent variables revealed varying degrees of (most often negative) skewness, as well as modest amount of excess kurtosis, with the exception of a single variable highly leptokurtic distribution. This finding was reinforced by significant values at the $p < .05$ level for non-word spelling for the treatment group, real-word spelling for the comparison group, and non-word reading for the treatment group on the Shapiro-Wilk test of normality. Table 4 provides descriptive statistics, including measures of skewness and kurtosis for each of the study's dependent variables by condition, and Table 5 provides the results of the Shapiro-Wilk test of normality.
### Table 4

**Descriptive Statistics by Condition for Each Dependent Variable and Interval-Level Covariate**

<table>
<thead>
<tr>
<th>DV</th>
<th>Condition</th>
<th>Measure</th>
<th>Statistic</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Real Word Reading</td>
<td>Comparison</td>
<td>$M$</td>
<td>42.901</td>
<td>1.2004</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$SD$</td>
<td>10.8035</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Skewness</td>
<td>-0.255</td>
<td>0.267</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Kurtosis</td>
<td>-0.606</td>
<td>0.529</td>
</tr>
<tr>
<td></td>
<td>Treatment</td>
<td>$M$</td>
<td>45.944</td>
<td>0.8706</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$SD$</td>
<td>8.2136</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Skewness</td>
<td>-0.466</td>
<td>0.255</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Kurtosis</td>
<td>0.562</td>
<td>0.506</td>
</tr>
<tr>
<td>Real Word Spelling</td>
<td>Comparison</td>
<td>$M$</td>
<td>20.383</td>
<td>0.6654</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$SD$</td>
<td>5.9887</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Skewness</td>
<td>0.572</td>
<td>0.267</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Kurtosis</td>
<td>0.162</td>
<td>0.529</td>
</tr>
<tr>
<td></td>
<td>Treatment</td>
<td>$M$</td>
<td>23.326</td>
<td>0.5661</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$SD$</td>
<td>5.3402</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Skewness</td>
<td>-0.136</td>
<td>0.255</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Kurtosis</td>
<td>0.419</td>
<td>0.506</td>
</tr>
<tr>
<td>Non-Word Reading</td>
<td>Comparison</td>
<td>$M$</td>
<td>18.136</td>
<td>0.5749</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$SD$</td>
<td>5.1739</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Skewness</td>
<td>-0.15</td>
<td>0.267</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Kurtosis</td>
<td>-0.512</td>
<td>0.529</td>
</tr>
<tr>
<td></td>
<td>Treatment</td>
<td>$M$</td>
<td>19.629</td>
<td>0.4322</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$SD$</td>
<td>4.077</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Skewness</td>
<td>-0.339</td>
<td>0.255</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Kurtosis</td>
<td>-0.279</td>
<td>0.506</td>
</tr>
<tr>
<td>Non-Word Spelling</td>
<td>Comparison</td>
<td>$M$</td>
<td>14.025</td>
<td>0.3451</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$SD$</td>
<td>3.1063</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Skewness</td>
<td>-0.229</td>
<td>0.267</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Kurtosis</td>
<td>0.341</td>
<td>0.529</td>
</tr>
<tr>
<td></td>
<td>Treatment</td>
<td>$M$</td>
<td>16.438</td>
<td>0.3212</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$SD$</td>
<td>3.03</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Skewness</td>
<td>-1.215</td>
<td>0.255</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Kurtosis</td>
<td>2.761</td>
<td>0.506</td>
</tr>
<tr>
<td>Reading Fluency</td>
<td>Comparison</td>
<td>$M$</td>
<td>45.593</td>
<td>2.2861</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$SD$</td>
<td>20.5747</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Skewness</td>
<td>-0.154</td>
<td>0.267</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Kurtosis</td>
<td>-0.337</td>
<td>0.529</td>
</tr>
<tr>
<td></td>
<td>Treatment</td>
<td>$M$</td>
<td>63.169</td>
<td>2.0015</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$SD$</td>
<td>18.8826</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Skewness</td>
<td>-0.581</td>
<td>0.255</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Kurtosis</td>
<td>0.105</td>
<td>0.506</td>
</tr>
</tbody>
</table>

*Note.* DV = dependent variable; $M$ = mean; $SD$ = standard deviation; and $SE$ = standard error. Skewness refers to excess skewness.
Table 5

**Shapiro-Wilk Test of Normality**

<table>
<thead>
<tr>
<th>DV</th>
<th>Condition</th>
<th>Statistic</th>
<th>df</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Read Word Reading</td>
<td>Comparison</td>
<td>.981</td>
<td>81</td>
<td>.264</td>
</tr>
<tr>
<td></td>
<td>Treatment</td>
<td>.982</td>
<td>89</td>
<td>.272</td>
</tr>
<tr>
<td>Real Word Spelling</td>
<td>Comparison</td>
<td>.964*</td>
<td>81</td>
<td>.024</td>
</tr>
<tr>
<td></td>
<td>Treatment</td>
<td>.985</td>
<td>89</td>
<td>.415</td>
</tr>
<tr>
<td>Non-Word Reading</td>
<td>Comparison</td>
<td>.984</td>
<td>81</td>
<td>.433</td>
</tr>
<tr>
<td></td>
<td>Treatment</td>
<td>.969*</td>
<td>89</td>
<td>.033</td>
</tr>
<tr>
<td>Non-Word Spelling</td>
<td>Comparison</td>
<td>.975</td>
<td>81</td>
<td>.121</td>
</tr>
<tr>
<td></td>
<td>Treatment</td>
<td>.912**</td>
<td>89</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Reading Fluency</td>
<td>Comparison</td>
<td>.984</td>
<td>81</td>
<td>.437</td>
</tr>
<tr>
<td></td>
<td>Treatment</td>
<td>.972</td>
<td>89</td>
<td>.051</td>
</tr>
</tbody>
</table>

Note. df = degrees of freedom; DV = dependent variable; Sig. = significance.
** Significant at the p < .01 level; * significant at the p < .05 level.

The author attempted to reduce the leptokurtosis and skewness of the relevant distributions by performing several transformations; however none was sufficient to normalize the data (i.e., none was sufficient to render the relevant results of the Shapiro-Wilk test of normality non-significant). The author also considered the use of non-parametric tests; however, no suitable non-parametric method of data analysis could be identified. Therefore, the author elected to proceed with data screening and, subsequently, the full MANCOVA. This decision was taken because multivariate analyses of variance (MANOVA) and related tests (e.g., MANCOVA) are considered robust to violations of normality if groups are of roughly comparable size, that is the size of the largest group is no more than 1.5 times larger than the size of the smallest group (Leech, Barrett, & Morgan, 2005).

To screen for collinearity among dependent variables, which might indicate that two or more variables measured the same construct, the author used IBM SPSS
CORELATE to perform a Pearson’s product-moment correlation analysis identifying the relationship between each dependent variable and all others. All of the variables were significantly correlated to all others at the p < .01 level, with correlation coefficients ranging in strength from $R^2 = .519$ to $R^2 = .795$. Based on these results, the author elected to perform a principal component analysis (PCA), a factor reduction procedure (Tabachnick & Fidell, 2013), to determine if any latent variables might explain the strong correlations among the dependent variables. Table 6 provides the results of the Pearson’s product-moment correlation analysis performed by the author.

Table 6

<table>
<thead>
<tr>
<th>Pearson’s Product-Moment Correlation Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>DV</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Real Word</td>
</tr>
<tr>
<td>Reading</td>
</tr>
<tr>
<td>$R^2$</td>
</tr>
<tr>
<td>Sig.</td>
</tr>
<tr>
<td>$N$</td>
</tr>
<tr>
<td>Real Word Spelling</td>
</tr>
<tr>
<td>$R^2$</td>
</tr>
<tr>
<td>Sig.</td>
</tr>
<tr>
<td>$N$</td>
</tr>
<tr>
<td>Non-Word Reading</td>
</tr>
<tr>
<td>$R^2$</td>
</tr>
<tr>
<td>Sig.</td>
</tr>
<tr>
<td>$N$</td>
</tr>
<tr>
<td>Non-Word Spelling</td>
</tr>
<tr>
<td>$R^2$</td>
</tr>
<tr>
<td>Sig.</td>
</tr>
<tr>
<td>$N$</td>
</tr>
<tr>
<td>Reading Fluency</td>
</tr>
<tr>
<td>$R^2$</td>
</tr>
<tr>
<td>Sig.</td>
</tr>
<tr>
<td>$N$</td>
</tr>
</tbody>
</table>

Note. DV = dependent variable; $R^2 = $ Pearson’s product moment correlation coefficient.

** Significant at the p < .01 level; * significant at the p < .05 level.
A PCA with Varimax (orthogonal) rotation was performed using IMB SPSS DATA REDUCTION on dependent variable data gathered from 170 complete cases. An examination of the Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy (KMO = .853) indicated that the data were suitable for factor analysis per the guidelines suggested by Tabatchnick and Fidell (2013). This finding was reinforced by the extremely high value generated by Bartlett's test of sphericity [$\chi^2(10) = 672.814, p < .001$]. Perhaps unsurprisingly, the results of the PCA were indicative of a single latent variable, or component, upon which the study’s five dependent variables loaded. This component, obviously associated with participants’ overall language arts achievement, explained over 76% of the variance across the five dependent variables.

While the results of the factor analysis pointed to a single latent variable underlying performance across the dependent variables measured, the author felt it would be undesirable to combine all of the dependent variables into one variable representative of achievement across measures. First, to conflate all of the variables into one would necessarily conflate their error, rendering subsequent analyses less precise. Second, the high degree of correlation among the dependent variables was unsurprising, as the intercorrelations\textsuperscript{26} reported for the WJ IV ACH (Schrank et al., 2014) tests used to measure four of five of the dependent variables ranged from $R^2 = .66$ to $R^2 = .82$, very closely mirroring the findings of the Pearson's product moment correlation analysis. Furthermore, the primary reason to avoid including highly correlated variables is that high levels of correlation among dependent variables can lead to a reduction in statistical power (Foster, Barkus, & Yavorosky, 2006); however, in situations in which multivariate analysis is required to reduce the possibility of a Type I error, such a loss of power can be

\textsuperscript{26} Among norm group participants aged six to eight years.
considered an acceptable trade-off (Foster et al., 2006). Therefore, the author chose to proceed with the full MANCOVA.

**Identification of covariates.** To determine whether significant differences existed between groups on pre-test achievement scores and/or demographic variables, univariate analyses of variance (ANOVA) were performed to compare mean scores between groups (i.e., treatment and comparison) for each dependent variable and demographic variable. The author elected to determine significance at the p < .10 level to ensure that any differences between groups that might impact dependent variable measures would be adequately accounted for in subsequent analyses. While the author understood that the introduction of additional covariates into the model would weaken statistical power, the author felt that a more conservative approach was warranted, given the lack of random assignment to groups. The ANOVA analyses led to the identification of several covariates. Significant differences favoring the treatment group were detected at the p < .10 level in the pre-test scores for real word reading \( F(1, 169) = 2.927, p = .089 \), and non-word reading \( F(1, 169) = 3.314, p = .070 \). Significant differences were also detected at the p < .10 level English language learner (ELL) status \( F(1,169) = 11.480, p = .001 \) and special education (SPED) status \( F(1,169) = 3.421, p = .066 \). Therefore, a decision was taken to add pre-test achievement, ELL status, and SPED status to the MANCOVA model as covariates.

**Data Analysis**

**Test procedure.** A multivariate analysis of covariance (MANCOVA) was performed to detect mean differences in (a) non-word reading, (b) real word reading, (c) non-word spelling, (d) real word spelling, and/or (e) reading fluency post-test.

---

27 Data analyses were performed using IBM SPSS GENERAL LINEAR MODEL unless otherwise noted.
achievement scores of students assigned to use the MVRC online reading intervention and those of students assigned to a business-as-usual comparison condition (regular classroom reading instruction alone), once the effect of differences in pre-test achievement scores and relevant demographic variables had been accounted for.

Tests of equality of variance and covariance. Levene’s test for equality of error variance was non-significant for each of the study’s dependent variables at the p < .05 level, as was Box’s test of equality of covariance matrices at the p < .05 level, indicating that the data did not violate the assumption of homogeneity of variance. Therefore, consistent with the recommendations of Tabachnick and Fidell (2013), the author chose to use Wilk’s lambda as the test statistic in subsequent multivariate analyses. The results of Box’s test of equality of covariance matrices and Levene’s test for equality of variance are displayed in Table 7 and Table 8, respectively.

Table 7

<table>
<thead>
<tr>
<th>Box's Test of Equality of Covariance Matrices</th>
</tr>
</thead>
<tbody>
<tr>
<td>Box's M</td>
</tr>
<tr>
<td>F</td>
</tr>
<tr>
<td>df 1</td>
</tr>
<tr>
<td>df 2</td>
</tr>
<tr>
<td>Sig.</td>
</tr>
</tbody>
</table>

Note. df = degrees of freedom; F = Fisher's F ratio; Sig. = significance.

Table 8

<table>
<thead>
<tr>
<th>Levene's Test of Equality of Error Variances</th>
</tr>
</thead>
<tbody>
<tr>
<td>DV</td>
</tr>
<tr>
<td>Real Word Reading</td>
</tr>
<tr>
<td>Real Word Spelling</td>
</tr>
<tr>
<td>Non-Word Reading</td>
</tr>
<tr>
<td>Non-Word Spelling</td>
</tr>
<tr>
<td>Reading Fluency</td>
</tr>
</tbody>
</table>

Note. df = degrees of freedom; F = Fisher's F ratio; Sig. = significance.
Multivariate tests. The results of the MANCOVA analysis revealed a significant main effect ($\lambda = .668$, $F[5, 161] = 16.014$, $p < .001$, multivariate $\eta^2 = .332$) of the intervention on the achievement scores of participants assigned to the treatment condition relative to those assigned to the business-as-usual comparison condition, once the effects of differences in pre-test achievement scores, special education (SPED) status, and English language learner (ELL) status had been accounted for. The main effects of ELL status ($\lambda = .868$, $F[5, 161] = 4.903$, $p < .001$, multivariate $\eta^2 = .132$) and pre-test scores ($\lambda = .212$, $F[5, 161] = 119.487$, $p < .001$, multivariate $\eta^2 = .788$) indicated significant effects on the combined dependent variables (DV). Special education status, however, had no significant effect on the combined DVs. Table 9 contains the results of the multivariate tests for the study’s independent variable and each of its covariates.

Table 9

Multivariate Tests

<table>
<thead>
<tr>
<th>IV/Covariate</th>
<th>Test</th>
<th>Value</th>
<th>$F$</th>
<th>Hypothesis $df$</th>
<th>Error $df$</th>
<th>Sig.</th>
<th>Partial $\eta^2$</th>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>SPED Status</td>
<td>Pillai's race</td>
<td>.056</td>
<td>1.920</td>
<td>5</td>
<td>161</td>
<td>.094</td>
<td>.056</td>
<td>.639</td>
</tr>
<tr>
<td></td>
<td>Wilks' Lambda</td>
<td>.944</td>
<td>1.920</td>
<td>5</td>
<td>161</td>
<td>.094</td>
<td>.056</td>
<td>.639</td>
</tr>
<tr>
<td>ELL Status</td>
<td>Pillai's Trace</td>
<td>.132**</td>
<td>4.903</td>
<td>5</td>
<td>161</td>
<td>&lt;.001</td>
<td>.132</td>
<td>.979</td>
</tr>
<tr>
<td></td>
<td>Wilks' Lambda</td>
<td>.868**</td>
<td>4.903</td>
<td>5</td>
<td>161</td>
<td>&lt;.001</td>
<td>.132</td>
<td>.979</td>
</tr>
<tr>
<td>Pre-Test Scores</td>
<td>Pillai's Trace</td>
<td>.788**</td>
<td>119.478</td>
<td>5</td>
<td>161</td>
<td>&lt;.001</td>
<td>.788</td>
<td>1.000</td>
</tr>
<tr>
<td></td>
<td>Wilks' Lambda</td>
<td>.212**</td>
<td>119.478</td>
<td>5</td>
<td>161</td>
<td>&lt;.001</td>
<td>.788</td>
<td>1.000</td>
</tr>
<tr>
<td>Condition</td>
<td>Pillai's Trace</td>
<td>.332**</td>
<td>16.014</td>
<td>5</td>
<td>161</td>
<td>&lt;.001</td>
<td>.332</td>
<td>1.000</td>
</tr>
<tr>
<td></td>
<td>Wilks' Lambda</td>
<td>.668**</td>
<td>16.014</td>
<td>5</td>
<td>161</td>
<td>&lt;.001</td>
<td>.332</td>
<td>1.000</td>
</tr>
</tbody>
</table>

Note. $df$ = degrees of freedom; $F$ = Fisher's $F$ ratio; IV = independent variable; Sig. = significance.
** Significant at the $p < .01$ level; * significant at the $p < .05$ level.
**Univariate tests.** Post-hoc ANCOVAs were performed to isolate the relationship between each dependent variable, the combined pre-test covariate, and the grouping variable (condition). The multivariate test results were confirmed by the post-hoc ANCOVAs, performed using Bonferroni’s correction for multiple comparisons, across four of the study’s five dependent variables: real word spelling ($F[1, 165] = 16.341, p < .001, \text{multivariate } \eta^2 = .090$), non-word reading ($F[1, 165] = 4.368, p = .038, \text{multivariate } \eta^2 = .026$), non-word spelling ($F[1, 165] = 29.212, p = .001, \text{multivariate } \eta^2 = .150$), and reading fluency ($F[1, 165] = 58.348, p < .001, \text{multivariate } \eta^2 = .261$). Statistically significant effects of the intervention, however, were not detected in real word reading ($F[1, 165] = 2.328, p = .129, \text{multivariate } \eta^2 = .014$).

Preliminary data analyses revealed significant differences between groups in special education status, English language learner (ELL) status, and pre-test achievement scores; however, tests of between-subjects effects revealed that only the pre-test score covariate was significant at the $p < .05$ level for all dependent variables. Special education status, by contrast, was significant at the $p < .05$ level relative to real word reading alone, while ELL status was significant at the $p < .05$ level only in relationship to non-word reading. Table 10 contains the parameter estimates for each dependent variable relative to the independent variable and each covariate. Table 11 displays pairwise comparisons between groups for each dependent variable.
<table>
<thead>
<tr>
<th>IV/Covariate</th>
<th>DV</th>
<th>df</th>
<th>$M^2$</th>
<th>$F$</th>
<th>Sig.</th>
<th>Partial $\eta^2$</th>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>SPED Status</td>
<td>Real Word Reading</td>
<td>1</td>
<td>109.027*</td>
<td>3.899</td>
<td>.050</td>
<td>.023</td>
<td>.501</td>
</tr>
<tr>
<td></td>
<td>Real Word Spelling</td>
<td>1</td>
<td>.570</td>
<td>.073</td>
<td>.787</td>
<td>&lt;.001</td>
<td>.058</td>
</tr>
<tr>
<td></td>
<td>Non-Word Reading</td>
<td>1</td>
<td>.158</td>
<td>.013</td>
<td>.909</td>
<td>&lt;.001</td>
<td>.051</td>
</tr>
<tr>
<td></td>
<td>Non-Word Spelling</td>
<td>1</td>
<td>13.429</td>
<td>2.481</td>
<td>.117</td>
<td>.015</td>
<td>.347</td>
</tr>
<tr>
<td></td>
<td>Reading Fluency</td>
<td>1</td>
<td>6.461</td>
<td>.045</td>
<td>.831</td>
<td>&lt;.001</td>
<td>.055</td>
</tr>
<tr>
<td>ELL Status</td>
<td>Real Word Reading</td>
<td>1</td>
<td>55.407</td>
<td>1.982</td>
<td>.161</td>
<td>.012</td>
<td>.288</td>
</tr>
<tr>
<td></td>
<td>Real Word Spelling</td>
<td>1</td>
<td>.018</td>
<td>.002</td>
<td>.962</td>
<td>&lt;.001</td>
<td>.050</td>
</tr>
<tr>
<td></td>
<td>Non-Word Reading</td>
<td>1</td>
<td>75.392*</td>
<td>6.220</td>
<td>.014</td>
<td>.036</td>
<td>.698</td>
</tr>
<tr>
<td></td>
<td>Non-Word Spelling</td>
<td>1</td>
<td>1.944</td>
<td>.359</td>
<td>.550</td>
<td>.002</td>
<td>.092</td>
</tr>
<tr>
<td></td>
<td>Reading Fluency</td>
<td>1</td>
<td>171.255</td>
<td>1.205</td>
<td>.274</td>
<td>.007</td>
<td>.194</td>
</tr>
<tr>
<td>Pre-Test Scores</td>
<td>Real Word Reading</td>
<td>1</td>
<td>8497.432**</td>
<td>303.921</td>
<td>&lt;.001</td>
<td>.648</td>
<td>1.000</td>
</tr>
<tr>
<td></td>
<td>Real Word Spelling</td>
<td>1</td>
<td>3627.206**</td>
<td>466.969</td>
<td>&lt;.001</td>
<td>.739</td>
<td>1.000</td>
</tr>
<tr>
<td></td>
<td>Non-Word Reading</td>
<td>1</td>
<td>1375.069**</td>
<td>113.452</td>
<td>&lt;.001</td>
<td>.407</td>
<td>1.000</td>
</tr>
<tr>
<td></td>
<td>Non-Word Spelling</td>
<td>1</td>
<td>529.461**</td>
<td>97.817</td>
<td>&lt;.001</td>
<td>.372</td>
<td>1.000</td>
</tr>
<tr>
<td></td>
<td>Reading Fluency</td>
<td>1</td>
<td>36566.981**</td>
<td>257.198</td>
<td>&lt;.001</td>
<td>.609</td>
<td>1.000</td>
</tr>
<tr>
<td>Condition</td>
<td>Real Word Reading</td>
<td>1</td>
<td>65.082</td>
<td>2.328</td>
<td>.129</td>
<td>.014</td>
<td>.329</td>
</tr>
<tr>
<td></td>
<td>Real Word Spelling</td>
<td>1</td>
<td>126.926**</td>
<td>16.341</td>
<td>&lt;.001</td>
<td>.090</td>
<td>.980</td>
</tr>
<tr>
<td></td>
<td>Non-Word Reading</td>
<td>1</td>
<td>52.940*</td>
<td>4.368</td>
<td>.038</td>
<td>.026</td>
<td>.547</td>
</tr>
<tr>
<td></td>
<td>Non-Word Spelling</td>
<td>1</td>
<td>158.118**</td>
<td>29.212</td>
<td>&lt;.001</td>
<td>.150</td>
<td>1.000</td>
</tr>
<tr>
<td></td>
<td>Reading Fluency</td>
<td>1</td>
<td>8295.640**</td>
<td>58.348</td>
<td>&lt;.001</td>
<td>.261</td>
<td>1.000</td>
</tr>
</tbody>
</table>

Note. $df$ = degrees of freedom; DV = dependent variable; $F$ = Fisher's $F$ ratio; IV = independent variable; $M^2$ = mean squared; Sig. = significance.

** Significant at the p < .01 level; * significant at the p < .05 level.
Table 11

*Pairwise Comparisons for Univariate Tests*

<table>
<thead>
<tr>
<th>Dependent Variable</th>
<th>Condition</th>
<th>Mean</th>
<th>Mean Difference</th>
<th>Standard Error</th>
<th>Sig.</th>
<th>Lower Bound</th>
<th>Upper Bound</th>
</tr>
</thead>
<tbody>
<tr>
<td>Real Word</td>
<td>Comparison</td>
<td>42.901</td>
<td>-1.297</td>
<td>.850</td>
<td>.129</td>
<td>-2.976</td>
<td>.382</td>
</tr>
<tr>
<td>Reading</td>
<td>Treatment</td>
<td>45.944</td>
<td>1.297</td>
<td>.850</td>
<td>.129</td>
<td>-.382</td>
<td>2.976</td>
</tr>
<tr>
<td>Real Word</td>
<td>Comparison</td>
<td>20.383</td>
<td>-1.811**</td>
<td>.448</td>
<td>&lt;.001</td>
<td>-2.696</td>
<td>-.927</td>
</tr>
<tr>
<td>Spelling</td>
<td>Treatment</td>
<td>23.326</td>
<td>1.811**</td>
<td>.448</td>
<td>&lt;.001</td>
<td>.927</td>
<td>2.696</td>
</tr>
<tr>
<td>Non-Word Reading</td>
<td>Comparison</td>
<td>18.136</td>
<td>-1.170*</td>
<td>.560</td>
<td>.038</td>
<td>-2.275</td>
<td>-.065</td>
</tr>
<tr>
<td>Treatment</td>
<td>19.629</td>
<td>1.170*</td>
<td>.560</td>
<td>.038</td>
<td>.065</td>
<td>2.275</td>
<td></td>
</tr>
<tr>
<td>Non-Word Reading</td>
<td>Comparison</td>
<td>14.025</td>
<td>-2.022**</td>
<td>.374</td>
<td>&lt;.001</td>
<td>-2.760</td>
<td>-1.283</td>
</tr>
<tr>
<td>Spelling</td>
<td>Treatment</td>
<td>16.438</td>
<td>2.022**</td>
<td>.374</td>
<td>&lt;.001</td>
<td>1.283</td>
<td>2.760</td>
</tr>
<tr>
<td>Reading Fluency</td>
<td>Comparison</td>
<td>45.593</td>
<td>-14.645**</td>
<td>1.917</td>
<td>&lt;.001</td>
<td>-18.430</td>
<td>-10.859</td>
</tr>
<tr>
<td>Treatment</td>
<td>63.169</td>
<td>14.645**</td>
<td>1.917</td>
<td>&lt;.001</td>
<td>10.859</td>
<td>18.430</td>
<td></td>
</tr>
</tbody>
</table>

Note. Sig. = significance.
** Significant at the p < .01 level; * significant at the p < .05 level.

**Additional post-hoc tests.** Post-hoc ANCOVAs were performed to isolate the relationship between each dependent variable, its corresponding pre-test covariate, and the grouping variable (condition). The first of these ANCOVAs was conducted to confirm the results of the previous analyses relative to the study’s real word reading dependent variable. The results of the ANCOVA indicated no statistically significant differences between groups at the p < .05 level on real word reading, once the effects of differences in pre-test achievement scores, special education status, and English language learner (ELL) status had been accounted for. This result was consistent with that of the previous analyses, thus confirming the author’s original finding. Also consistent with previous univariate analyses, the dependent variable was significantly affected by the pre-test score covariate at the p < .05 level; however, in contrast to previous univariate

---

28 It was not possible to perform Bonferroni’s correction for multiple comparisons in the initial analyses. Because a total of five comparisons were performed, only significance values of p < .01 were treated as truly significant.
analyses, the SPED status covariate was not significant at the p < .05 level. Table 12 contains the parameter estimates for the dependent variable relative to the independent variable and each covariate, while Table 13 displays pairwise comparisons between groups for the dependent variable.

Table 12

**Between-Subjects Effects for Real Word Reading**

<table>
<thead>
<tr>
<th>IV/Covariate</th>
<th>df</th>
<th>$M^2$</th>
<th>$F$</th>
<th>Sig.</th>
<th>Partial $\eta^2$</th>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>SPED Status</td>
<td>1</td>
<td>84.106</td>
<td>3.406</td>
<td>.067</td>
<td>.020</td>
<td>.450</td>
</tr>
<tr>
<td>ELL Status</td>
<td>1</td>
<td>27.339</td>
<td>1.107</td>
<td>.294</td>
<td>.007</td>
<td>.182</td>
</tr>
<tr>
<td>Pre-Test Scores</td>
<td>1</td>
<td>9019.412**</td>
<td>365.273</td>
<td>&lt;.001</td>
<td>.688</td>
<td>1.000</td>
</tr>
<tr>
<td>Condition</td>
<td>1</td>
<td>38.506</td>
<td>1.559</td>
<td>.214</td>
<td>.009</td>
<td>.237</td>
</tr>
</tbody>
</table>

Note. $df$ = degrees of freedom; $M^2$ = mean squared; $F$ = Fisher's $F$ ratio; IV = independent variable; $M^2$ = mean squared; Sig. = significance.

** Significant at the p < .01 level.

Table 13

**Pairwise Comparisons for Real Word Reading**

<table>
<thead>
<tr>
<th>Dependent Variable</th>
<th>Condition</th>
<th>Mean Difference</th>
<th>Standard Error</th>
<th>Sig.</th>
<th>Lower Bound</th>
<th>Upper Bound</th>
<th>95% Confidence Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>Real Word Reading</td>
<td>Comparison</td>
<td>-.998</td>
<td>.799</td>
<td>.214</td>
<td>-2.575</td>
<td>.580</td>
<td></td>
</tr>
<tr>
<td>Real Word Reading</td>
<td>Treatment</td>
<td>.998</td>
<td>.799</td>
<td>.214</td>
<td>-2.575</td>
<td>.580</td>
<td></td>
</tr>
</tbody>
</table>

Note. Sig. = significance.

The second ANCOVA, examining differences between groups on real word spelling confirmed the results of previous analyses. The ANCOVA indicated that there were significant differences between groups at the p < .05 level in real word reading, once the effects of differences in pre-test achievement scores, special education status, and English language learner (ELL) status had been accounted for. Also consistent with previous univariate analyses, only the pre-test score covariate significantly affected the
dependent variable at the p < .05 level. Table 14 displays the parameter estimates for the dependent variable relative to the independent variable and each covariate, and Table 15 shows pairwise comparisons between groups for the dependent variable.

Table 14

<table>
<thead>
<tr>
<th>IV/Covariate</th>
<th>df</th>
<th>M^2</th>
<th>F</th>
<th>Sig.</th>
<th>Partial η2</th>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>SPED Status</td>
<td>1</td>
<td>48.601</td>
<td>3.762</td>
<td>.054</td>
<td>.022</td>
<td>.487</td>
</tr>
<tr>
<td>ELL Status</td>
<td>1</td>
<td>.258</td>
<td>.020</td>
<td>.888</td>
<td>&lt;.001</td>
<td>.052</td>
</tr>
<tr>
<td>Pre-Test Scores</td>
<td>1</td>
<td>2784.994**</td>
<td>215.557</td>
<td>&lt;.001</td>
<td>.565</td>
<td>1.000</td>
</tr>
<tr>
<td>Condition</td>
<td>1</td>
<td>240.537**</td>
<td>18.617</td>
<td>&lt;.001</td>
<td>.101</td>
<td>.990</td>
</tr>
</tbody>
</table>

Note. df = degrees of freedom; F = Fisher's F ratio; IV = independent variable; M^2 = mean squared; Sig. = significance. ** Significant at the p < .01 level.

Table 15

<table>
<thead>
<tr>
<th>Dependent Variable</th>
<th>Condition</th>
<th>Mean Difference</th>
<th>Standard Error</th>
<th>Sig.</th>
<th>Lower Bound</th>
<th>Upper Bound</th>
</tr>
</thead>
<tbody>
<tr>
<td>Real Word Spelling</td>
<td>Comparison</td>
<td>-2.473**</td>
<td>.573</td>
<td>&lt;.001</td>
<td>-3.604</td>
<td>-1.341</td>
</tr>
<tr>
<td></td>
<td>Treatment</td>
<td>2.473**</td>
<td>.573</td>
<td>&lt;.001</td>
<td>1.341</td>
<td>3.604</td>
</tr>
</tbody>
</table>

Note. Sig. = significance. ** Significant at the p < .01 level.

The results of an ANCOVA examining differences between groups on non-word reading were inconsistent with previous analyses. The ANCOVA failed to demonstrate significant differences between groups at the p < .05 level in non-word reading, once the effects of differences in pre-test achievement scores, special education status, and English language learner (ELL) status had been accounted for. Table 16 displays the parameter estimates for the dependent variable relative to the independent variable and each covariate, and Table 17 shows pairwise comparisons between groups for the dependent variable.
variable.

Table 16

Between-Subjects Effects for Non-Word Reading

<table>
<thead>
<tr>
<th>IV/Covariate</th>
<th>df</th>
<th>$M^2$</th>
<th>$F$</th>
<th>Sig.</th>
<th>Partial $\eta^2$</th>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>SPED Status</td>
<td>1</td>
<td>11.600</td>
<td>.882</td>
<td>.349</td>
<td>.005</td>
<td>.154</td>
</tr>
<tr>
<td>ELL Status</td>
<td>1</td>
<td>26.825</td>
<td>2.040</td>
<td>.155</td>
<td>.012</td>
<td>.295</td>
</tr>
<tr>
<td>Pre-Test Scores</td>
<td>1</td>
<td>1196.359**</td>
<td>90.997</td>
<td>&lt;.001</td>
<td>.354</td>
<td>1.000</td>
</tr>
<tr>
<td>Condition</td>
<td>1</td>
<td>37.548</td>
<td>2.856</td>
<td>.093</td>
<td>.017</td>
<td>.390</td>
</tr>
</tbody>
</table>

Note. $df =$ degrees of freedom; $F =$ Fisher's $F$ ratio; $M^2 =$ mean squared; Sig. = significance.

** Significant at the p < .01 level.

Table 17

Pairwise Comparisons for Non-Word Reading

<table>
<thead>
<tr>
<th>Dependent Variable</th>
<th>Condition</th>
<th>Mean Difference</th>
<th>Standard Error</th>
<th>Sig.</th>
<th>Lower Bound</th>
<th>Upper Bound</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-Word Reading</td>
<td>Comparison</td>
<td>-.989</td>
<td>.585</td>
<td>.093</td>
<td>-2.145</td>
<td>.166</td>
</tr>
<tr>
<td></td>
<td>Treatment</td>
<td>.989</td>
<td>.585</td>
<td>.093</td>
<td>-.166</td>
<td>2.145</td>
</tr>
</tbody>
</table>

Note. Sig. = significance.

An ANCOVA examining differences between groups on non-word spelling also mirrored previous analyses. The results revealed significant effects of the intervention on non-word spelling, once the effects of differences in pre-test achievement scores, special education status, and English language learner (ELL) status had been accounted for.

While pre-test scores remained significant at the p < .05 level, ELL status and SPED status were non-significant. Table 18 displays the parameter estimates for the dependent variable relative to the independent variable and each covariate, and Table 19 shows pairwise comparisons between groups for the dependent variable.
Table 18

**Between-Subjects Effects for Non-Word Spelling**

<table>
<thead>
<tr>
<th>IV/Covariate</th>
<th>df</th>
<th>$M^2$</th>
<th>$F$</th>
<th>Sig.</th>
<th>Partial $\eta^2$</th>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>SPED Status</td>
<td>1</td>
<td>21.676</td>
<td>3.234</td>
<td>.074</td>
<td>.019</td>
<td>.432</td>
</tr>
<tr>
<td>ELL Status</td>
<td>1</td>
<td>19.943</td>
<td>2.976</td>
<td>.086</td>
<td>.018</td>
<td>.403</td>
</tr>
<tr>
<td>Pre-Test Scores</td>
<td>1</td>
<td>316.785**</td>
<td>47.269</td>
<td>&lt;.001</td>
<td>.223</td>
<td>1.000</td>
</tr>
<tr>
<td>Condition</td>
<td>1</td>
<td>206.257**</td>
<td>30.777</td>
<td>&lt;.001</td>
<td>.157</td>
<td>1.000</td>
</tr>
</tbody>
</table>

Note. $df =$ degrees of freedom; $F =$ Fisher's $F$ ratio; $M^2 =$ mean squared; Sig. = significance.

** Significant at the $p < .01$ level.

Table 19

**Pairwise Comparisons for Non-Word Spelling**

<table>
<thead>
<tr>
<th>Dependent Variable</th>
<th>Condition</th>
<th>Mean Difference</th>
<th>Standard Error</th>
<th>Sig.</th>
<th>Lower Bound</th>
<th>Upper Bound</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-Word Spelling</td>
<td>Comparison</td>
<td>-2.296**</td>
<td>.414</td>
<td>&lt;.001</td>
<td>-3.112</td>
<td>-1.479</td>
</tr>
<tr>
<td></td>
<td>Treatment</td>
<td>2.296**</td>
<td>.414</td>
<td>&lt;.001</td>
<td>1.479</td>
<td>3.112</td>
</tr>
</tbody>
</table>

Note. Sig. = significance.

** Significant at the $p < .01$ level.

The results of the ANCOVA evaluating differences between groups on reading fluency reflected the findings of previous analyses. The ANCOVA revealed significant effects of the intervention on reading fluency, once the effects of differences in pre-test achievement scores, special education status, and English language learner (ELL) status had been accounted for. Unlike previous univariate analyses, however, the ANCOVA revealed significant effects at the $p < .05$ level for all the covariates (pre-test scores, ELL status, and SPED status), thus supporting their introduction into the model. Table 20 displays the parameter estimates for the dependent variable relative to the independent variable and each covariate, and Table 21 shows pairwise comparisons between groups for the dependent variable.
Table 20

**Between-Subjects Effects for Reading Fluency**

<table>
<thead>
<tr>
<th>IV/Covariate</th>
<th>df</th>
<th>( M^2 )</th>
<th>( F )</th>
<th>Sig.</th>
<th>Partial ( \eta^2 )</th>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>SPED Status</td>
<td>1</td>
<td>798.743*</td>
<td>5.857</td>
<td>.017</td>
<td>.034</td>
<td>.672</td>
</tr>
<tr>
<td>ELL Status</td>
<td>1</td>
<td>663.482*</td>
<td>4.865</td>
<td>.029</td>
<td>.028</td>
<td>.592</td>
</tr>
<tr>
<td>Pre-Test Scores</td>
<td>1</td>
<td>38039.169**</td>
<td>278.937</td>
<td>&lt;.001</td>
<td>.627</td>
<td>1</td>
</tr>
<tr>
<td>Condition</td>
<td>1</td>
<td>9691.512**</td>
<td>71.067</td>
<td>&lt;.001</td>
<td>.300</td>
<td>1.000</td>
</tr>
</tbody>
</table>

Note. \( df \) = degrees of freedom; \( F \) = Fisher's \( F \) ratio; IV = independent variable; \( M^2 \) = mean squared; Sig. = significance.
** Significant at the \( p < .01 \) level; * significant at the \( p < .05 \) level.

Table 21

**Pairwise Comparisons for Reading Fluency**

<table>
<thead>
<tr>
<th>Dependent Variable</th>
<th>Condition</th>
<th>Mean Difference</th>
<th>Standard Error</th>
<th>Sig.</th>
<th>Lower Bound</th>
<th>Upper Bound</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reading Fluency</td>
<td>Comparison</td>
<td>-15.720**</td>
<td>1.865</td>
<td>&lt;.001</td>
<td>-19.402</td>
<td>-12.039</td>
</tr>
<tr>
<td></td>
<td>Treatment</td>
<td>15.720**</td>
<td>1.865</td>
<td>&lt;.001</td>
<td>12.039</td>
<td>19.402</td>
</tr>
</tbody>
</table>

Note. Sig. = significance.
** Significant at the \( p < .01 \) level.

Fidelity of Implementation

**Product usage logs.** Product usage logs generated by the MVRC software indicated that active participant use of the intervention exceeded 90% of the time assigned (a mean of 44 hours of active use) among treatment group participants who remained in the study until post-test.

**Behavioral observations.** Over the course of the intervention, treatment group participants were observed 45 times while using the MVRC online intervention, or approximately twice per week. Observers used a novel planned activity check instrument to record behavioral observations. Data from this instrument were analyzed by dividing the number of students displaying engagement behavior at each interval by the total
number of students observed and multiplying the dividend by 100 to determine the percentage of students displaying engagement behavior. This information was averaged for the entire twenty-minute observation period and then aggregated and averaged by month. Inter-observer agreement was determined by dividing the number of agreements between observers per interval by the total number of agreements plus disagreements and multiplying the dividend by 10.

The results of the behavioral observations indicated that treatment group participants displayed engagement behavior over 90% of the intervals observed while using the MVRC online intervention. Inter-observer agreement, however, was within the acceptable range at just under 84%. Table 22 displays the results of the planned activity check.

Table 22

<table>
<thead>
<tr>
<th></th>
<th>Percentage of Engagement Behavior</th>
<th>Inter-Observable Agreement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oct-13</td>
<td>91.50%</td>
<td>84.50%</td>
</tr>
<tr>
<td>Nov-13</td>
<td>91%</td>
<td>77%</td>
</tr>
<tr>
<td>Dec-13</td>
<td>95%</td>
<td>86%</td>
</tr>
<tr>
<td>Jan-14</td>
<td>93%</td>
<td>81.50%</td>
</tr>
<tr>
<td>Feb-14</td>
<td>91%</td>
<td>85.50%</td>
</tr>
<tr>
<td>Mar-14</td>
<td>92%</td>
<td>88%</td>
</tr>
<tr>
<td>Mean</td>
<td>91.50%</td>
<td>83.67%</td>
</tr>
</tbody>
</table>
CHAPTER FIVE: RESEARCH REPORT

In this chapter, the author presents the results of a formal report documenting the present study and its findings. The report is presented at a publishable paper. Thus, consistent with journal conventions, the format differs from that of previous chapters.

A quasi-experimental research design was used to evaluate the efficacy of MindPlay Virtual Reading Coach (MVRC), an ICT-based reading intervention, in addition to regular daily language instruction provided by a classroom teacher. After attrition, participants included 170 students enrolled in eight second-grade classrooms (four classrooms in each school) in two public elementary schools in the southwestern United States. Examiners obtained reading achievement data from each participating student. Measures included tests from the Woodcock-Johnson Tests of Achievement (WJ IV ACH), as well as the Test of Silent Word Reading Fluency (TOSWRF-2). A multivariate analysis of covariance (MANCOVA) was used to determine whether there were significant mean differences in (a) non-word reading, (b) real word reading, (c) non-word spelling, (d) real word spelling, and/or (e) reading fluency scores favoring students assigned to use the MVRC online reading intervention, once the effects of differences in pre-test achievement scores and relevant demographic variables had been accounted for. Analyses revealed a significant main effect ($\lambda = .668, F[5, 161] = 16.014, p < .0001$, multivariate $\eta^2 = .332$) of the intervention on achievement scores of participants assigned to the treatment condition, which was confirmed across three of the study’s independent variables: real word spelling ($F[1, 165] = 16.341, p < .0001$, multivariate $\eta^2 = .090$), non-word spelling ($F[1, 165] = 29.212, p < .0001$, multivariate $\eta^2 = .150$), and reading fluency ($F[1, 165] = 58.348, p < .0001$, multivariate $\eta^2 = .261$).
The Effects of the Use of an ICT-Based Reading Intervention on Students’ Achievement in Grade Two

Literacy and its component skills, the ability to read with fluency and comprehension and write fluently and coherently, are essential to educational attainment across domains: they “[bridge] the gap between learning to read and reading to learn” (Duke, Bennett-Armistead, & Roberts, 2003, p. 226) and provide the key that opens the door to a world of textually-based knowledge. The American system of education, however, has not yet achieved its potential in ensuring that as many Americans as possible enjoy the benefits of literacy. The findings of the National Assessment of Adult Literacy revealed that 43% of adults in the United States scored at basic or below basic levels in prose literacy, or the ability to understand, summarize, make simple inferences, determine cause and effect, and recognize an author’s purpose when presented with texts of moderate density (Kutner, Greenberg, Jin, Boyle, Hsu, & Dunleavy, 2007). Results of the National Assessment of Educational Progress (NAEP) painted an even bleaker picture of American youth. As of 2013, 65% fourth-grade students scored at basic or below-basic levels of grade-level literacy (National Center for Education Statistics, 2013). Research suggests that once children have reached this point in their education, when the focus of instruction has shifted from learning to read to reading to learn (Duke et al., 2003), they are at increased risk for academic failure (Felton & Pepper, 1995; Juel, 1988), often struggling to acquire the content knowledge necessary for academic success.

Reading failure poses a serious threat to a child’s future educational, professional, and social success. This conclusion is well supported in the literature. Kennely and Monrad (2007) identified a statistically significant correlation between low reading
scores and school dropout, and researchers have consistently found that youngsters with reading-related difficulties are disproportionally represented in the juvenile detention system (Rutherford, Bullis, Anderson, & Griller-Clark, 2002; Shelley-Tremblay, O’Brien, & Langhinrichsen-Rohling, 2007), placing them at increased risk for future criminal behavior and social dysfunction.

The vast majority of children at risk for illiteracy can be taught to read with fluency and comprehension, provided they receive developmentally appropriate instruction in the sound-symbol correspondences of spoken and written language (Ehri, Nunes, Willows, Schuster, Yaghoub-Zadeh, & Shanahan, 2001; Snow, Griffin, & Burns, 2005). In particular, sequential instruction in code-based skills, including explicit, systematic phonics has been shown to positively affect the reading and writing abilities of students with reading-related challenges (Ehri et al., 2001; Hatcher, Hulme, & Snowling, 2004; Torgerson, Brooks, & Hall, 2006). For these techniques to work, however, teachers must have the knowledge and skills necessary to implement them.

**Teacher Preparation and Professional Development**

Despite a large body of literature showing what works, many teacher preparation programs have failed to provide teachers with the knowledge and skills required to meet the needs of beginning readers. In a 2006 review of course syllabi and textbooks, Walsh, Glaser, and Dunne-Wilcox revealed that only 15% of American colleges of education provided pre-service teachers with even minimal exposure to the science underlying reading instruction, and a smaller percentage of reading textbooks were found to be acceptably comprehensive in their coverage of reading instruction. Correspondingly, Snow et al. (2005) found that a large proportion of special and general education teachers
were unprepared to provide evidence-based reading interventions to their students.

Because teacher preparation programs have failed to provide the majority of elementary educators with the training required to effectively meet the needs of beginning readers, efforts must be undertaken to identify and promote means to ensure that all students nevertheless have access to high quality reading instruction. Systematic reading instruction using information and communication technologies (ICT) has been enthusiastically trumpeted as a means by which to promote reading achievement (Savage et al., 2013), and such instruction often requires little or no direct intervention on the part of the classroom teacher (Bishop & Edwards Santoro, 2006). ICT-based instructional programs have been widely adopted in classroom contexts, “generally with an underlying expectation that student learning can improve … through supportive skill instruction with practice” (Cassady & Smith, 2005, p. 362). This sentiment was mirrored in the National Reading Panel’s (NPR) report of 2000, which characterized ICT-based reading instruction as a potentially promising development, allowing students greater opportunity to “interact instructionally with text” than typically offered by conventional instruction alone (Ch. 6, p. 8).

Some key potential advantages of beginning reading instruction using information and computer technologies are (a) explicit, systematic instruction in the sound-symbol correspondences of spoken and written language (Camilli et al., 2003; Ehri et al., 2001; Torgerson et al., 2006), (b) multimodal instruction to promote recall and retention (Low & Sweller, 2005; Moreno & Mayer, 2007), (c) formative feedback to guide learning and activate prior knowledge (Narciss, 2013), (d) interactivity to promote attention and engagement (Sims, 2000, 2003), and (e) opportunities for mastery learning to enhance
achievement (Guskey, 2007, 2012).

The question of whether or not the majority of ICT-based beginning reading interventions actually leverage the potential advantages of the medium remains unresolved (Edwards Santoro & Bishop, 2010; Grant et al., 2012), and ICT-based reading instruction remains poorly theorized and inadequately researched (Savage et al., 2013), particularly concerning studies involving participants aged eight years and younger (Lankshear & Knobel, 2003). This dearth of quality research is exemplified by the results of recent syntheses of ICT-based reading intervention literature. In a meta-analysis of over forty studies designed to evaluate the efficacy of ICT-based reading interventions in promoting early literacy achievement, Blok et al. (2002) found only three studies (Barker and Torgersen [1995] and two sub-studies contained in Foster, Erickson, Foster, Brinkman, and Torgesen [1994]) investigating a sequential, code-based English-language reading intervention. While the authors (Barker & Torgesen, 1995; Foster et al., 1994) documented statistically significant effects of their respective interventions favoring participants in the treatment group, the software used (iterations of DaisyQuest) is now obsolete. All of the other studies included in the meta-analysis performed by Blok et al. (2002) were either conducted in a language other than English or involved narrowly-focused computer-assisted interventions, such as the visual and auditory presentation of text, speech feedback, or virtual flashcards. Though Blok et al. (2002) reported a modest corrected overall effect size estimate of +0.19 across studies, they cautioned that this result should be interpreted with care, as many of the studies evaluated were poorly designed, and several compared the effects of multiple software applications, rather than comparing ICT-based interventions to non-computer mediated approaches.
The caution urged by Blok et al. (2002) appears well founded in the light of the results of other syntheses, whose authors applied more stringent inclusion criteria (e.g., requiring experimental designs or matched quasi-experimental designs with both pre- and post-test measures). In a 2003 synthesis of the randomized controlled trial (RCT) literature in the field, Torgerson and Zhu identified a dozen studies meeting inclusion criteria. Of those, only one (Mitchell & Fox, 2001) included a sequential, code-based ICT-based reading intervention delivered to participants in the early primary grades, and the authors of that study reported no statistically significant effect of the intervention. These results were paralleled by a synthesis of literature commissioned by the National Science Foundation and performed by Kulik in 2003. While the author (Kulik, 2003) identified nine studies of computer-based reading interventions meeting inclusion criteria, only three of those showed statistically significant positive effects favoring treatment group participants, and all three were published in the early 1990s and evaluated the effects of now obsolete software packages.

In a more recent synthesis, Slavin, Lake, Chambers, Cheung, and Davis (2009) identified nineteen studies of computer-assisted reading programs meeting selection criteria; however, only eight were peer-reviewed studies including a now extant sequential ICT-based reading intervention delivered to students in kindergarten through grade three. Of those eight, four were embedded in a larger study commissioned by the Institute of Education Sciences (IES) (Campuzano, Dynarski, Agodini, Rall, & Pendleton, 2009). Of the four interventions evaluated by Campuzano et al. (2009), none

---

29 While Mitchell and Fox (2001) found no significant differences favoring the ICT-based reading intervention group when compared to comparable teacher-delivered intervention group, they did find significant differences favoring the ICT-based reading intervention group when compared to a no-treatment control.
were found to have statistically significant positive effects on participants’ reading achievement. Unlike Campuzano et al. (2009), the authors of three of the four remaining studies reviewed by Slavin et al. (2009) reported statistically significant positive effects of their respective interventions, with effect sizes ranging from +0.17 to +1.05. While otherwise well-designed, the quasi-experimental study in which significant differences were not found favoring the treatment group in measures of word reading (Paterson, Jacobs Henry, O'Quin, Ceprano, & Blue, 2003), lacked a pre-test measure for that construct, putting its findings concerning reading achievement into question. Notably, both of the still extant software packages for which the authors provided statistically significant evidentiary support, Waterford Early Learning Program (Cassady & Smith, 2005; Tracey & Young, 2005) and Lexia Learning Systems software (Macaruso, Hook, & McCabe, 2006), involved a strong emphasis on systematic phonics instruction and the sequential presentation of code-based skills.

**Purpose**

With some notable exceptions (e.g., Macaruso, Hook, & McCabe, 2006; McMurray, 2013; Savage et al., 2013; Savage, Abrami, Hipps, & Deault, 2009; Savage, Erten, Abrami, Hipps, Comaskey, & van Lierop, 2010), relatively little high quality experimental or quasi-experimental research has been published examining the effects of ICT-based reading interventions on beginning reading achievement (Blok et al., 2002; Kulik, 2003; Slavin et al., 2009; Torgerson & Zhu, 2003). Prominent voices in the field have suggested that teachers and education authorities remain wary of adopting any ICT-based reading program until it has a consistent base of high quality evidentiary support
(Slavin et al., 2009; Torgerson, 2007; Torgerson & Zhu, 2003). Through the present study, the authors wish to fill a gap in the existing ICT-based beginning reading intervention literature, while addressing issues of research design and intervention quality that have been inadequately explored in previous intervention research.

**Research Question**

In the present study, the following research question was addressed: Are there significant mean differences in (a) non-word reading, (b) real word reading, (c) non-word spelling, (d) real word spelling, and/or (e) reading fluency post-test achievement scores of students assigned to use the MVRC online reading intervention in addition to regular classroom reading instruction and those of students from a business-as-usual comparison condition, once the effects of differences in pre-test achievement scores and relevant demographic variables have been accounted for?

**Method**

**Participants and Settings**

**Participants.** Participants included 209 students enrolled in eight, second grade classrooms in two public elementary schools in the southwestern United States. Of those, 107 were assigned to the treatment condition, and 102 were assigned to a business-as-usual comparison condition. Owing to attrition and illness, 39 participants had incomplete data sets following post-testing. Ultimately, the treatment condition comprised 89 complete cases, and the comparison condition comprised 81, representing 170 complete cases, or 81.34% of original cases. While overall data loss was below 20%,

---

30 Among the recommendations of Torgerson (2007) and Torgerson and Zhu (2003) was the conduct of randomized controlled trials (RCTs). Owing to the small sample size in the present study, it was not possible to perform random assignment to groups at the classroom level. Therefore, the authors selected to statistically control for pre-test differences between groups.
data loss did not impact each condition equally, as the treatment group retained 83.12% of original cases, while the comparison group retained 79.41% of original cases, producing a differential attrition rate of 3.71%.

**Participant demographics.** All participants attended schools in the same district and within the same postal code. At the time of testing, 81.6% of students in the district were identified as Hispanic, 12.1% as white, 4.9% as Native America, 2.7% as African American, and 0.6% as Asian or Pacific Islander. Over nine-in-ten (93.2%) students were eligible to receive free or reduced price school meals, and 6.2% were homeless. District-wide, nine percent of students were classified as English language learners (ELL), and an additional 5.3% had been reclassified as fluent English speakers.\(^{31}\) Additionally, 11.8% of students within the district had an individualized education program (IEP), indicating a documented disability.\(^{32}\)

The percentage of students in the sample who had a documented disability tracked very closely with district demographics, at 12.3%. Sample demographics, however, varied markedly from district demographics in ELL status, as 37.4% of sample group members were classified as ELLs, and an additional 8.8% were reclassified as fluent English speakers. Data concerning ethnicity and markers of socioeconomic status were not made available at the individual case level in order to protect the privacy of participants.

**Assignment to groups.** Because of the potential for disruption to classroom routines, random assignment at the student level was impossible. Assignment to groups

---

\(^{31}\) This reclassification process occurred when students achieved a satisfactory level of English-language proficiency per standardized testing.

\(^{32}\) Data were derived from district publications. A reference to the source of this data has not been included, so as to protect the identity of participating schools.
was therefore performed at the school level. One school, including each of its two participating classrooms, was assigned to the treatment condition, and the other school, including each of its two participating classrooms, was assigned to the comparison condition. The school whose third grade students (the youngest grade tested) performed more poorly (39% pass rate in reading) relative to the other (53% pass rate in reading) on the state standardized test of in the 2012-2013 school year was assigned to the treatment condition, and the school that performed better (53% pass rate in reading) relative to the other (39% pass rate in reading) on the state standardized test of reading in the 2012-2013 school year was assigned to the comparison condition. Group assignment was performed in this manner in order to ensure that any demographic advantages would favor the comparison group.

**Intervention settings.** The research was conducted over the 2013-2014 school year. During treatment, treatment group students received MVRC online reading instruction in their schools’ computer labs during a fixed period each day, Monday through Thursday. No other students were permitted in the lab during treatment, and the environment was made as reasonably free of distractions as possible. Participating students were provided with individual computers, monitors, and headphones. Prior to the beginning of treatment, MindPlay Educational Software for Reading personnel visited the school sites and confirmed the adequacy of the schools’ computer hardware and Internet connectivity.

**Materials**

**Mindplay Virtual Reading Coach (MVRC).** The MVRC (MindPlay Educational

---

33 Data were derived from district publications. A reference to the source of this data has not been included, so as to protect the identity of participating schools.
Software for Reading, 2015) intervention is a multi-component reading and language arts curriculum delivered via the Internet. It contains explicit instruction in (1) phonemic awareness, (2) phonics, (3) fluency, (4) vocabulary, and (5) comprehension in alignment with the recommendations of the National Reading Panel (NRP) (2000), as well as instruction in grammar and concepts about print. It provides sequenced instruction consistent with the models of alphabetic and phonological development proposed by Ehri (2005) and Pufpaff (2009), respectively, and satisfies the large majority of the content criteria outlined by Grant et al. (2012), as well as all of the instructional and interface design criteria identified by Bishop and Edwards Santoro (2006).

Measures

**Dependent variable measures.** Dependent variable measures included the Test of Silent Word Reading Fluency, Second Edition (TOSWRF-2; Mather, Hammill, Allen & Roberts, 2014) and tests from the Woodcock Johnson Tests of Achievement, Fourth Edition (WJ IVACH; Schrank, Mather, & McGrew, 2014). These measures were chosen because they are considered highly valid measures of their respective constructs, and they have high levels of test-retest reliability, reducing the likelihood of regression to the mean. Table 23 provides an overview of the description of each dependent variable measure performed.

---

34 Copies of these instruments were provided to the researchers by the publishers prior to publication and public release.
### Table 23

**Tests of Reading Achievement**

<table>
<thead>
<tr>
<th>Instrument</th>
<th>DV</th>
<th>Description of Measure</th>
<th>Administration</th>
</tr>
</thead>
<tbody>
<tr>
<td>Letter-Word Identification</td>
<td>Real Word Reading</td>
<td>Reading aloud letters and real words from a list</td>
<td>Individual</td>
</tr>
<tr>
<td>Word Attack</td>
<td>Non-Word Reading</td>
<td>Reading aloud nonsense words from a list</td>
<td>Individual</td>
</tr>
<tr>
<td>Spelling</td>
<td>Real Word Spelling</td>
<td>Spelling of orally administered real words</td>
<td>Group</td>
</tr>
<tr>
<td>Spelling of Sounds</td>
<td>Non-Word Spelling</td>
<td>Spelling of orally administered nonsense words</td>
<td>Group</td>
</tr>
<tr>
<td>Test of Silent Word Reading Fluency, Second Edition (TOSWRF-2; Mather et al., 2014)</td>
<td>Reading Fluency</td>
<td>Rapidly drawing lines to indicate the boundaries between connected real words</td>
<td>Group</td>
</tr>
</tbody>
</table>

*Note.* DV = dependent variable. The Letter-Word Identification, Word Attack, Spelling, and Spelling of Sounds instruments are tests of the fourth edition of the Woodcock-Johnson Tests of Achievement (WJ IV ACH; Schrank et al., 2014).

**Rationale for dependent variable measure selection.** The dependent variable measures selected for this intervention were chosen both because of their high reliabilities and because they were designed to measure code-based skills foundational to proficient reading. Tests of real word and non-word reading and spelling were selected to ensure that the study included measures of the identification and production of both phonetically regular and phonetically irregular words. Identification and production of the former are indicative of graphophonemic knowledge, while identification and production of the latter are indicative of sight word knowledge and familiarity with the word patterns of English. A test of reading fluency was selected both as a broad measure of decoding ability and because reading rate and accuracy are essential to reading comprehension, and
there is a strong correlation between reading fluency and comprehension (Pikulski & Chard, 2005), which is the ultimate goal of reading.

**Fidelity measures.** To measure fidelity of implementation, both product usage logs and behavioral observation measures were used. Two measures of fidelity of implementation were selected to evaluate not only whether the participants were using the intervention for the duration specified, but also whether they were actively engaged while using the intervention. The active engagement of participants, it was reasoned, would be a more nuanced measure of the acceptability of the intervention among participants than would simple duration of use, thus serving as a metric for an important facet of social validity.

**Product usage logs.** The frequency and duration of individual student's use of the MVRC online reading intervention were measured using software usage logs generated by the MVRC product. These logs included data concerning the duration of use of product components in minutes, the type and number of activities successfully completed, and individual students’ progress toward achievement targets. Only active use of the MVRC product was reported in the software usage logs, as the MVRC online intervention automatically logs users out of the program after three minutes of inactivity.

**Behavioral observations.** A 10-data point planned activity check, using partial-interval recording method, was employed to measure student engagement behavior biweekly throughout the study. This instrument was developed by a doctoral candidate in Special Education at the University of Arizona. Student engagement behavior was operationalized as physical position (head up, eyes facing screen), headphone use, keyboard use, and screen activity. To administer this measure, two observers
simultaneously scanned each classroom at two-minute intervals for 20 minutes and
documented the number of participants who appeared to be engaged, as well as the
number of participants who did not appear to be engaged, per the criteria previously
outlined. Data collected with this instrument were used to document the percentage of
treatment group participants who displayed engagement behavior at each observed
interval. Observers varied their observation schedules each week to reduce the likelihood
of participants' modifying their behavior in anticipation of observation.

**Intervention Procedure**

**Routine classroom instruction.** All participants across both groups received
Success for All language arts instruction in the classroom setting. Success for All is a
comprehensive, code-based, standards-aligned reading curriculum that includes
structured lessons, cooperative group activities, and regular assessments (Slavin &
Madden, 2012). Success for All classroom activities included intensive daily reading
blocks of 90 minutes duration, as well as systematic phonics instruction, guided practice,
and ongoing formative evaluation. Activities are scripted and highly structured, so there
was little variability in instruction among participating classrooms.

**Intervention instruction.** The research was conducted over the 2013-2014 school
year. Participating students assigned to the treatment group received MVRC online
reading instruction in addition to Success for All reading and language arts instruction.
Participants assigned to the treatment group used the software for 30 minutes each day,
Monday through Thursday, for a total of two hours per week throughout the regular
school year (mid-September through mid-April), with the exception of holidays, school
functions, and mandatory state testing days. Classroom teachers brought students to the
computer lab at the assigned time and assisted with student log in when required; however, they did not assist students in any other way.

**Comparison instruction.** Participating students assigned to the business-as-usual comparison condition received Success for All reading and language arts instruction but did not receive the MVRC intervention. In addition to regular language arts instruction, participants in the comparison group also received two hours of supplementary reading instruction weekly from their classroom teachers. This instruction employed materials and instructional techniques consistent with the Success for All curriculum and routine classroom practice.

**Teacher training.** Teachers of the treatment group classrooms underwent training prior to the implementation of the intervention. Training included a one-hour webinar and a four-hour in-person seminar delivered by MindPlay personnel. The MVRC online reading intervention requires very little participation on the part of classroom teachers, as the content and difficulty of the intervention are adaptive in response to individual student performance. Ultimate achievement targets are aligned with Common Core standards; however, the content and instruction delivered to individual students are determined solely by embedded diagnostic assessments (i.e., they cannot be modified by teachers). Therefore, teacher training focused on the underlying pedagogical and theoretical structure of the program, the structure and content of the embedded lessons and activities, facilitating and supporting student access and use of the intervention, and accessing and understanding the data contained in student and classroom reports generated by the software. Teachers were also provided with a teacher’s guide to accompany the program.
Assessment Procedure

**Procedures to ensure accuracy of administration.** Examiners included a graduate student in special education, who had over five years of experience working with children in schools and two years of experience administering standardized achievement tests, as well as several retired special educators, who had extensive experience working with children in educational settings and administering standardized achievement tests. Prior to the administration of reading achievement dependent variable measures, examiners met to review administration. In addition, all examiners received supplemental training in the administration of the WJ IV ACH from one of the tests’ co-authors.

**Procedures to ensure accuracy of scoring.** Following data collection, test protocols were re-scored by a second examiner to ensure accuracy of scoring. While formal data concerning scoring disagreements and agreements were not collected or retained, examiners reported very few disagreements when rescoring protocols.

**Administration of dependent variable measures.** Pre-test dependent variable data collection was performed over a one-week period during the first 30 days of the academic year, and post-test dependent variable data collection was performed over a one-week period during the last 30 days of the academic year. Prior to dependent variable data collection, examiners endeavored to establish rapport with participating students. Individual data collection was performed in quiet, low-distraction rooms in the participants’ school; group data collection was performed in the participants’ classrooms. Data collection was staggered over several days to avoid fatigue-related effects, and measures were counter-balanced to avoid order effects.
The *Spelling and Spelling of Sounds* tests of the WJ IV ACH (Schrank et al., 2014), as well as the TOSWRF-2 (Mather et al., 2014), were administered to participating students in whole-class groups, whereas the *Word Attack* and *Letter-Word Identification* tests of the WJ IV ACH were administered individually and in a manner consistent with standardized administration protocols. Group-administered tests were administered beginning with the first item in each assessment and ending with items likely to be well beyond the performance of any participant. Consequently, basal (start point) requirements were satisfied for all participants tested, and no participant failed to reach a ceiling (end point) in their performance on any assessment.

**Data Analysis**

**Measurement of data.** All dependent variable and pretest covariate data collected for this study were measured as obtained, raw scores using discrete, quantitative scales. Obtained, raw scores for each post-test achievement measure comprised the study’s dependent variables. A pre-test covariate was created by taking the mean of sample-referenced $z$-scores across pre-tests.\(^{35}\) The conversion of scores was undertaken prior to averaging to ensure that all data reflected equivalent scales of measurement. Such a procedure was necessary because the number of items varied across tests, as did item difficulty, underlying constructs, and scoring procedures.

**Missing data.** Owing to attrition and illness, 39 participants had incomplete data sets following post-testing. Ultimately, the treatment condition comprised 89 complete cases, and the comparison condition comprised 81, representing 170 complete cases, or

\(^{35}\) The authors chose to create a single pre-test achievement covariate, rather than add the pre-test achievement scores for each dependent variable a separate covariate to the MANCOVA model. This decision was taken to reduce the number of variables added to the MANCOVA model, thereby increasing statistical power.
81.34% of original cases. Because more than 10% of the original cases were missing data, the authors chose to exclude cases listwise from analyses. Imputation of missing data was considered and rejected consistent with the recommendations of Tabatchnick and Fidell (2013), who suggested imputation only when fewer than 10% of cases were missing data and data were missing at random.

**Identification of covariates.** To determine whether significant differences existed between groups on pre-test achievement scores and/or demographic variables, univariate analyses of variance (ANOVA) were performed to compare mean scores between groups (i.e., treatment and comparison) for each dependent variable and demographic variable. The authors elected to determine significance at the p < .10 level to ensure that any differences between groups that might impact dependent variable measures would be adequately accounted for in subsequent analyses. The ANOVA analyses led to the identification of several covariates. Significant differences favoring the treatment group were detected at the p < .10 level in the pre-test scores for real word reading \(F(1, 169) = 2.927, p = .089\), and non-word reading \(F(1, 169) = 3.314, p = .070\). Significant differences were also detected at the p < .10 level English language learner (ELL) status \(F(1,169) = 11.480, p = .001\) and special education (SPED) status \(F(1,169) = 3.421, p = .066\). Therefore, a decision was taken to add pre-test achievement, ELL status, and SPED status to the MANCOVA model as covariates.

**Test Procedure**

A multivariate analysis of covariance (MANCOVA) was performed to detect mean differences in (a) non-word reading, (b) real word reading, (c) non-word spelling, (d) real word spelling, and/or (e) reading fluency post-test achievement scores of students
assigned to use the MVRC online reading intervention and those of students assigned to a business-as-usual comparison condition (regular classroom reading instruction alone), once the effect of differences in pre-test achievement scores and relevant demographic variables had been accounted for.\textsuperscript{36}

**Tests of equality of variance and covariance.** Levene’s test for equality of error variance was non-significant for each of the study’s dependent variables at the \( p < .05 \) level, as was Box’s test of equality of covariance matrices at the \( p < .01 \) level, indicating that the data did not violate the assumption of homogeneity of variance. Therefore, consistent with the recommendations of Tabachnick and Fidell (2013), the authors chose to use Wilk’s lambda as the test statistic in subsequent multivariate analyses.

**Multivariate tests.** The results of the MANCOVA analysis revealed a significant main effect (\( \lambda = .668, F[5, 161] = 16.014, p < .001, \text{multivariate } \eta^2 = .332 \)) of the intervention on the achievement scores of participants assigned to the treatment condition relative to those assigned to the business-as-usual comparison condition, once the effects of differences in pre-test achievement scores, special education (SPED) status, and English language learner (ELL) status had been accounted for. The main effects of ELL status (\( \lambda = .868, F[5, 161] = 4.903, p < .001, \text{multivariate } \eta^2 = .132 \)) and pre-test scores (\( \lambda = .212, F[5, 161] = 119.487, p < .001, \text{multivariate } \eta^2 = .788 \)) indicated significant effects on the combined dependent variables (DV). Special education status, however, had no significant effect on the combined DVs. Table 2 contains the results of the multivariate tests for the study’s independent variable and each of its covariates.

\textsuperscript{36} Data analyses were performed using IBM SPSS GENERAL LINEAR MODEL unless otherwise noted.
Table 24

Multivariate Tests

<table>
<thead>
<tr>
<th>IV/ Covariate</th>
<th>Test</th>
<th>Value</th>
<th>F</th>
<th>Hypothesis df</th>
<th>Error df</th>
<th>Sig.</th>
<th>Partial η²</th>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>SPED Status</td>
<td>Pillai's race</td>
<td>.056</td>
<td>1.920</td>
<td>5</td>
<td>161</td>
<td>.094</td>
<td>.056</td>
<td>.639</td>
</tr>
<tr>
<td></td>
<td>Wilks' Lambda</td>
<td>.944</td>
<td>1.920</td>
<td>5</td>
<td>161</td>
<td>.094</td>
<td>.056</td>
<td>.639</td>
</tr>
<tr>
<td>ELL Status</td>
<td>Pillai's Trace</td>
<td>.132**</td>
<td>4.903</td>
<td>5</td>
<td>161</td>
<td>&lt;.001</td>
<td>.132</td>
<td>.979</td>
</tr>
<tr>
<td></td>
<td>Wilks' Lambda</td>
<td>.868**</td>
<td>4.903</td>
<td>5</td>
<td>161</td>
<td>&lt;.001</td>
<td>.132</td>
<td>.979</td>
</tr>
<tr>
<td>Pre-Test Scores</td>
<td>Pillai's Trace</td>
<td>.788**</td>
<td>119.478</td>
<td>5</td>
<td>161</td>
<td>&lt;.001</td>
<td>.788</td>
<td>1.000</td>
</tr>
<tr>
<td></td>
<td>Wilks' Lambda</td>
<td>.212**</td>
<td>119.478</td>
<td>5</td>
<td>161</td>
<td>&lt;.001</td>
<td>.788</td>
<td>1.000</td>
</tr>
<tr>
<td>Condition</td>
<td>Pillai's Trace</td>
<td>.332**</td>
<td>16.014</td>
<td>5</td>
<td>161</td>
<td>&lt;.001</td>
<td>.332</td>
<td>1.000</td>
</tr>
<tr>
<td></td>
<td>Wilks' Lambda</td>
<td>.668**</td>
<td>16.014</td>
<td>5</td>
<td>161</td>
<td>&lt;.001</td>
<td>.332</td>
<td>1.000</td>
</tr>
</tbody>
</table>

Note. df = degrees of freedom; F = Fisher's F ratio; IV = independent variable; Sig. = significance.
** Significant at the p < .01 level; * significant at the p < .05 level.

Univariate tests. Post-hoc ANCOVAs were performed to isolate the relationship between each dependent variable, the combined pre-test covariate, and the grouping variable (condition). The multivariate test results were confirmed by the post-hoc ANCOVAs, performed using Bonferroni’s correction for multiple comparisons, across four of the study’s five dependent variables: real word spelling ($F[1, 165] = 16.341, p < .001, \text{ multivariate } \eta^2 = .090$), non-word reading ($F[1, 165] = 4.368, p = .038, \text{ multivariate } \eta^2 = .026$), non-word spelling ($F[1, 165] = 29.212, p = .001, \text{ multivariate } \eta^2 = .150$), and reading fluency ($F[1, 165] = 58.348, p < .001, \text{ multivariate } \eta^2 = .261$). Statistically significant effects of the intervention, however, were not detected in real word reading ($F[1, 165] = 2.328, p = .129, \text{ multivariate } \eta^2 = .014$). While absolute differences were detected favoring the treatment group, the findings lacked power, and the probability of error was unacceptably high. Table 25 displays tests of between-subjects effects, while
Table 26 displays pairwise comparisons between groups for each dependent variable.

Table 25

<table>
<thead>
<tr>
<th>IV/Covariate</th>
<th>DV</th>
<th>df</th>
<th>M^2</th>
<th>F</th>
<th>Sig.</th>
<th>Partial η^2</th>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>SPED Status</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Real Word Reading</td>
<td>1</td>
<td>109.027*</td>
<td>3.899</td>
<td>.050</td>
<td>.023</td>
<td>.501</td>
<td></td>
</tr>
<tr>
<td>Real Word Spelling</td>
<td>1</td>
<td>.570</td>
<td>.073</td>
<td>.787</td>
<td>&lt;.001</td>
<td>.058</td>
<td></td>
</tr>
<tr>
<td>Non-Word Reading</td>
<td>1</td>
<td>.158</td>
<td>.013</td>
<td>.909</td>
<td>&lt;.001</td>
<td>.051</td>
<td></td>
</tr>
<tr>
<td>Non-Word Spelling</td>
<td>1</td>
<td>13.429</td>
<td>2.481</td>
<td>.117</td>
<td>.015</td>
<td>.347</td>
<td></td>
</tr>
<tr>
<td>Reading Fluency</td>
<td>1</td>
<td>6.461</td>
<td>.045</td>
<td>.831</td>
<td>&lt;.001</td>
<td>.055</td>
<td></td>
</tr>
<tr>
<td><strong>ELL Status</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Real Word Reading</td>
<td>1</td>
<td>55.407</td>
<td>1.982</td>
<td>.161</td>
<td>.012</td>
<td>.288</td>
<td></td>
</tr>
<tr>
<td>Real Word Spelling</td>
<td>1</td>
<td>.018</td>
<td>.002</td>
<td>.962</td>
<td>&lt;.001</td>
<td>.050</td>
<td></td>
</tr>
<tr>
<td>Non-Word Reading</td>
<td>1</td>
<td>75.392*</td>
<td>6.220</td>
<td>.014</td>
<td>.036</td>
<td>.698</td>
<td></td>
</tr>
<tr>
<td>Non-Word Spelling</td>
<td>1</td>
<td>1.944</td>
<td>.359</td>
<td>.550</td>
<td>.002</td>
<td>.092</td>
<td></td>
</tr>
<tr>
<td>Reading Fluency</td>
<td>1</td>
<td>171.255</td>
<td>1.205</td>
<td>.274</td>
<td>.007</td>
<td>.194</td>
<td></td>
</tr>
<tr>
<td><strong>Pre-Test Scores</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Real Word Reading</td>
<td>1</td>
<td>8497.432**</td>
<td>303.921</td>
<td>&lt;.001</td>
<td>.648</td>
<td>1.000</td>
<td></td>
</tr>
<tr>
<td>Real Word Spelling</td>
<td>1</td>
<td>3627.206**</td>
<td>466.969</td>
<td>&lt;.001</td>
<td>.739</td>
<td>1.000</td>
<td></td>
</tr>
<tr>
<td>Non-Word Reading</td>
<td>1</td>
<td>1375.069**</td>
<td>113.452</td>
<td>&lt;.001</td>
<td>.407</td>
<td>1.000</td>
<td></td>
</tr>
<tr>
<td>Non-Word Spelling</td>
<td>1</td>
<td>529.461**</td>
<td>97.817</td>
<td>&lt;.001</td>
<td>.372</td>
<td>1.000</td>
<td></td>
</tr>
<tr>
<td>Reading Fluency</td>
<td>1</td>
<td>36566.981**</td>
<td>257.198</td>
<td>&lt;.001</td>
<td>.609</td>
<td>1.000</td>
<td></td>
</tr>
<tr>
<td><strong>Condition</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Real Word Reading</td>
<td>1</td>
<td>65.082</td>
<td>2.328</td>
<td>.129</td>
<td>.014</td>
<td>.329</td>
<td></td>
</tr>
<tr>
<td>Real Word Spelling</td>
<td>1</td>
<td>126.926**</td>
<td>16.341</td>
<td>&lt;.001</td>
<td>.090</td>
<td>.980</td>
<td></td>
</tr>
<tr>
<td>Non-Word Reading</td>
<td>1</td>
<td>52.940*</td>
<td>4.368</td>
<td>.038</td>
<td>.026</td>
<td>.547</td>
<td></td>
</tr>
<tr>
<td>Non-Word Spelling</td>
<td>1</td>
<td>158.118**</td>
<td>29.212</td>
<td>&lt;.001</td>
<td>.150</td>
<td>1.000</td>
<td></td>
</tr>
<tr>
<td>Reading Fluency</td>
<td>1</td>
<td>8295.640**</td>
<td>58.348</td>
<td>&lt;.001</td>
<td>.261</td>
<td>1.000</td>
<td></td>
</tr>
</tbody>
</table>

Note. df = degrees of freedom; DV = dependent variable; F = Fisher's F ratio; IV = independent variable; M^2 = mean squared; Sig. = significance.

** Significant at the p < .01 level; * significant at the p < .05 level.
Table 26

*Pairwise Comparisons for Univariate Tests*

<table>
<thead>
<tr>
<th>Dependent Variable</th>
<th>Condition</th>
<th>Mean</th>
<th>Mean Difference</th>
<th>Standard Error</th>
<th>Sig.</th>
<th>Lower Bound</th>
<th>Upper Bound</th>
</tr>
</thead>
<tbody>
<tr>
<td>Real Word Reading</td>
<td>Comparison</td>
<td>42.901</td>
<td>-1.297</td>
<td>.850</td>
<td>.129</td>
<td>-2.976</td>
<td>3.82</td>
</tr>
<tr>
<td></td>
<td>Treatment</td>
<td>45.944</td>
<td>1.297</td>
<td>.850</td>
<td>.129</td>
<td>-3.82</td>
<td>2.976</td>
</tr>
<tr>
<td>Real Word Spelling</td>
<td>Comparison</td>
<td>20.383</td>
<td>-1.811**</td>
<td>.448</td>
<td>&lt;.001</td>
<td>-2.976</td>
<td>-.927</td>
</tr>
<tr>
<td></td>
<td>Treatment</td>
<td>23.326</td>
<td>1.811**</td>
<td>.448</td>
<td>&lt;.001</td>
<td>.927</td>
<td>2.696</td>
</tr>
<tr>
<td>Non-Word Reading</td>
<td>Comparison</td>
<td>18.136</td>
<td>-1.170*</td>
<td>.560</td>
<td>.038</td>
<td>-2.275</td>
<td>-.065</td>
</tr>
<tr>
<td></td>
<td>Treatment</td>
<td>19.629</td>
<td>1.170*</td>
<td>.560</td>
<td>.038</td>
<td>.065</td>
<td>2.275</td>
</tr>
<tr>
<td>Non-Word Spelling</td>
<td>Comparison</td>
<td>14.025</td>
<td>-2.022**</td>
<td>.374</td>
<td>&lt;.001</td>
<td>-2.760</td>
<td>-1.283</td>
</tr>
<tr>
<td></td>
<td>Treatment</td>
<td>16.438</td>
<td>2.022**</td>
<td>.374</td>
<td>&lt;.001</td>
<td>1.283</td>
<td>2.760</td>
</tr>
<tr>
<td>Reading Fluency</td>
<td>Comparison</td>
<td>45.593</td>
<td>-14.645**</td>
<td>1.917</td>
<td>&lt;.001</td>
<td>-18.430</td>
<td>-10.859</td>
</tr>
<tr>
<td></td>
<td>Treatment</td>
<td>63.169</td>
<td>14.645**</td>
<td>1.917</td>
<td>&lt;.001</td>
<td>10.859</td>
<td>18.430</td>
</tr>
</tbody>
</table>

Note. Sig. = significance.
** Significant at the p < .01 level; * significant at the p < .05 level.

**Additional post-hoc tests.** Additional post-hoc ANCOVAs were performed to isolate the relationship between each dependent variable, its corresponding pre-test covariate, and the grouping variable (condition). By controlling for pre-test achievement individually for each dependent variable, rather than using the mean pre-test score covariate employed in the full MANCOVA, the authors sought to lend more precision to the analyses and ensure that the specific effects of pre-test achievement on each dependent variable were adequately accounted for.

The post-hoc ANCOVAs confirmed the results of the full MANCOVA and post-hoc univariate tests with only one exception: the results of an ANCOVA examining differences between groups on non-word reading were inconsistent with previous analyses. The ANCOVA failed to demonstrate significant differences between groups at the p < .01 level (reduced from p < .05 using Bonferroni’s procedure to account for
multiple comparisons) in non-word reading, once the effects of differences in pre-test achievement scores, special education status, and English language learner (ELL) status had been accounted for. Because the ANCOVA analysis more precisely isolated the relationship between the real word reading variable, its corresponding pre-test covariate, and the grouping variable (condition) than did the univariate tests performed using the combined pre-test achievement covariate, the authors chose to exclude non-word reading from their report of significant findings. Table 27 and Table 28 provide *Pairwise Comparisons for Post-Hoc ANCOVAs* and *Between-Subjects Effects for Post-Hoc ANCOVAs*, respectively.

Table 27

*Pairwise Comparisons for Post-Hoc ANCOVAs*

<table>
<thead>
<tr>
<th>Dependent Variable</th>
<th>Condition</th>
<th>Mean Difference</th>
<th>Std. Error</th>
<th>Sig.</th>
<th>Lower Bound</th>
<th>Upper Bound</th>
</tr>
</thead>
<tbody>
<tr>
<td>Real Word Reading</td>
<td>Comparison</td>
<td>-.998</td>
<td>.799</td>
<td>.214</td>
<td>-2.575</td>
<td>.580</td>
</tr>
<tr>
<td></td>
<td>Treatment</td>
<td>.998</td>
<td>.799</td>
<td>.214</td>
<td>-.580</td>
<td>2.575</td>
</tr>
<tr>
<td>Real Word Spelling</td>
<td>Comparison</td>
<td>-2.473**</td>
<td>.573</td>
<td>&lt;.001</td>
<td>-3.604</td>
<td>-1.341</td>
</tr>
<tr>
<td></td>
<td>Treatment</td>
<td>2.473**</td>
<td>.573</td>
<td>&lt;.001</td>
<td>1.341</td>
<td>3.604</td>
</tr>
<tr>
<td>Non-Word Reading</td>
<td>Comparison</td>
<td>-.989</td>
<td>.585</td>
<td>.093</td>
<td>-2.145</td>
<td>.166</td>
</tr>
<tr>
<td></td>
<td>Treatment</td>
<td>.989</td>
<td>.585</td>
<td>.093</td>
<td>-.166</td>
<td>2.145</td>
</tr>
<tr>
<td>Non-Word Spelling</td>
<td>Comparison</td>
<td>-2.296**</td>
<td>.414</td>
<td>&lt;.001</td>
<td>-3.112</td>
<td>-1.479</td>
</tr>
<tr>
<td></td>
<td>Treatment</td>
<td>2.296**</td>
<td>.414</td>
<td>&lt;.001</td>
<td>1.479</td>
<td>3.112</td>
</tr>
<tr>
<td>Reading Fluency</td>
<td>Comparison</td>
<td>-2.296**</td>
<td>.414</td>
<td>&lt;.001</td>
<td>-3.112</td>
<td>-1.479</td>
</tr>
<tr>
<td></td>
<td>Treatment</td>
<td>2.296**</td>
<td>.414</td>
<td>&lt;.001</td>
<td>1.479</td>
<td>3.112</td>
</tr>
</tbody>
</table>

Note. *df* = degrees of freedom; *M^2* = mean squared; Sig. = significance; Std. Error = standard error.

** Significant at the p < .01 level.
Table 28

*Between-Subjects Effects for Post-Hoc ANCOVAs*

<table>
<thead>
<tr>
<th>Dependent Variable</th>
<th>df</th>
<th>$M^2$</th>
<th>$F$</th>
<th>Sig.</th>
<th>Partial $\eta^2$</th>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>Real Word Reading</td>
<td>1</td>
<td>38.506</td>
<td>1.559</td>
<td>.214</td>
<td>.009</td>
<td>.237</td>
</tr>
<tr>
<td>Real Word Spelling</td>
<td>1</td>
<td>240.537**</td>
<td>18.617</td>
<td>&lt;.001</td>
<td>.101</td>
<td>.990</td>
</tr>
<tr>
<td>Non-Word Reading</td>
<td>1</td>
<td>37.548</td>
<td>2.856</td>
<td>.093</td>
<td>.017</td>
<td>.390</td>
</tr>
<tr>
<td>Non-Word Spelling</td>
<td>1</td>
<td>206.257**</td>
<td>30.777</td>
<td>&lt;.001</td>
<td>.157</td>
<td>1.000</td>
</tr>
<tr>
<td>Reading Fluency</td>
<td>1</td>
<td>9691.512**</td>
<td>71.067</td>
<td>&lt;.001</td>
<td>.300</td>
<td>1.000</td>
</tr>
</tbody>
</table>

Note. df = degrees of freedom; $F$ = Fisher's $F$ ratio; $M^2$ = mean squared; Sig. = significance.
** Significant at the p < .01 level; * significant at the p < .05 level.

**Fidelity of Implementation**

**Product usage logs.** Product usage logs generated by the MVRC software indicated that active participant use of the intervention exceeded 90% of the time assigned (a mean of 44 hours of active use) among treatment group participants who remained in the study until post-test.

**Planned activity check.** Over the course of the intervention, treatment group participants were observed 45 times while using the MVRC online intervention, or approximately twice per week. Observers used a novel planned activity check instrument to record behavioral observations. Data from this instrument were analyzed by dividing the number of students displaying engagement behavior at each interval by the total number of students observed and multiplying the dividend by 100 to determine the percentage of students displaying engagement behavior. This information was averaged for the entire twenty-minute observation period and then aggregated and averaged by
month. Inter-observer agreement was determined by dividing the number of agreements between observers per interval by the total number of agreements plus disagreements and multiplying the dividend by 10.

The results of the behavioral observations indicated that treatment group participants displayed engagement behavior over 90% of the intervals observed while using the MVRC online intervention. Inter-observer agreement, however, was within the acceptable range at just under 84%. Table 29 displays the results of the planned activity check.

Table 29

<table>
<thead>
<tr>
<th></th>
<th>Percentage of Engagement Behavior</th>
<th>Inter-Observable Agreement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oct-13</td>
<td>91.50%</td>
<td>84.50%</td>
</tr>
<tr>
<td>Nov-13</td>
<td>91%</td>
<td>77%</td>
</tr>
<tr>
<td>Dec-13</td>
<td>9.50%</td>
<td>86%</td>
</tr>
<tr>
<td>Jan-14</td>
<td>93%</td>
<td>81.50%</td>
</tr>
<tr>
<td>Feb-14</td>
<td>91%</td>
<td>85.50%</td>
</tr>
<tr>
<td>Mar-14</td>
<td>92%</td>
<td>88%</td>
</tr>
<tr>
<td>Mean</td>
<td>91.50%</td>
<td>83.67%</td>
</tr>
</tbody>
</table>

**Discussion**

The results of this analysis suggest a very robust effect of the MVRC intervention on participants’ reading fluency and spelling achievement gains. The multivariate effect size for the intervention overall was in the large range (multivariate $\eta^2 = .332$), as was the effect size for reading fluency (multivariate $\eta^2 = .261$) and non-word spelling (multivariate $\eta^2 = .150$), while real word spelling (multivariate $\eta^2 = .090$) was in the
While significant effects of the intervention were not detected in isolated word reading tasks (i.e., non-word reading and real word reading), non-significant differences favoring students in the treatment group were detected. It is possible participants in both groups had reached saturation in this aspect of decoding, as such tasks are heavily emphasized in the Success for All curriculum, and participants in both groups made statistically significant gains in word reading measures from pre-test to post test. It should also be noted that unlike the Success for All curriculum, the MVRC intervention includes formal instruction in orthographic patterns and word origins, features that are likely responsible for the large gains in real word and non-word spelling among treatment group participants. Furthermore, the MVRC intervention places a strong emphasis on the development of reading fluency, with sentence reading fluency, pause-assisted fluency, passage reading fluency, and eye tracking activities of gradually increasing difficulty. The importance of fluency building activities to the MVRC intervention may explain the particularly large gains in reading fluency favoring treatment group participants. This finding is particularly significant, as fluency is a broader measure of reading ability than simple decoding, and reading rate and accuracy (i.e., fluency) are essential to reading comprehension (Pikulski & Chard, 2005).

The large size of effects attributable to the MVRC intervention is noteworthy in a field where most intervention research reports modest, if any, effects of the intervention. Notably, the MVRC web-based reading intervention contains systematic, developmentally appropriate instruction in phonics and other code-based skills, for which there is strong evidentiary support in other research contexts. These features, similar to

37 Rules of thumb for determining the magnitude of effects were derived from Rovai, Baker, and Ponton (2014).
those of the promising ABRACADABRA intervention (Savage et al., 2009; Savage et al., 2010; Savage et al., 2013), are likely responsible for its success.

**Limitations**

The authors of the present study wish to acknowledge several important limitations. Chief among these is the lack of equivalence across demographic factors between the treatment and comparison groups. While the authors employed statistical controls to correct for the differences, further research with random assignment to groups or a non-post-hoc matched design should be performed. Also absent was observational or documentary data related to teachers’ instructional practices. Because the number of participating classrooms was modest, this information would have been helpful in detecting potential confounds. Finally, outcome measures included only narrowly focused tasks of word reading, word spelling, and reading fluency. Each of these tasks measured aspects of basic reading and spelling skills, but none assessed the synthesis of these skills in either reading comprehension or sentence or passage writing. Further research should include broader measures, in particular measures of reading comprehension. While decoding ability is essential to fluent reading, the ultimate goal of reading is the comprehension of connected text.

**Conclusions and Implications for Future Research**

The prevalence of illiteracy in the United States is indicative of systemic failure in primary reading instruction. Despite the implementation of federal programs such as No Child Left Behind and Race to the Top, fourth grade student reading achievement on the NAEP improved only negligibly between 2003 (Grigg, Daane, Jin, & Campbell) and 2013 (National Center for Education Statistics). Because literacy is critical to informed
participation in the institutions and processes vital to democratic governance, systemic change is clearly warranted. Such change, however, will require adequate time, resources, and political will. It is therefore necessary to consider interim strategies and interventions for the promotion of strong reading skills.

Computer adaptive online reading instruction, like that provided by the MVRC online reading intervention, holds great promise, in that it is able to tailor instruction and formative feedback to individual performance in a manner that would be difficult for a classroom teacher providing large group instruction. Though computer programs lack the insight of the experienced teacher, error analyses and responsive content presentation of the type performed by MVRC allow for increasingly sophisticated and nuanced interactions between students and instructional software. Computer software, no matter how sophisticated, is unlikely ever to replace qualified teachers, but the results of the present study suggest that it may provide a beneficial adjunct to traditional classroom instruction.

The present study points to MVRC as a potentially promising intervention; however, the research base in ICT-based early reading interventions remains paper-thin. High-quality intervention studies, including replication research, should be performed to further evaluate MVRC, as well as other ICT-based reading interventions commonly used in American public schools. This research should be geared not only toward product evaluation, but also to the identification of optimal instructional practices. Particular attention should be paid to the duration and intensity of instruction, level of instructional integration, and level of support provided to teachers and students. Furthermore, researchers should endeavor to identify factors affecting differential levels of
responsiveness among students and differential levels of instructional integration and treatment fidelity among teachers. Finally, research should be performed in accordance with an established framework for high quality educational research and reporting.
APPENDIX A: ANCHORED MATRIX FOR EVALUATING EXPERIMENTAL RESEARCH (AMEER)

**Rubric for Essential Quality Indicators**

<table>
<thead>
<tr>
<th>Criterion/Score</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Quality Indicators for Describing Participants</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. Adequate information was provided to determine/confirm whether the participants demonstrated the disabilities or difficulties addressed.</td>
<td>No information was provided to determine/confirm whether the participants demonstrated the disabilities or difficulties addressed.</td>
<td>Inadequate information was provided to determine/confirm whether the participants demonstrated the disabilities or difficulties addressed.</td>
<td>Adequate information was provided to determine/confirm whether the participants demonstrated the disabilities or difficulties addressed.</td>
<td>Ample information was provided to determine/confirm whether the participants demonstrated the disabilities or difficulties addressed.</td>
</tr>
<tr>
<td>2. Appropriate procedures were used to increase the probability that participants were comparable across conditions.</td>
<td>No procedures were used to increase the probability that participants were comparable across conditions.</td>
<td>Inadequate procedures were used to increase the probability that participants were comparable across conditions.</td>
<td>Adequate procedures were used to increase the probability that participants were comparable across conditions.</td>
<td>Robust procedures were used to increase the probability that participants were comparable across conditions.</td>
</tr>
<tr>
<td>3. Adequate information describing important characteristics of the intervention providers was included, and appropriate procedures to increase the probability that intervention providers were comparable across conditions were used.</td>
<td>No information describing important characteristics of intervention providers was included, or procedures to increase the probability that intervention providers were comparable across conditions were not documented.</td>
<td>Inadequate information describing important characteristics of intervention providers was included, or inadequate procedures were used to increase the probability that intervention providers were comparable across conditions.</td>
<td>Adequate information describing important characteristics of intervention providers was included, and adequate procedures were used to increase the probability that intervention providers were comparable across conditions.</td>
<td>Ample information describing important characteristics of the intervention providers was included, and robust procedures were used to increase the probability that intervention providers were comparable across conditions.</td>
</tr>
<tr>
<td>Criterion/Score</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>----------------</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>Quality Indicators for Implementation of the Intervention and Description of Comparison Conditions</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. The intervention was clearly described and specified.</td>
<td>The intervention was not described or specified.</td>
<td>The intervention was inadequately described or specified.</td>
<td>The intervention was adequately described and specified.</td>
<td>The intervention was clearly and coherently described and specified.</td>
</tr>
<tr>
<td>2. Fidelity of implementation was described and assessed.</td>
<td>Fidelity of implementation was not described or assessed.</td>
<td>Fidelity of implementation was inadequately described or assessed.</td>
<td>Fidelity of implementation was adequately described and assessed.</td>
<td>Fidelity of implementation was clearly and coherently described and assessed.</td>
</tr>
<tr>
<td>3. The nature of services provided in comparison conditions was described and documented.</td>
<td>The nature of services provided in comparison conditions was not described or documented.</td>
<td>The nature of services provided in comparison conditions was inadequately described or documented.</td>
<td>The nature of services provided in comparison conditions was adequately described and documented.</td>
<td>The nature of services provided in comparison conditions was clearly and coherently described and documented.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Criterion/Score</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quality Indicators for Outcome Measures</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. Multiple measures were used to provide an appropriate balance between measures closely aligned with the intervention and measures of generalized performance</td>
<td>Multiple measures were not used.</td>
<td>Multiple measures were used but were of inadequate quality or appropriateness to provide a balance between measures closely aligned with the intervention and measures of generalized performance</td>
<td>Multiple measures were used, and measures were of adequate quality and appropriateness to provide a balance between measures closely aligned with the intervention and measures of generalized performance</td>
<td>Multiple high quality measures were used, and measures were carefully selected to provide a balance between measures closely aligned with the intervention and measures of generalized performance.</td>
</tr>
<tr>
<td>2. Outcomes for capturing the intervention’s effects were measured at the appropriate times.</td>
<td>No information was provided concerning the timing of measurement</td>
<td>Outcomes for capturing the intervention’s effects were measured with inadequate frequency or</td>
<td>Outcomes for capturing the intervention’s effects were measured with adequate frequency and were measured at appropriate times.</td>
<td>Outcomes for capturing the intervention’s effects were measured with ample frequency and were measured at appropriate times.</td>
</tr>
<tr>
<td>Criterion/Score</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>--------------------------------------------------------------------------------</td>
<td>-------------------------------------------------------------------</td>
<td>-------------------------------------------------------------------</td>
<td>-------------------------------------------------------------------</td>
<td>-------------------------------------------------------------------</td>
</tr>
<tr>
<td>Quality Indicators for Data Analysis</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. The data analysis techniques chosen were appropriate and linked in an integral fashion to key research questions and hypotheses.</td>
<td>No information was provided concerning data analysis techniques.</td>
<td>The data analysis techniques chosen were inappropriate or were inadequately linked to key research questions and hypotheses.</td>
<td>The data analysis techniques chosen were appropriate and were adequately linked to key research questions and hypotheses.</td>
<td>The data analysis techniques chosen were highly appropriate and linked in an integral fashion to key research questions and hypotheses.</td>
</tr>
<tr>
<td>2. The unit of analysis chosen was clearly linked to key research questions, hypotheses, and statistical analyses.</td>
<td>No information was provided concerning the unit of analysis.</td>
<td>The unit of analysis chosen was inadequately linked to key research questions, hypotheses, or statistical analyses.</td>
<td>The unit of analysis chosen was adequately linked to key research questions, hypotheses, and statistical analyses.</td>
<td>The unit of analysis chosen was clearly and coherently linked to key research questions, hypotheses, and statistical analyses.</td>
</tr>
<tr>
<td>3. Did the research report include not only inferential statistics but also effect size calculations?</td>
<td>No information was provided concerning effect size.</td>
<td>Effect size was not reported for all key measures.</td>
<td>Effect size was reported for all key measures, but the cell size was inadequate to demonstrate statistical power.</td>
<td>Effect size was reported for all key measures, and the cell size was adequate to demonstrate statistical power.</td>
</tr>
</tbody>
</table>
### Rubric for Desirable Quality Indicators

<table>
<thead>
<tr>
<th>Criterion/Score</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Data were provided concerning attrition rates. Severe overall attrition documented and explained. Attrition was comparable across samples. Overall attrition was less than 30%.</td>
<td>No information was provided concerning attrition.</td>
<td>Data were provided concerning attrition rates, but severe overall attrition was inadequately explained; attrition was not comparable across samples; or overall attrition was more than 30%.</td>
<td>Adequate data were provided concerning attrition rates; severe overall attrition was adequately explained; attrition was comparable across samples; and overall attrition was less than 30%.</td>
<td>Data were provided concerning attrition rates; severe overall attrition was clearly explained; attrition was comparable across samples; and overall attrition was less than 20%.</td>
</tr>
<tr>
<td>2. Information was provided concerning not only internal consistency reliability but also test-retest reliability and inter-rater reliability (when appropriate) for outcome measures.</td>
<td>No information was provided concerning internal consistency reliability, test-retest reliability, or inter-rater reliability.</td>
<td>Inadequate information was provided concerning internal consistency reliability, test-retest reliability, or inter-rater reliability.</td>
<td>Adequate information was provided concerning not only internal consistency reliability but also test-retest reliability and inter-rater reliability (when appropriate) for outcome measures.</td>
<td>Ample information was provided concerning not only internal consistency reliability but also test-retest reliability and inter-rater reliability (when appropriate) for outcome measures.</td>
</tr>
<tr>
<td>3. Outcomes capturing the intervention's effects were measured beyond an immediate posttest.</td>
<td>Outcomes capturing the intervention's effects were not measured beyond an immediate posttest.</td>
<td>No evidence of the criterion-related validity or construct validity of the measures was provided.</td>
<td>Inadequate evidence of the criterion-related validity or construct validity of the measures was provided.</td>
<td>Adequate evidence of the criterion-related validity and construct validity of the measures was provided.</td>
</tr>
<tr>
<td>4. Evidence of the criterion-related validity and construct validity of the measures was provided.</td>
<td>No evidence of the criterion-related validity or construct validity of the measures was provided.</td>
<td>Adequate evidence of the criterion-related validity and construct validity of the measures was provided.</td>
<td>Ample evidence of the criterion-related validity and construct validity of the measures was provided.</td>
<td>Ample evidence of the criterion-related validity and construct validity of the measures was provided.</td>
</tr>
<tr>
<td>Criterion/Score</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>--------------------------------------------------------------------------------</td>
<td>------------------------------------------------------------------</td>
<td>------------------------------------------------------------------</td>
<td>------------------------------------------------------------------</td>
<td>------------------------------------------------------------------</td>
</tr>
<tr>
<td>5. Fidelity of implementation was described and assessed in terms of surface (the expected intervention was implemented) and quality (how well the intervention was implemented) features.</td>
<td>Fidelity of implementation was not described and assessed in terms of surface (the expected intervention was implemented) and quality (how well the intervention was implemented) features.</td>
<td>Fidelity of implementation was inadequately described or assessed in terms of surface (the expected intervention was implemented) or quality (how well the intervention was implemented) features.</td>
<td>Fidelity of implementation was adequately described and assessed in terms of surface (the expected intervention was implemented) and quality (how well the intervention was implemented) features.</td>
<td>Fidelity of implementation clearly described and assessed in terms of surface (the expected intervention was implemented) and quality (how well the intervention was implemented) features.</td>
</tr>
<tr>
<td>6. Documentation of the nature of instruction or series provided in comparison conditions was provided.</td>
<td>Documentation of the nature of instruction or series provided in comparison conditions was not provided.</td>
<td>Inadequate documentation of the nature of instruction or series provided in comparison conditions was provided.</td>
<td>Adequate documentation of the nature of instruction or series provided in comparison conditions was provided.</td>
<td>Ample documentation of the nature of instruction or series provided in comparison conditions was provided.</td>
</tr>
<tr>
<td>7. The research report included audio or videotape excerpts that captured the nature of the intervention.</td>
<td>The research report did not include audio or videotape excerpts that captured the nature of the intervention.</td>
<td>The research report included audio or videotape excerpts that captured the nature of the intervention.</td>
<td>The research report included audio or videotape excerpts that captured the nature of the intervention.</td>
<td>The research report included audio or videotape excerpts that captured the nature of the intervention.</td>
</tr>
<tr>
<td>8. Results were presented in a clear, coherent fashion.</td>
<td>Results were presented in an unclear or incoherent fashion.</td>
<td>Results were presented in an inadequately clear or coherent fashion.</td>
<td>Results were presented in an adequately clear and coherent fashion.</td>
<td>Results were presented in a highly clear and coherent fashion.</td>
</tr>
</tbody>
</table>
APPENDIX B: MINDPLAY VIRTUAL READING COACH

COMPONENT DESCRIPTIONS

The major components of the MVRC intervention are diagnostic assessment, mastery instructional activities, and proficiency instructional activities. A brief overview of each component and its subcomponents follow.

Component Descriptions

Diagnostic assessment. All students using the MVRC intervention are subjected to RAPS360 diagnostic assessment in reading comprehension, phoneme segmentation, word meaning, word recognition, phonetic reading, visual tracking, passage reading fluency, sentence reading fluency, and pause assisted fluency. Comprehension assessments generate grade equivalency scores and text Lexile measures. Vocabulary activities are presented orally, to ensure a pure measure of vocabulary and avoid confounds arising from poor reading ability. These assessments are computer adaptive and can take from five to 30 minutes, depending on a student’s performance. A student with less proficiency in any skill will reach a ceiling more quickly than a student with a greater degree of proficiency, thus reducing the time required for testing and the likelihood of fatigue or frustration. Figure 2, taken with permission, details the components and assessment paths of the RAPS360 diagnostic assessment. Figure 3, taken with permission, features a screen shot of a vocabulary assessment item embedded within RAPS360.
Figure 2. RAPS 360 Full Diagnostic Assessment Paths

Figure 2. RAPS 360 full diagnostic assessment paths. Taken with permission from MindPlay Educational Software for Reading.
The phonemic awareness mastery component of the MVRC intervention targets pre-reading and emerging reading skills. Activities focus on the identification of individual sounds: consonants and vowels, then blends and diphthongs, and finally syllables and whole words. Place and manner of articulation are introduced and taught explicitly. Instructional activities include phonemic counting, phonemic segmentation, phonemic blending, phonemic substitution, syllabic counting, syllabic segmentation, and rhyming. Instruction is presented in response to student performance, with activities of increasing difficulty introduced as the student masters foundational content. Figure 4, taken with permission, is a screen shot of a phonemic

**Phonemic awareness mastery.** The phonemic awareness mastery component of the MVRC intervention targets pre-reading and emerging reading skills. Activities focus on the identification of individual sounds: consonants and vowels, then blends and diphthongs, and finally syllables and whole words. Place and manner of articulation are introduced and taught explicitly. Instructional activities include phonemic counting, phonemic segmentation, phonemic blending, phonemic substitution, syllabic counting, syllabic segmentation, and rhyming. Instruction is presented in response to student performance, with activities of increasing difficulty introduced as the student masters foundational content. Figure 4, taken with permission, is a screen shot of a phonemic
segmentation activity typical of the phonemic awareness mastery component of the MVRC intervention. Note the video window at the center left. The individual in the video window is a reading coach, and instruction and remediation are delivered via video interactions with the reading coach.

**Figure 4.** Example MVRC Phonemic Awareness Mastery Activity

![Example MVRC Phonemic Awareness Mastery Activity](image)

*Figure 4. Example phonemic awareness mastery activity. Taken with permission from MindPlay Educational Software for Reading.*

**Phonics mastery.** The phonics mastery component of the MVRC intervention includes adaptive instruction in all of the sound-symbol correspondences of standard American English. Each phonics mastery lesson contains multiple interactive activities linked to embedded assessments. Based upon student performance on those assessments, MVRC introduces lessons of gradually increasing difficulty, while providing targeted
feedback and remediation designed to improve areas of weakness and reinforce areas of strength. Vowel and consonant graphemes and digraphs are taught explicitly, as are diphthongs, blends and syllable types. In addition to teaching phoneme-grapheme correspondences, the place and manner of phonemic production are reinforced, in order to further develop phonemic awareness. Orthographic patterns are taught along with word origins and the spelling conventions associated with specific word families. Additional instructional activities involve keyboard positioning, alternate spelling patterns, irregular word recognition, and identification of homonyms. Figure 5, taken with permission, features a word identification task typical of instruction within the phonics mastery component of the MVRC intervention. In this activity, the student is presented with a printed consonant-vowel-consonant (CVC) pattern word. The reading coach articulates three words that vary in only a single phoneme. The reading coach then directs the student to select the video in which the articulated word matches the printed word.
Grammar and meaning mastery. The grammar and meaning mastery component of the MVRC intervention is designed to teach the patterns and underlying order of English grammar, rather than a set of discrete rules governing English usage. Content is introduced systematically in response to student performance on embedded assessments. Lessons provide explicit definitions, explanations, and examples, along with metacognitive and mnemonic strategies to increase retention. Students first sort, identify, and manipulate parts of speech, then phrases, sentences, and passages of increasing difficulty. Content follows in a logical sequence, such that a lesson on conjunctions, for example, will be followed by a lesson on compound sentences. Topics
covered in the grammar and meaning mastery component of the MVRC intervention include parts of speech, verb tense and agreement, sentence parts, clauses, sentence types, punctuation, capitalization, and paragraph structure. Figure 6, taken with permission, features a task typical of grammar instruction within the grammar and meaning mastery component of the MVRC intervention, and Figure 7, taken with permission, displays a typical punctuation activity. In the first activity, the student must select the correct conjugation in response to a question posed by the virtual reading coach, while in the second, the student must appropriately punctuate each sentence in response to verbal and pictorial cues.

**Figure 6.** Example MVRC Grammar Activity

*Figure 6. Example MVRC grammar activity. Taken with permission from MindPlay Educational Software for Reading.*
**Vocabulary proficiency.** The vocabulary proficiency component of the MVRC intervention is designed both to introduce and reinforce new words and to teach strategies to derive meaning from context and structure. Embedded assessments determine a student’s initial level of performance, and the intervention delivers and adapts instruction in response to ongoing performance. Vocabulary proficiency activities include associating words and images, identifying and matching synonyms, recognizing antonyms, matching antonyms, organizing words into lexical clusters, discriminating words with similar meanings, using context to derive word meanings, using word origins to derive word meanings, using word origins to derive word spellings, and using memory.
strategies to recall newly acquired words. Figure 8, taken with permission, features an advanced vocabulary proficiency activity, in which the student is asked to separate a series of related words strung together in a word chain.

**Figure 8.** Example MVRC Vocabulary Activity

*Figure 8. Example MVRC vocabulary activity. Taken with permission from MindPlay Educational Software for Reading.*

**Fluency and comprehension proficiency.** The fluency and comprehension proficiency component of the MVRC intervention contains activities designed to promote reading rate and accuracy, as well as understanding of written texts of gradually increasing complexity. Text presentation varies in response to student performance, and texts are selected from among a library of over 1,000 passages, ranging from
kindergarten to grade 12 in difficulty. Fluency and comprehension proficiency activities include eye tracking exercises, in which the student tracks and identifies rapidly moving symbols; high frequency word practice, in which the student is required to quickly identify the most commonly used words of English; passage fluency, in which the student must quickly read and comprehend passages of increasing complexity; pause-assisted fluency, in which the student must quickly read and comprehend sentences of increasing complexity; and comprehension, in which the student reads novel passages and makes textual inferences and answers questions relative to passage content. Figure 9, taken with permission, features an eye tracking activity, in which the student is asked to watch numbers as they move across the screen while counting the number of times a particular number appears. Figure 10, taken with permission, displays a comprehension activity, in which the student is asked to infer characteristics of the main character based upon story content, while Figure 11 depicts a fluency activity, in which a student must rapidly read a text and respond to questions designed to measure comprehension.
Figure 9. Example MVRC Eye Tracking Activity

Figure 9. Example MVRC eye tracking activity. Taken with permission from MindPlay Educational Software for Reading.
Figure 10. Example MVRC Reading Comprehension Activity

Once in a Lifetime

Beautiful French countryside flew past the window as the car sped along the winding road. Rebecca and her family were on vacation in the south of France. Her parents had explained that it was a once in a lifetime trip, and now Rebecca could see why.

Based on this story, what does Rebecca like?

Rebecca likes turtles.

Rebecca likes history.

Rebecca likes traveling.

Figure 10. Example MVRC reading comprehension activity. Taken with permission from MindPlay Educational Software for Reading.
**Figure 11.** Example MVRC Reading Fluency Activity

Ask several times a day. This keeps the egg warm all over and keeps the chicks from sticking to the shell.

The book says the eggs must have the right heat, light, air, and dampness. I know Lady did that right because in 21 days, the chicks hatched.

When I went to the hen house, I heard peeping sounds coming from the eggs. I wanted to see an egg hatch, but Lady had Them covered.

At the same time, my class hatched 12 chicken eggs. We put these eggs in a machine called an incubator. The machine kept the right heat, light, and dampness.

*Figure 11.* Example MVRC reading fluency activity. Taken with permission from MindPlay Educational Software for Reading.
APPENDIX C: PLANNED ACTIVITY CHECK

Planned Activity Check (PLA-check) for Fidelity

Date: ______________ Site _______ Start Time: ______(am/pm) End Time: ______(am/pm)
Observer: ___________________________

Directions: Note the total number of students (participants) in the classroom and the time at which the observation begins. At the designated 2 minute interval, scan slowly from left to right (or right to left; whichever gives the observer a better view) and classify the students according to any of the off-task behaviors. The scan should take no longer than 30 seconds. Count the number of off-task students and subtract this number from the total number of students in the classroom. Each scan should take about 30 – 45 seconds, but not longer than a minute. Continue scanning at designated 2 minute intervals for a total of 20 minutes (10 times).

Off-task Behaviors:

a. Student is not wearing headphones over their ears or in both ears.

b. Student is engaged in other activities besides MVRC (e.g. head down, looking away from computer, talking with another student, has program open on computer other than MVRC).

Please note: Regular instructional interaction (e.g. student talking with teacher) or regular everyday occurrences (e.g. leaving the room to use the restroom, getting called to the front office) do not count as being off task. These are events that are normal occurrences in the classroom and do not impact the delivery of the intervention. In the rare instance these occurrences happen beyond the norm for a particular student (e.g. leaving the classroom several times for various and possibly unnecessary reasons), note the name of the student, and the number of times they leave the room. This will be analyzed separately.

Number of participants ______

<table>
<thead>
<tr>
<th>Time (e.g. Time 1=T1)</th>
<th>Off-Task</th>
<th>On-Task</th>
</tr>
</thead>
<tbody>
<tr>
<td>T1</td>
<td>_______</td>
<td>_______</td>
</tr>
<tr>
<td>T2</td>
<td>_______</td>
<td>_______</td>
</tr>
<tr>
<td>T3</td>
<td>_______</td>
<td>_______</td>
</tr>
<tr>
<td>T4</td>
<td>_______</td>
<td>_______</td>
</tr>
<tr>
<td>T5</td>
<td>_______</td>
<td>_______</td>
</tr>
<tr>
<td>T6</td>
<td>_______</td>
<td>_______</td>
</tr>
<tr>
<td>T7</td>
<td>_______</td>
<td>_______</td>
</tr>
<tr>
<td>T8</td>
<td>_______</td>
<td>_______</td>
</tr>
<tr>
<td>T9</td>
<td>_______</td>
<td>_______</td>
</tr>
<tr>
<td>T10</td>
<td>_______</td>
<td>_______</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>_______</td>
<td>_______</td>
</tr>
</tbody>
</table>
REFERENCES


Slavin, R. E. & Madden, N. A. (2012). *Success for All: Summary of research on achievement outcomes* (revised). Baltimore, MD: Johns Hopkins University, Center for Research and Reform in Education.


