A Meta-Analysis of Research-Based Reading Interventions with English Language Learners

Maritza Torres
University of Denver

Follow this and additional works at: https://digitalcommons.du.edu/etd

Part of the Educational Administration and Supervision Commons, Educational Leadership Commons, and the Educational Psychology Commons

Recommended Citation
Torres, Maritza, "A Meta-Analysis of Research-Based Reading Interventions with English Language Learners" (2016). Electronic Theses and Dissertations. 1165.
https://digitalcommons.du.edu/etd/1165

This Dissertation is brought to you for free and open access by the Graduate Studies at Digital Commons @ DU. It has been accepted for inclusion in Electronic Theses and Dissertations by an authorized administrator of Digital Commons @ DU. For more information, please contact jennifer.cox@du.edu,dig-commons@du.edu.
Meta-Analysis of Research-Based Reading Interventions with English Language Learners

A Dissertation
Presented to
the Faculty of the Morgridge College of Education
University of Denver

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

by
Maritza Torres
June 2016
Advisor: Dr. Susan Korach
ABSTRACT

The Response to Intervention (RTI) model, introduced as part of the reauthorization of the Individuals with Disabilities Education Act in 2004, is a proactive process of early interventions and evidence-based instruction for all students. RTI has additional intensive and individualized interventions to prevent student underachievement, including students at risk for academic failure and culturally and linguistically diverse students (Vellutino et al., 1996; Donovan & Cross, 2002; Francis, Rivera, Lesaux, Kieffer, & Rivera, 2006; Fuchs & Fuchs, 2006; Vellutino, Scanlon, Small, & Fanuele, 2006; Al Otaiba et al., 2009). Klingner and Edwards (2006) suggest that the needs of culturally and linguistically diverse students differ from the general population of students. Research indicates challenges with RTI implementation with English Language Learners (ELLs) (Klingner, 2010). There is a growing body of research on RTI implementation; however, evidence-based interventions are not applicable to all students and the impact of interventions on ELL students is not clear.

The purpose of the study was to determine if there was empirical support for Tier I, Tier II, and Tier III research-based reading interventions that produce improvement in reading for ELLs. This meta-analysis included twenty-seven studies published from 2005.
through 2013 that quantitatively examined the effects of research-based reading interventions for ELLs as part of the RTI model.

The meta-analysis raised questions about the dominance of Tier II interventions in the research, the lack of difference between treatment and control groups, and the teacher’s background and context. This study was expecting to find a difference between the treatment and control groups receiving RTI interventions but instead it revealed large effect sizes for control and treatment groups across interventions except for Tier II interventions targeting reading comprehension. Therefore, before adopting Tier I and Tier II reading programs for ELL students, education leaders need to carefully examine results of these interventions with this subgroup. A key element of the culturally and linguistically responsive RTI model is the need for teachers with culturally responsive practices and knowledge about the needs of ELLs (Klingner & Edwards, 2006). The primary studies targeted the essential reading components proposed by the National Reading Panel, conducted trainings about the implementation of the intervention, and used rigorous methods to ensure fidelity of the intervention but there was not clear evidence of linguistically and culturally responsive practices. This finding suggests that future research with ELLs and RTI should address the preparation of teachers or personnel delivering the interventions and investigate possible moderators that can explain the heterogeneity among effects sizes.
ACKNOWLEDGMENTS

To my parents and family: thank you for your love and patience during this long journey. Your encouragement and kind words were essential to accomplishing this dream.

To my dissertation committee: thank you for your ongoing support and advice. It has been an honor to work under your guidance. I would like to thank and acknowledge my advisor, Dr. Susan Korach, for helping me analyze my research with English Language Learners through a policy and leadership perspective. I would like to thank Dr. Ellen Miller-Brown for sharing your expertise about English Language Learners, literacy, and Response to Intervention. Thank you to Dr. Antonio Olmos for your invaluable support with the methodology and statistical analysis. To Dr. Michelle Vanchu-Orosco, my sincerest gratitude for your dedication and patience. Thank you for spending long hours sharing your expertise about meta-analysis. I will never forget our “boot camps.”
TABLE OF CONTENTS

CHAPTER ONE ......................................................................................................................1
Introduction .........................................................................................................................1
Statement of the Problem .....................................................................................................6
Purpose of the Study .............................................................................................................9
Research Questions ..........................................................................................................11
Definitions ............................................................................................................................11

CHAPTER TWO ....................................................................................................................18
Literature Review .................................................................................................................18
Ethnic Composition and ELL Representation in the United States School System ..........18
Achievement Gap ................................................................................................................21
English Language Learners .................................................................................................22
Reading Components .........................................................................................................25
Systems of Support for ELL’s Students .............................................................................31

CHAPTER THREE ..................................................................................................................40
Methods ................................................................................................................................40
Purpose of Study ................................................................................................................40
Research Questions ............................................................................................................41
Meta-analysis .......................................................................................................................41
Criteria for Selection of Studies .........................................................................................42
  Distinguishing features of the study .................................................................................42
  Research respondents .......................................................................................................43
  Key variables .....................................................................................................................43
  Research methods .............................................................................................................43
  Cultural and linguistic range ............................................................................................44
  Time frame .........................................................................................................................44
  Publication type .................................................................................................................44
  Exclusion criteria ..............................................................................................................44
Finding Relevant Literature ...............................................................................................45
Coding Study Characteristics and Empirical Findings .........................................................47
Statistical Analysis .............................................................................................................51

CHAPTER FOUR ....................................................................................................................55
Results ..................................................................................................................................55
Study Characteristics ...........................................................................................................56
Analysis by Tier ....................................................................................................................65
Tier I .....................................................................................................................................67
  Mean difference analysis with post-test data .................................................................67
  Mean gain analysis with tier I pre-post studies treatment versus control group ..........69
Tier II ................................................................. 71
  Mean difference analysis with post-test data .................................... 71
  Mean gain analysis with tier I pre-post studies treatment versus control group ..... 72
Analyses by Reading Components ......................................................... 75
  Reading Fluency ........................................................................... 75
  Alphabets: Phonological Awareness and Phonics .................................. 76
  Reading Comprehension .................................................................. 79
Publication Bias ................................................................................. 82

CHAPTER FIVE .................................................................................... 85
Discussion ............................................................................................ 85
  Dominance of tier II interventions ..................................................... 91
  Lack of a difference between treatment and control group .................... 91
  Teacher’s Background and Context ................................................... 92
Limitations .......................................................................................... 93
Conclusions .................................................................................... 93

REFERENCES ..................................................................................... 99

APPENDICES ................................................................................... 114

Appendix A ....................................................................................... 114
Appendix B ....................................................................................... 115
Appendix C ....................................................................................... 134
Appendix D ....................................................................................... 138
Appendix E ....................................................................................... 143
Appendix F ....................................................................................... 144
Appendix G ....................................................................................... 145
Appendix H ....................................................................................... 146
Appendix I ....................................................................................... 148
Appendix J ....................................................................................... 149
Appendix K ....................................................................................... 150
Appendix L ....................................................................................... 151
Appendix M ....................................................................................... 152
Appendix N ....................................................................................... 155
LIST OF TABLES

TABLE 1 .................................................................................................................. 4
TABLE 2 .................................................................................................................. 19
TABLE 3 .................................................................................................................. 21
TABLE 4 .................................................................................................................. 58
TABLE 5 .................................................................................................................. 60
TABLE 6 .................................................................................................................. 62
TABLE 7 .................................................................................................................. 63
TABLE 8 .................................................................................................................. 64
TABLE 9 .................................................................................................................. 65
TABLE 10 ............................................................................................................... 67
TABLE 11 ............................................................................................................... 68
TABLE 12 ............................................................................................................... 70
TABLE 13 ............................................................................................................... 71
TABLE 14 ............................................................................................................... 73
TABLE 15 ............................................................................................................... 75
TABLE 16 ............................................................................................................... 77
TABLE 17 ............................................................................................................... 79
TABLE 18 ............................................................................................................... 80
TABLE 19 ............................................................................................................... 81
LIST OF FIGURES

FIGURE 1 .........................................................................................................................4
FIGURE 2 .........................................................................................................................19
FIGURE 3 .........................................................................................................................69
FIGURE 4 .........................................................................................................................72
FIGURE 5 .........................................................................................................................83
FIGURE 6 .........................................................................................................................83
CHAPTER ONE

Introduction

The U.S. Department of Education (2003) indicates that difficulties in reading are the most common reason Latino students receive special education services and limit their participation in the workplace and in society (Al Otaiba et al., 2009). The Institute of Education Sciences selected a National Literacy Panel of thirteen experts to synthesize quantitative and qualitative research on the development of literacy in language-minority students (August & Shanahan, 2006). As stated by the National Literacy Panel (2006), difficulties reading and writing proficiently in English hinder the participation of language minority students in American schools, workforce, and society (August & Shanahan, 2006).

Prior to the reauthorization of the Individuals with Disabilities Education Act (IDEA) in 2004, experts viewed the special education system as a “wait-to-fail” model instead of a system that provides students with high quality evidence-based intervention within the regular education system (Martín, 2014). Traditionally, students had to wait until a significant discrepancy between reading achievement and intelligence was demonstrated to receive reading interventions (Donovan & Cross, 2002). The Response to Intervention (RTI), introduced as part of the reauthorization of IDEA in 2004, provides a proactive process of early interventions and evidence-based instruction to all students.
with additional intensive and individualized interventions to prevent student underachievement, including students at risk for academic failure and culturally and linguistically diverse students (Vellutino et al., 1996; Donovan & Cross, 2002; Francis, Rivera, Lesaux, Kieffer, & Rivera, 2006; Fuchs & Fuchs, 2006; Vellutino, Scanlon, Small, & Fanuele, 2006; Al Otaiba et al., 2009).

RTI is a multi-tiered model of intervention with graduated levels of support. Tier I of RTI encompasses universal, high-quality instruction and assessment in regular education for all students (Hazelkorn, Bucholz, Goodman, Duffy, & Brady, 2011). Tier II focuses on specialized interventions for students who are not making adequate progress in the core program or in Tier I (Hazelkorn et al., 2011). Tier III focuses on students who are presenting reading difficulties and did not respond to Tier I and II interventions. Tier III is based on individual student’s needs, and it provides intensive and sustained intervention with frequent progress monitoring (Vaughn & Roberts, 2007). Interventions at each level, or tier, should be based on scientific evidence of effectiveness (Klingner & Edwards, 2006). Additionally, there can be within each of these levels of intervention more than one intervention (National Center on Response to Intervention, 2010). The main difference between the tiers is the intensity of the interventions and the frequency of the measurements (Reschly, 2005).

Under the RTI model, as part of the eligibility process, the special education team rules out that the cause of poor academic achievement and possible specific learning disability is not due to other factors such as visual, hearing, or motor disability; intellectual disability; emotional/ behavioral disability; cultural factors; attendance
problems and/or high mobility rate; classroom behavior; environmental or economic factors; or limited English proficiency. However even with RTI, the misidentification of English Language Learners (ELLs) for special education still persists due to different factors including language, assessment, and instruction (Marchand-Martella, Klingner, & Martella, n.d.; McCardle, Mele-McCarthy, & Leos, 2005; Skiba et al., 2008).

The retention and dropout rates of ELLs are more prevalent compared to non-ELL students (Zehler et al., 2003; Orosco & Klingner, 2010). The National Center for Education Statistics (2005) reported that 73% of ELL children in the fourth grade scored below the “basic” level of reading, suggesting that a significant number did not acquire even partial mastery of the skills required for grade level work (as cited in Farver, Lonigan, & Eppe, 2009). In 2007, non-ELLs in the fourth grade scored 36 points higher than ELLs on the National Assessment of Educational Progress (NAEP) for reading and 25 points higher in math. The achievement gap between ELL and non-ELL eighth grade students was 42 points in reading and 37 points in math (Goldenberg, 2008). The Alliance for Excellent Education (2012) states

nationally, millions of students in grades 7–12 are at risk of dropping out of high school because of low literacy skills, poor attendance, and class failure. Unfortunately, many of these students come from groups that are underserved and underrepresented: students of color, high-mobility students (including foster, migrant, and homeless students), English language learners, students with disabilities, and low-income students. (Alliance for Excellent Education, p.1; cited by Marchand-Martella, Klingner & Martella, n.d.)

According to NAEP (2011), results in reading and mathematics from 2002 through 2009 indicate that the academic achievement of African American and Hispanic students, including ELLs, is significantly lower than White students. From 1992 to 2009,
there were no significant changes in the size of the Hispanic-White reading gap for students at grades 4 and 8 (Figure 1) (NAEP, 2011a). In 2009, the reading achievement gap between Hispanic and White students was 25 points at grade 4 and 24 points at grade 8 (Table 1).

![Achievement Gap Trend between Hispanic and White Student](image)


<table>
<thead>
<tr>
<th></th>
<th>White-Hispanic</th>
<th>Non-ELL Hispanic</th>
<th>White-Non-ELL Hispanic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mathematics</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Grade 4</td>
<td>21</td>
<td>19</td>
<td>14</td>
</tr>
<tr>
<td>Grade 8</td>
<td>26</td>
<td>34</td>
<td>19</td>
</tr>
<tr>
<td>Reading</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Grade 4</td>
<td>25</td>
<td>29</td>
<td>15</td>
</tr>
<tr>
<td>Grade 8</td>
<td>24</td>
<td>39</td>
<td>15</td>
</tr>
</tbody>
</table>

**Table 1. 2009 Hispanic-White Achievement**

High quality research-based instruction and interventions are important components of RTI to prevent academic and behavioral difficulties and address the needs
of students who are not making expected progress (World-Class Instructional Design and Assessment, 2013). The success of RTI with culturally and linguistically diverse students might be positively impacted by the prevalence of research-based interventions validated with this population and that are culturally responsive (Klingner & Edwards, 2006). Klingner and Edwards (2006) suggest that the needs of culturally and linguistically diverse students differ from the general population of students. These authors developed a revised RTI model that emphasizes culturally responsive practices and evidence-based interventions at each level (Klingner, 2010). Tier I includes two important components: (a) evidenced-based interventions validated with diverse populations and (b) teachers who have developed culturally responsive practices and have knowledge about the needs of ELLs. Educators that work with ELLs need preparation to understand different factors that influence ELL students’ learning and interactions including their sociocultural background and language acquisition process, as well as teaching methods for English as a second language. Tier II is characterized by more intensive interventions when culturally and linguistically diverse students are not responding to Tier I methods. Tier III may include a referral to a Child Study Team made up of diverse personnel with a wide range of skills. This team helps to determine if the student referred has had meaningful interventions and opportunities to learn. The Child Study Team can determine if the student needs more intensive ongoing support or perhaps special education services.¹

¹ Please see Appendix A.
Statement of the Problem

Klingner and Edwards (2006) suggest that the needs of culturally and linguistically diverse students differ from the general population of students. Research indicates challenges with RTI implementation with English Language Learners (ELLs) (Klingner, 2010). There is a growing body of research on RTI implementation; however, evidence-based interventions are not applicable to all students and the impact of interventions on ELL students is not clear. There are additional factors to consider for a successful implementation of this model with this population (Orosco & Klingner 2010). Orosco and Klingner’s (2010) research indicates three important challenges with RTI implementation with ELLs. First, the preparation of educators to work with ELLs requires understanding of second language acquisition that helps teachers to differentiate between language acquisition and learning disabilities as well as a training that provides effective instructional and assessment practices. Second, the tendency of school personnel to find weaknesses within the child and overlook the environment and instructional context that affect the student. Third, another challenge with RTI, and the focus of this study, is the assumption that evidence-based interventions are applicable to all students: a “one size fits all” model (Orosco & Klingner, 2010, p. 271). Klingner and Edwards (2006) recommend to validate interventions with students in this case ELLs that are part of the target population and disaggregate the results to examine the differences across students from different backgrounds (Klingner & Edwards, 2006). The present study examined research on reading interventions and focused on evidence-based literacy interventions for ELLs implemented as part of the RTI model.
The assumption that evidence-based interventions are applicable to all students can create inappropriate referrals leading to misidentification of ELLs in special education (Orosco & Klingner, 2010). For example, the RTI model requires schools collect data and monitor progress to demonstrate whether students are responding to research-based interventions (Gresham et al., 2005). If a student is not demonstrating adequate progress, educators need to evaluate the instruction before they assume a problem within the child (Ortiz, 1997; Klingner & Edwards, 2006). Schools need to evaluate if the tiers of intervention are structurally sound, implemented as intended, and if the general population as well as specific subgroups are achieving successful outcomes (Fien et al., 2010). Instruction and interventions that are developed and implemented without consideration of the specific language and learning needs of ELL students could impact their performance (Marchand-Martella, Klingner, & Martella, n.d.). Language proficiency and dominance are important variables that can influence intervention results (Ortiz, 1997; Klingner & Edwards, 2006). ELLs are not all the same: these students have different levels of English language acquisition that can impact their rates of improvement. As stated by Howell, Fox, and Morehead (1993), by age six, a native English speaker has already learned 13,000 words and has basic grammar before he or she enters school. ELLs learning to read in a second language begin the process with a very different knowledge base because they have limited exposure to phonology and vocabulary in English and less background knowledge related to English text passages (Nelson, 2003).

An extensive body of research reports that ELL students typically require at least five years to catch up to native speakers in academic language proficiency.
Students whose first language is English are not standing still waiting for ELL students to catch up. Every year, they make gains in reading, writing, and vocabulary abilities. So ELL students have to run faster to bridge the gap. (Cummins, n.d. p. 3)

Besides considering language dominance and proficiency, it is recommended that school staff gather information about different factors that can impact ELLs’ academic and linguistic development and response to instruction and intervention. These factors include the learning environment, academic achievement and instruction, oral language and literacy, personal and family, physical and psychological, previous schooling, and cross cultural factors (Hamayan, 2013; World-Class Instructional Design and Assessment, 2013). For instance, findings from the National Literacy Panel (2006) suggest that similar approaches to teach reading and writing, especially instruction that provides substantial exposure to the essential reading components including phonemic awareness, phonics, fluency, vocabulary, and reading comprehension, are effective with non-ELLs and ELLs but not sufficient with ELLs. High quality instruction for ELLs must address oral language development. “The need to develop stronger English-language proficiency to become literate in English argues for an early, ongoing, and intensive effort to develop this oral proficiency” (August & Shanahan, 2006, p. 5) including vocabulary and background knowledge in English (August & Shanahan, 2006). Based on this information, it is clear that learning to read in a second language requires additional instructional approaches than those utilized with English-only students (August & Shanahan, 2006; Cummins, 2007; Ortis & Klingner, 2010). This is because ELLs can struggle with phonological awareness in English because “some phonemes may not be present in students’ native language and, therefore, may be difficult to distinguish
auditorily from similar sounds” (Ortis & Klingner, 2010, p. 271). Despite these differences, recommendations for teaching ELLs to read focus on the similarities between learning to read in the first and second language overlooking important distinctions (Gersten et al., 2007).

**Purpose of the Study**

The purpose of the study is to (a) determine whether there is empirical support for research-based reading interventions that produce improvement in reading for ELLs and (b) estimate the strength of this relationship.

There is research on ELL students and reading interventions, however, the impact of RTI with ELL students is not clear. Orosco and Klingner (2010) conducted a study to evaluate the implementation of RTI in an urban elementary school with a large percentage of Latino ELLs struggling in reading. They found that misalignment between assessment and instruction, negative school culture, problems with teacher preparation, and limited resources negatively impacted the implementation of RTI with ELLs.

A previous meta-analysis conducted by the National Reading Panel (2000) evaluated different methods for teaching reading and concluded that phonemic awareness, phonics, reading fluency, vocabulary, and reading comprehension are critical components for teaching reading to young children as well as adolescents. However, the National Reading Panel (2000) stated these findings “did not address issues relevant to second language learning” (p. 3). Later, the National Literacy Panel (2006) confirmed the benefits of these components with ELL students but stated high quality instruction for ELLs must include substantial support of oral language development in English (August
& Shanahan, 2006, p. 5). The National Literacy Panel also concluded instructional approaches need adjustments to provide more benefits to ELLs (e.g., more work with specific phonemes in English that do not exist in the student’s home language) and vocabulary and background knowledge in English need to be addressed intensively with ELLs.

The results of the present study provided an understanding of the implementation and effectiveness of RTI with ELLs and revealed implications for policy and school-based leadership. Therefore, the current study sought, classified, and analyzed the existing research of reading interventions with ELLs since the implementation of RTI.

This study used meta-analysis to classify existing studies into three different tiers of intervention, Tier I, Tier II, and Tier III, and aggregate and compare findings from different studies. This study included experimental and quasi-experimental studies that quantitatively examine the effects of interventions with ELL students from Kindergarten through 8th grade attending public schools, in English speaking countries that are implementing the RTI model. Eligible studies reported at least one quantitative test of reading including phonemic awareness, phonics, reading fluency, vocabulary, and /or reading comprehension.

Because this method focuses on the aggregation and comparison of findings, the present meta-analysis included results of different studies that present similar constructs and relationships and similar statistical forms of analyses (Lipsey & Wilson, 2001). To calculate an estimated effect size of the impact means and standard deviations or significance test results are necessary (Lipsey & Wilson, 2001; Vanchu-Orosco, 2012).
Research Questions

This study addressed the following research questions:

1. Is there empirical support for effects of Tier I interventions on reading for ELL students?
2. Is there empirical support for effects of Tier II interventions on reading for ELL students?
3. Is there empirical support for effects of Tier III interventions on reading for ELL students?

Definitions

The definitions that apply to this study include the terms ELL, the Response to Intervention model (RTI), and the meta-analytic techniques. The RTI definition includes the tiers of intervention, and research-based interventions. This study reviewed the existing research on reading interventions; therefore, the essential components for developing reading need to be included. Terms for meta-analysis include effect size, mean effect, Q statistic, fixed-effects model, random-effects model, and publication bias.

- English Language Learners (ELLs) — the National Literacy Panel (August & Shanahan, 2006) defines this term as “students who come from language backgrounds other than English and whose proficiency is not developed enough to where they can profit fully from English only instruction” (Fien et al., 2011, p. 143). The majority of ELLs in the United States speak Spanish (Zehler et al., 2003), but there are differences within this group regarding country of origin,
ethnicity, socioeconomic background, immigration status, generation (August & Hakuta, 1997), educational background, and literacy in native language.

• Response to Intervention — The National Center on Response to Intervention (2014) defines RTI as a system that incorporates assessment and intervention to enhance students’ academic achievement and behavior. Within this model, schools implement evidence-based interventions, use assessment and data to identify students at-risk, apply progress monitoring tools, and adjust the intensity and type of intervention based on the students’ response to the intervention.

• Tiers of Intervention — RTI is a multi-tiered model of prevention and intervention that provides to students with more intensive instructional support during each successive tier (Stecker, 2007). Interventions at each level, or tier, should be based on scientific evidence of effectiveness (Klingner & Edwards, 2006), with the main difference between the tiers being “intervention intensity and measurement precision” (Reschly, 2005, p. 511). More intensive interventions and support are necessary when students at-risk demonstrate lack of response in previous tiers (Stecker, 2007).

• Tier I — encompasses universal screening, classroom based-instruction, and assessment in the general education classroom with all students (Vaughn & Roberts, 2007; Hazelkorn et al., 2011). Tier I includes scientifically based reading instruction and curriculum with emphasis on the essential reading components (phonemic awareness, phonics, reading fluency, reading comprehension, and
vocabulary) and benchmark assessments three times per year (Vaughn & Roberts, 2007).

- Tier II — focuses on specialized and targeted interventions for students who are not making adequate progress in the core program or in Tier I (Hazelkorn et al., 2011). The students at-risk receive targeted instruction to help close the gap between their current performance and their expected performance. The specialized, scientifically based instruction can be 20-30 minutes in addition to Tier I. The progress monitoring or assessments occur twice a month to guarantee optimal progress and learning.

- Tier III — interventions at this level provide intensive scientifically based instruction to students with significant difficulty in reading that did not respond sufficiently to Tier I and Tier II. The small group instruction may be provided for 50 minutes per session. Progress monitoring occurs at least twice per month (Vaughn & Roberts, 2007).

- The Culturally and Linguistically Response RTI Model — this model emphasizes culturally responsive practices and evidence-based interventions at each level (Klingner, 2010). Tier I includes two important components: (1) evidenced-based interventions validated with diverse populations and (2) teachers who have developed culturally responsive practices and have knowledge about the needs of ELLs. Tier II is characterized by culturally intensive interventions when diverse students are not responding to Tier I methods. Tier III includes a referral to a Child Study Team made up of diverse personnel with diverse expertise pertaining
to the ELLs. This team helps to determine if the student referred has had meaningful interventions and opportunities to learn. The Child Study Team then determines if the student needs more intensive, ongoing support, or perhaps, special education services (Klingner, 2010).

- Research-based interventions or evidence-based intervention — a core component of RTI is defined as an intervention for which data from scientific, rigorous research designs have demonstrated (or empirically validated) the efficacy of the intervention. That is, within the context of a group or single-subject experiment or a quasi-experimental study, the intervention is shown to improve the results for students who receive the intervention. (National Center on Response to Intervention, 2014, p. 6)

In terms of reading, the National Reading Panel identified alphabolics including phonemic awareness and phonics, reading fluency, reading comprehension, and vocabulary as essential components for developing reading (National Reading Panel, 2000). While there are multiple definitions of these essential reading components, this study uses the definitions promoted by the National Reading Panel as a framework.

- Phonemic awareness (PA) — defined as the ability to manipulate, blend and segment sounds or phonemes in oral syllables and words. Unlike phonics instruction, PA does not rely on letter-sound relations when teaching students to read and spell.

- Phonics — phonics instruction focuses on letter sound correspondence and spelling patterns to teach students how to read and spell.
• Reading fluency — the ability to read orally with “speed, accuracy, and proper expression” (National Reading Panel, p. 3) facilitating reading comprehension.

• Reading comprehension — The National Reading Panel referred to the definition by Durkin (1993). This author viewed comprehension as an active and intentional thinking process “during which meaning is constructed through interactions between text and reader” (National Reading Panel, p.4-39). Besides being an interactive process, National Reading Panel notes that reading comprehension is a cognitive process that requires complex skills, involves the understanding of vocabulary, and needs the preparation of educators so they can support students on developing this skill (National Reading Panel, 2000)

• Vocabulary — there are two types of vocabulary: expressive and receptive. Expressive vocabulary refers to words individuals produce for verbal and written communication. Receptive vocabulary refers to the words individuals recognize by listening and reading. This component was classified by the National Reading Panel as critical in understanding the development of reading comprehension. Both reading comprehension and vocabulary involve the meaning of text at different levels. Vocabulary is tied to individual words and comprehension to larger units (National Reading Panel, 2000).

• Effect size — defined “as an index of the direction and magnitude of association between two variables and may include differences between groups, correlation between two variables, and contingencies between two dichotomies” (Card, 2012, p. 87). The effect size statistic must represent quantitative findings in a
standardized form allowing the researcher to conduct comparisons and analysis across studies (Lipsey & Wilson, 2001). The effect size allows the researcher to calculate its standard error and give more weight to studies that have small standard errors than those with large standard errors or less precise estimates (Card, 2012).

- Mean effect size — the most important index of central tendency in a meta-analysis is mean effect size. It allows researchers to describe the typical effect sizes for a particular study. “The mean effect size is calculated by computing the product of each study’s effect size by its weight, summing these products across studies, and dividing this value by the sum across studies” (Card, 2012, p. 181).

- Heterogeneity test — involves calculating the Q value and shows the amount of heterogeneity of effect sizes across studies. \( I^2 \) or, index of the magnitude of heterogeneity, is used to determine the percentage of variability among effect sizes. \( I^2 \) of 25%, 50% and 75% are small, medium, and large effect sizes, respectively (Card, 2012).

- Fixed-effect model — under this model it is assumed that all the studies have in common a single effect (Borenstein, Hedges, Higgins, & Rothstein, 2009; Card, 2012). Card (2012) recommends the fixed effect model when the Q-test for the distribution is not statistically significant.

- Random-effects model — “the true effects in the studies are assumed to have been sampled from the distribution of true effects” (Borenstein et al., 2009, p. 74)
The random-effects model allows the researcher to extrapolate results to the general population (A. Olmos, personal communication, November, 2015).

- **Publication bias** — refers to the possibility that studies, which did not find statistically significant effects, are more likely to be unpublished than studies reporting positive effects. The problem with publication bias is that the published literature may not be representative of the studies conducted on a topic and can yield a stronger overall effect size than if all studies were included as part of the meta-analysis (Card, 2012, p. 257). There are different methods to manage publication bias including moderator analyses, funnel plots, Trim and Fill, Failsafe number, and Trim and Fill (Card, 2012).
CHAPTER TWO

Literature Review

This literature review contains an overview of research on English Language Learner (ELLs), the Response to Intervention (RTI) model, and research-based interventions with ELLs. The first section will focus on the ethnic composition in the United States and representation of ELLs in the U.S. school system, the achievement gap, and then characteristics of ELLs and second language acquisition. The second section will address research regarding the essential reading components and research on reading skills for ELL students. The following section will provide an overview of systems of support for ELL students including the RTI model.

Ethnic Composition and ELL Representation in the United States School System

The number of ELL students continues to grow in the United States and their educational needs cannot be overlooked (Nelson, 2003). From 1980 to 2009, the largest population growth rate was for Hispanics compared with Whites and Blacks across the United States. For the period 1984-2011, the Hispanic school enrollment increased from approximately 9% to 24% (Figure 2) (PEW Research Center, 2012). Between 2008 and 2025, the Hispanic population is expected to grow to 21% of the U.S. population (National Center for Education Statistics, 2010).
### Table 2. Enrollment Rates by Ethnicity between 2000-2001 and 2007-2008 in the United States

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Hispanic</td>
<td>17%</td>
<td>21%</td>
</tr>
<tr>
<td>White</td>
<td>61%</td>
<td>56%</td>
</tr>
<tr>
<td>African Americans</td>
<td>17%</td>
<td>17%</td>
</tr>
<tr>
<td>Asian</td>
<td>4%</td>
<td>5%</td>
</tr>
<tr>
<td>American Indians</td>
<td>1%</td>
<td>1%</td>
</tr>
</tbody>
</table>

Between 2000-2001 and 2007-2008 (Table 2), the percentage of White students decreased from 61% to 56%, Asian students increased from 4% to 5%, and African American students and American Indian students’ rate of enrollment remained stable with 17% and 1%, respectively (National Center for Education Statistics, 2010).

Racial ethnic composition varies from state to state. In 2008, the West had the highest percentage of Hispanics, Asians, and Native Hawaiians or other Pacific Islanders.
Among the 50 states, New Mexico had the highest percentage of Hispanics (45%), followed by California (37%) and Texas (36%) (National Center for Education Statistics, 2010). Based on data from Common Core of Data (Table 3), in the school year 2011-2012, Hispanics represented 31.9% of the school population in Colorado. White students represented 56.1%, Black students 4.8, Asian/Pacific Islander students 3.32%, and American Indian/AK Native students 0.84% (National Center for Education Statistics, 2014a). In the same school year, Alaska, California, Colorado, Hawaii, Nevada, New Mexico, Oregon, and Texas served the highest percentage of ELL students in public schools (10% or more of public school students). In the school years 2011-2012 and 2012-2013, California had the highest percentage of ELLs, 23% and 22% respectively. In the school years 2011-2012 and 2012-2013, Nevada had 19.6% and 15.7% ELL students, New Mexico (16% and 15.8%), Texas (14.9% and 15.1%), and Oregon (11.3% and 8.9%). In Colorado 12% of students were ELLs in the school years 2011-2012 and 2012-2013 (National Center for Education Statistics, 2014b).
Achievement Gap

The demographic of the student population in the United States has changed over time; however, the discrepancies between the academic achievement of White students and non-White students persist. “The achievement gap occurs when one group of students outperforms another group and the difference in average scores for the two groups is statistically significant and larger than the margin error” (The National Assessment of Educational Progress, 2011, p. 1). NAEP (2011) stated that singular assessments goals measure students’ performance by identifying gaps and trends over time but fail to explain causation of the achievement gap. Barton and Coley (2009), on the other hand, investigated conditions and experiences at school, home, and beyond school that are correlated with the achievement gap. Their findings suggest minority and low-income students are less likely to receive instruction from certified and experienced teachers, have less access to technology, attend large classes, and worry about feeling

<table>
<thead>
<tr>
<th>Ethnicity</th>
<th>Colorado</th>
<th>U.S. Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>White</td>
<td>479,288</td>
<td>501,619</td>
</tr>
<tr>
<td>Hispanic</td>
<td>272,490</td>
<td>229,825</td>
</tr>
<tr>
<td>Black</td>
<td>40,932</td>
<td>153,382</td>
</tr>
<tr>
<td>Asian/Pacific Islander</td>
<td>28,339</td>
<td>49,100</td>
</tr>
<tr>
<td>American Indian/AK Native</td>
<td>7,143</td>
<td>10,724</td>
</tr>
<tr>
<td>Two or More Races</td>
<td>26,073</td>
<td>24,908</td>
</tr>
<tr>
<td>Total Students</td>
<td>854,265</td>
<td>971,013</td>
</tr>
</tbody>
</table>

safe at school. Other factors associated with the achievement gap are high teacher
absence and turnover, high mobility of students, low birth weight, environmental
damage, poor nutrition, single-parent homes, and excessive TV watching, among others
(NAEP, 2011).

Access to literacy resources could also explain differences in interactions,
behaviors, and achievement for young children, challenging the assumption that all
children have equal access to literacy resources (Neuman & Celano, 2001). A 3-year
comparative study conducted in Philadelphia in two low-income and two middle-income
neighborhoods examined the role of community access to print in children’s development
of early literacy skills. Access to print was defined as reading resources for purchase,
quality of signs, public spaces for reading, and books in child care centers, school
libraries, and public libraries. This study acknowledged substantial differences in the
availability of print resources for children who live in low- or middle-income
communities.

**English Language Learners**

The National Literacy Panel defines this term as “students who come from
language backgrounds other than English and whose proficiency is not developed enough
to where they can profit fully from English only instruction” (Fien et al., 2011, p. 143).
ELL is also defined as “an active learner of the English language who may benefit from
various types of language support programs. This term is used mainly in the U.S. to
describe K–12 students” (National Council of Teachers of English, 2008, p. 2). English
language learners are a diverse group. The majority of ELLs in the United States speak
Spanish (Zehler et al., 2003), but there are differences within this group regarding country of origin, ethnicity, socioeconomic background, immigration status, generation educational background, and literacy in native language (August & Hakuta, 1997).

Klingner (2010) presents a distinction between learners who are sequential bilinguals and simultaneous bilinguals. Sequential bilinguals learn their native language (L1) first and later acquire a second language (L2). Simultaneous bilinguals acquire both languages at the same time. The distinction between a learning disability and language acquisition is more challenging to identify with simultaneous bilinguals.

The process of acquiring a second language has been extensively investigated by Cummins (1991). He identified two interrelated components of language proficiency: basic interpersonal communication skills (BICS) or conversational fluency and cognitive academic language proficiency (CALP) or “conceptual linguistic proficiency.” These components distinguish the different periods that it takes to develop conversational skills compared with grade-appropriate academic proficiency in that language. Gibbons (2006) used the terms playground language and classroom language to differentiate the everyday language and the language of schooling.

BICS, or conversational fluency, are the skills needed to function in everyday interpersonal contexts; it is often acquired to a functional level in two years (Cummins, 1991). For Gibbons (2006), playground language highly relies on visual and physical contexts (e.g. gestures and body language) and enables children to develop a social life and interact in different social situations such as making friends and playing games, assisting in language acquisition.
CALP proficiency is needed to function in academic settings including reading about a new subject, reading a lecture without visual cues, writing a report, and taking a standardized test. Empirical evidence demonstrates that ELLs take at least four years to develop English academic skills and this includes socioeconomically advantaged immigrant students (Cummins, 1997). According to Cummins (1997), there are two important dimensions that influence the second language (L2) acquisition process, the attribute-based and input-based. The attribute-based refers to the individual’s cognitive and personality variables including the foundation or cognitive resources ELLs bring from their first language. The input-based refers to the level of exposure to L2.

The language of schooling requires higher order thinking skills, such as hypothesizing, evaluating, inferring, generalizing, predicting, or classifying (Gibbons, 2006). Krashen and Lee Brown (2007) hypothesized that humans acquire language and develop literacy by understanding messages, not by intentionally learning about rules of grammar and vocabulary. In this case, the role of reading is a powerful form of comprehensible input for the development of academic language and content knowledge. These authors suggest that there are three important components in the area of CALP: (1) knowledge of academic language; (2) knowledge of specialized subject matter; and (3) strategies. Knowledge of academic language refers to the special language used in schools and the professional life whereas the knowledge of specialized subject matter deals with subject content such as algebra, history, or science (Krashen & Lee Brown, 2007). In order to improve comprehension of a text, Krashen and Lee Brown (2007) identified strategies such as “narrow reading” or reading about a single subject or the
same author and the use of background knowledge. Cummins (1997) emphasizes that the failure to consider this distinction between BICS and CALP results in discriminatory psychological assessments and inappropriate programming for ELLs.

**Reading Components**

The National Reading Panel (2000) evaluated different methods for teaching reading and concluded that alphabetics including phonemic awareness and phonics, reading fluency, vocabulary, and reading comprehension are critical components for teaching reading to young children as well as adolescents. As previously mentioned, this report did not address ELLs.

Phonemic awareness (PA) is defined as the ability to manipulate, blend, and segment sounds or phonemes in oral syllables and words. Unlike phonics instruction, PA does not rely on letter-sound relations when teaching students to read and spell. Correlational studies have demonstrated PA is a strong predictor of how well children learn to read in early years of instruction (National Reading Panel, 2000). The meta-analysis conducted by the National Reading Panel reported strong evidence across experimental studies that PA training significantly improves reading, phonemic awareness, and spelling (2000).

Another essential component for reading is phonics. Phonics instruction focuses on letter-sound correspondence and spelling patterns to teach students how to read and spell. Systematic approach presents phonics in a planned sequence and within an explicit phonics method and within the incidental approach the teacher addresses phonics when given the opportunity and as part of the text (National Reading Panel, 2000). The
National Reading Panel (2000) concluded that systematic phonics instruction significantly improves the ability to decode and spell in first grade students and older children. Across grades, good readers improved spelling with phonics instruction but these benefits were more substantial with younger students. The systematic phonics instruction also demonstrated benefits for low achieving students and students with learning disabilities (National Reading Panel, 2000).

The National Reading Panel referred to the definition by Durkin (1993). This author viewed comprehension as an active and intentional thinking process “during which meaning is constructed through interactions between text and reader” (National Reading Panel, p. 4-39). Besides being an interactive process, the National Reading Panel notes reading comprehension is a cognitive process that requires complex skills, involving the understanding of vocabulary (National Reading Panel, 2000). Cummins (n.d.) stated comprehension involves not only vocabulary or understanding the meaning of text but also how words are organized in sentences and paragraphs to produce meaning. The National Reading Panel (2000) concluded that comprehension improves when students relate print materials to prior experiences and knowledge and build mental representations. Studies show that using a combination of techniques such as comprehension monitoring, cooperative learning, use of graphic and semantic organizers, question answering, question generation, story structure, and summarization improves reading comprehension and yields to better results in standardized tests of reading comprehension.
Both reading comprehension and vocabulary involve the meaning of text at different levels. Vocabulary is tied to individual words and comprehension to larger units. There are two types of vocabulary: expressive and receptive. An individual for verbal and written communication relates expressive vocabulary to words produce. Receptive vocabulary is the words individuals recognize by listening and reading. The reading study by the National Reading Panel recognizes the importance of vocabulary for reading but suggested that vocabulary instruction does not lead to improvements in reading.

Reading fluency has to do with understanding and comprehension resulting in reading with appropriate expressiveness or decoding speed and accuracy (Samuels & Farstrup, 2006). The National Reading Panel defined reading fluency as the ability to read orally with “speed, accuracy, and proper expression” (p. 3-5) facilitating reading comprehension. The meta-analysis conducted by National Reading Panel reported that repeated oral reading guided by teachers, peers, or parents improves word recognition, fluency, and comprehension for good readers and for those with reading difficulties across grade levels and settings. There was no clear evidence of the effects of independent silent reading on reading fluency and other skills but suggesting that independent silent reading is not effective if used for students that have not developed basic reading skills (National Reading Panel, 2000).

Reading fluency has been identified as a critical component in reading instruction for elementary grade students (Rasinski, Homan, & Biggs, 2008). Research has found that measures of reading fluency including reading speed or measures of students’ prosodic oral reading were associated with reading comprehension and with reading
achievement in general. Students referred for reading support usually are struggling with fluency more than word recognition or comprehension (Rasinski & Paddack, 1998). For example, researchers have found that some students can decode words accurately, understand the meaning of these words, and are capable of listening and understanding, but their reading is “slow, unexpressive, and laborious” (Rasinski, Homan, & Biggs, 2008, p. 193). Along the same lines, Stanovich (1980) defined an interactive compensatory explanation of reading fluency. The main difference between a good and a poor reader is the way he or she processed text while reading. Poor readers had more difficulty using automatic attention-free, bottom-up processes for word decoding. Repeated readings helped readers to develop automaticity in word processing (Baker et al., 2008). The automaticity and efficient word recognition frees up resources that can be applied to comprehension (Baker et al., 2008).

Linan-Thompson, Cirino, & Vaughn (2007) indicate that expected growth and rates of progress vary for ELL students as for non-English Learners. Benchmarks and rates of progress also vary within the group of ELLs with different levels of proficiency in the second language (L2). Al Otaiba et al. (2009) examined a statewide database in Florida of high-poverty schools with 5,000 Latino students across the second and third grades. The purpose was to identify differences in proficiency levels and growth for oral reading fluency of Latino students who were not proficient in English and receiving English as a second language (ESL), proficient in English, and proficient enough to be exited from ESL services. Within each proficiency group, these authors examined the differences in fluency among subgroups of children in general education, students
identified with learning disabilities, and students with speech and language delays. This study demonstrated that throughout the second and third grades, oral reading fluency scores consistently distinguished students with learning disabilities from their general education peers regardless of English proficiency. All the participants received instruction only in English and attended at least two years in a public school in the United States. Latino students who never received ESL or special education services began second grade reading more fluently than any other group (53 words correct per minute). In third grade, the general education students started the school year reading 61 words per minute with a weekly growth of 1.31 words per minute. On average, Latino students receiving speech and language services started second grade at grade level and presented higher fluency scores than students with learning disabilities LD.

Latino students with LD showed the slowest rates of progress, starting at 0.61 words per minute to a weekly gain of 0.92 words per minute (in 2nd grade) and lowest oral reading fluency, 29 words per minute with growth rates of 1.5 (September-December) in third grade. Because students with learning disabilities demonstrate different growth trends compared with their peers, these findings suggest that oral reading fluency can be an effective way to screen students and measure effects of RTI to support schools in the process of identifying Latino students needing more intensive instruction (Al Otaiba et al., 2009).

The National Reading Panel (2000) did not address reading in English for ELLs but a synthesis conducted by the National Literacy Panel (2006) investigated the development of literacy in language-minority students. The Institute of Education
Sciences selected a National Literacy Panel of thirteen experts to synthesize quantitative and qualitative research with this population (August & Shanahan, 2006). According to this study, similar approaches to teach reading and writing, especially instruction that provides substantial exposure to the essential reading components including phonemic awareness, phonics, fluency, vocabulary, and reading comprehension, are effective with non-ELLS and ELLs but not sufficient with ELLs. High quality instruction for ELLs must include “early, ongoing and intensive” (August & Shanahan, p. 5) support of oral language development in English and vocabulary and background knowledge in English need to be addressed intensively with ELLs. Oral proficiency in English is related to reading comprehension and writing abilities, more specifically ability to define words in English, listening comprehension, and syntactic skills. Instructional approaches need adjustments to provide more benefits to ELLs (e.g., more work with specific phonemes in English that do not exist in the student’s home language).

The National Literacy Panel concluded that the development of literacy in English is influenced by different factors including age, language proficiency, cognitive skills, previous learning experiences, English oral proficiency, and differences between English and the first language. The type of instruction is another important factor. For instance, a significant number of studies suggested that students receiving bilingual instruction perform higher on measures of English reading proficiency than students instructed only in English. Another important factor is the positive influence of oral language and literacy in ELL’s first language such as higher order vocabulary can provide advantages to ELLs and facilitate development of speech discrimination and production, vocabulary,
and intraword segmentation in a second language (August & Shanahan, 2006.) Good literacy and oral language skills in L1 are advantageous and facilitate L2 skills.

In addition, multiple studies indicated language minority students classified with learning disabilities can perform at grade level with appropriate instruction, and there was a strong agreement on conducting assessment in ELLs' first language and English when examining eligibility for special education (August & Shanahan, 2006).

**Systems of Support for ELLs Students**

According to the National Institute of Child and Health and Human Development-Early Child Care Research Network (2003), reading difficulties are the core problem for the majority of ELL students receiving special education (Al Otaiba et al., 2009). Farver et al. (2009) identified phonological awareness, print knowledge (letter identification and understanding of basic print concepts), and oral language (vocabulary and grammar) as three key skills in the preschool period that are predictive of reading ability at school-age. These skills help children to read sooner and may prevent reading disabilities. They found evidence to support the importance of early interventions especially with ELL students that are learning to read in first language (L1), in both languages, or only in English. According to Farver et al. (2009), some studies favored English-only instruction (Baker & De Kanter, 1981; Rossell & Baker, 1996) and others favored bilingual instruction. Goldenberg (2008) stated, “teaching students to read in their first language promotes higher levels of reading achievement in English” (p. 14). Educational experts argue that bilingual instruction and dual-language immersion programs provide techniques that help ELLs both learn English and attain academic success by providing
instruction to these students in the language they understand the best (Wright, 2005).
After four to seven years in dual-language programs, bilingually instructed children
outperform their monolingual English-speaking peers in academic achievement across
subjects (Thomas & Collier, 2002).

Another program for ELL students is English as a Second Language (ESL). There
are two types of ESL programs, ESL pull out and ESL content in the mainstream. In the
ESL pull out programs students work with the ESL resource teacher on developing
listening and speaking skills in English; however, during the ESL class ELLs miss
instruction. These classes tend to mix students of different ages and proficiency levels in
English, and across various subjects (Thomas & Collier, 2002). The ESL content
programs integrate both content and language simultaneously to make lessons
comprehensible for ELLs. The teacher or language specialist uses visuals, contexts, and
modified texts to present concepts and skills for a specific subject (Thomas & Collier,
2002).

Farver et al. (2009) conducted an experimental study with 94 Spanish-speaking
preschoolers contrasting three groups, a control group with a high-scope curriculum (n =
32), an intervention group receiving an emergent literacy intervention (Literacy Express
Preschool Curriculum) in English only (n = 31), and another intervention group that
initially received the Literacy Express Preschool Curriculum in Spanish transitioning to
English (n = 31). Both the Preschool Comprehensive Test of Phonological and Print
Processing in English and Spanish were administered before and after the intervention.
Results from this study indicated that children in both the English-only group and the
transitional group obtained significantly higher English language assessment scores for Receptive Vocabulary, Definitional Vocabulary, Blending, Elision, and Print Knowledge than the participants in the control group. In addition, children in the transitional group performed better than children in the English-only group in the areas of Definitional Vocabulary and Print Knowledge in English and on the Spanish-language measures. These findings support the idea that small group interventions in traditional settings and in the first language (L1) can be an effective way of improving literacy skills with ELLs.

The synthesis of research conducted by Cheung and Slavin (2012) investigated effective reading programs for Spanish dominant ELL students. These researchers reviewed twenty-two qualifying studies, from 1980 to 2010, and classified the interventions into two main categories: whole-school or whole-class program or small group or one-to-one supplemental intervention. Based on this synthesis the most favorable programs for Spanish-speaking students were: Success for All with specific adaptations for English language development (ES = .35); two types of cooperative learning, Bilingual Cooperative Integrated Reading Composition (ES = .54) and Peer Assisted Learning Strategy (ES = .36); and Direct Instruction (ES = .28). This study concluded that the most effective interventions provide substantial professional development and coaching for teachers and cooperative learning, which provides opportunities for ELL students to practice English in a meaningful context.

Another system of support for ELL students is the RTI model which was introduced as part of the reauthorization of IDEA in 2004 as an alternative to identify a learning disability and states “a local education agency may use a process that determines
if the child responds to scientific, research based interventions as part of the evaluation procedures” (IDEA, 2004, 614 (b), p. 6). As previously mentioned, RTI is a proactive process that provides evidence-based instruction to all students with additional intensive and individualized interventions to prevent student underachievement, including culturally and linguistically diverse students (Vellutino et al., 1996; Donovan & Cross, 2002; Francis et al., 2006; Vellutino et al., 2006; Al Otaiba et al., 2009;).

RTI is a multi-tiered model of prevention and intervention with graduated levels of support (Stacker, 2007). Interventions at each level, or tier, should be based on scientific evidence of effectiveness (Klingner & Edwards, 2006), with the main difference between the tiers being “intervention intensity and measurement precision” (Reschly, 2005, p. 511). Tier I of RTI encompasses universal screening, classroom based-instruction, and assessment in the general education classroom with all students (Vaughn & Roberts, 2007; Hazelkorn et al., 2011). Tier I includes scientifically based reading instruction and curriculum with emphasis on the essential reading components (phonemic awareness, phonics, reading fluency, reading comprehension, and vocabulary) and benchmark assessments three times per year (Vaughn & Roberts, 2007).

Tier II focuses on specialized or targeted interventions for students who are not making adequate progress in the core program or in Tier I (Hazelkorn et al., 2011). The students at-risk receive targeted instruction to help close the gap between their current performance and their expected performance. The specialized, scientifically based instruction can be 20-30 minutes in addition to Tier I. The progress monitoring or assessments occur twice a month to guarantee optimal progress and learning. Tier III is
considered to be the most sustained and intensive of all the levels and is focused on individual student need. Tier III provides intensive scientifically based instruction to students with significant difficulty in reading that did not respond sufficiently to Tier I and Tier II. The small group is provided for 50 minutes per session. Progress monitoring occurs at least twice per month (Vaughn & Roberts, 2007).

Fletcher and Vaughn (2009) pointed out an increase in overall academic achievement scores and reduction of special education referrals by districts with successful implementation of RTI models. These authors recommended examining the outcomes in relation to historical data so it is clear that RTI models support students who are at-risk of academic difficulties. Other benefits with the implementation of RTI include a significant decrease in the placement rates of minority students in special education (Batsche, Kavale, & Kovaleski, 2006) and a significant increase in the rate of response of minority students to early, intensive instruction (VanDerHeyden, Witt, & Gilbertson, 2005).

The results from different studies of evidence-based interventions and RTI with ELLs show some advantages of RTI with ELLs but also reveal some limitations of this model with this population. Han (2009) conducted a meta-analysis of evidence-based reading instruction for ELLs from pre-school through sixth grade. This study included 29 studies from peer-reviewed journal from 1967 through 2009. Dissertations, reports and conference presentations were not included. The overall effect of reading instruction was moderate (ES 0.50). Keyword method, proactive reading, and peer-assisted learning strategies were identified as promising practices. This study identified more than 10
programs that address phonemic awareness and phonics instructions for pre-school
through second grade at both Tier I and Tier II but indicated there are limited vocabulary
instructional programs available for ELLs at Tier I. The results of this study indicate the
correlation between quality and effect size was not statistically significant. The direction
of the correlation was negative suggesting a significant decrease of the mean of the effect
sizes with the increase of quality of studies or efforts to maintain the rigor of research
design. The author used two data sets, one data set comparing ELL treatment groups to
ELL control groups and another data set comparing ELLs to L1 students or at-risk ELLs
to not-at-risk ELLs. The intercept was 0.50 (t=7.15, p < .01) and the effect of reading
instruction for ELLs was moderate (ES = 0.50) with the first data set with 35 samples and
178 effect sizes. The overall mean effect of instructional programs was 0.07 suggesting
the programs did not produce significant different effects for ELLs or at-risk students
compared to for non-ELLs and not at risk students.

Orosco and Klingner (2010) conducted a study to evaluate the implementation of
RTI in an urban elementary school with a large percentage of Latino ELLs struggling in
reading. Based on their findings, ELLs have appeared to have different learning needs
than non-ELLs. Orosco and Klingner indicated that districts need policies based on socio-
culturally guided assessment and instruction: teachers need interventions validated with
English language learners and interventions that have empirical evidence of effectiveness
with ELLs and educators that work with ELLs need preparation to understand the
language acquisition process, bilingual education, and teaching methods for English as a
second language.
Klingner, Artiles, and Méndez Barletta (2006) conducted a synthesis to investigate the difference between ELLs with a learning disability and students who struggle with literacy due to limited proficiency in English. This synthesis included studies about ELLs with learning disabilities, kindergarten through 12, and ELLs struggling with reading. These authors concluded more research is needed to identify the learning needs of underachieving ELLs. One of the problems with the identification of learning disabilities with ELL is the focus on finding a deficit within the student instead of evaluating the context and instructional factors. The authors identified the following factors that support a successful RTI model with ELLs: a learning environment where literacy is considered a sociocultural practice (Artiles 2002), where cultural and linguistic diversity are valued, (Ortiz, 1997, 2002; Nieto, 2004; Baca, 2012), and where teachers know instructional practices that are tailored for ELLs.

The eligibility of students for special education under the category specific learning disability (SLD) has been one of the most controversial changes to the Individuals with Disabilities Education Improvement Act (Batsche, Kavale & Kovaleski, 2006). Documenting if a child is responding to a scientific research-based intervention allows the identification of a specific learning disability. IDEA (2004) states that “the Local Education Agencies shall not take into consideration whether a child has a severe discrepancy between achievement and intellectual ability in oral expression, listening, comprehension, reading” (Pub. L. No. 108–446 § 614 [b][6][A]).

One of the most important tools in identifying specific learning disabilities is called curriculum-based measurement (CBM). Deno (1985) developed this measurement
system to help special educators monitor students’ progress in basic skills and improve quality of instruction. Fuchs, Fuchs, and Stecker (2005) define CBM as a type of progress monitoring that is scientifically validated. This classroom-based assessment is used to evaluate academic competence (e.g. reading, math, and spelling), track academic development, and enhance academic achievement. The National Center on Student Progress Monitoring (2011) reports that through this scientifically based practice, an assessment of the student’s performance and effectiveness of the intervention can be conducted. As well, Howell and Shinn (2002) emphasize that CBM can be useful for educators to make decisions about students’ instructional needs. These authors describe four important characteristics of CBM: students are monitored on ongoing basis, tests are typically short, tests measure a key skill, and tests use passages of similar difficulty.

CBM assumes student progress can be monitored to aid in determining the quality and intensity of instruction, all other things being equal. Dominguez de Ramírez and Shapiro (2006) suggested that CBM is an effective tool to assess ELL literacy skills in both native and second language instruction. A study with 165 students across grades 1-5 in bilingual (N = 68) and general education classrooms (N = 97) suggested that growth rates may not be equivalent across general education students and Spanish-speaking students. Oral reading fluency was assessed in English and Spanish with a CBM three times a year. Significant main effects were reported for time and group. When comparing general and Spanish-speaking ELLs for reading passages in English, all students demonstrated significant growth between October and May in English oral reading fluency, and general regular education students read more fluently than Spanish-speaking
students. A significant interaction between group and time suggested that general education students presented greater growth in reading fluency than Spanish-speaking students. A significant interaction between time and group was also significant when comparing general education students’ reading in English and the Spanish-speaking students reading in Spanish. General education students made more substantial progress in English than Spanish-speaking students did in Spanish (Domínguez de Ramírez and Shapiro, 2006).
CHAPTER THREE

Methods

The present study explored research on reading for ELL students to provide a better understanding of the implementation and effectiveness of the RTI model with ELLs and reveal implications for policy and school-based leadership. Therefore, the current study searched, classified, and analyzed the existing research on reading interventions with ELLs within the general population of reading interventions. This study classified existing studies into three different tiers of intervention, Tier I, Tier II, and Tier III and examined the effects of reading interventions with ELLs.

Meta-analysis was used to aggregate and compare findings of research studies (Lipsey & Wilson, 2001) and the effects on reading achievement of different research-based reading interventions. Meta-analysis, as a research technique, allows the researcher to estimate the effect size for each study and combine those estimates across studies, yielding stronger statistical power than individual studies (Lipsey & Wilson, 2001).

Purpose of the Study

The purpose of the study was to: (a) determine whether there is empirical support for research-based reading interventions that produce improvement in reading for ELLs, and (b) estimate the strength of this relationship.
Research Questions

This study addresses the following research questions:

1. Is there empirical support for effects of Tier I interventions on reading for ELL students?
2. Is there empirical support for effects of Tier II interventions on reading for ELL students?
3. Is there empirical support for effects of Tier III interventions on reading for ELL students?

Meta-analysis

Meta-analysis is a methodological and statistical approach that allows the researcher to formulate inferences about a larger population of studies by comparing and systematically synthesizing results from a sample of empirical studies in which individual studies are the unit of analysis (Card, 2012). The researcher can only include studies that meet certain pre-specified criteria (Borenstein et al., 2009).

Meta-analysis, as a method to “summarize, integrate, and interpret sets of scholarly works” (Lipsey & Wilson, 2001, p. 2), can only include empirical research studies with quantitative findings that present descriptive and inferential statistics (Lipsey & Wilson, 2001). Meta-analysis uses an effect size to standardize findings from the unit of analysis allowing the researcher to aggregate results so each study contributes to the overall mean effect size (Lipsey & Wilson, 2001).
Criteria for Selection of Studies

The criteria for selection of studies, or inclusion and exclusion criteria, in a meta-analysis uses specific characteristics of the population of research studies whose findings are to be examined and summarized (Lipsey & Wilson, 2001). The inclusion and exclusion criteria provides information regarding the features of studies that will be included or rejected in the meta-analysis, and allows the researcher to define the population of studies that will be used to drawn conclusions about research (Card, 2012).

The following categories were considered when developing the inclusion and exclusion criteria: “(a) distinguishing features of the study (b) research respondents (c) key variables (d) research methods (e) cultural and linguistic range (f) time frame (g) publication type” (Lipsey & Wilson, 2001, p. 16-17).

**Distinguishing features of the study.** Eligible studies must involve the use of research-based interventions to improve reading. Research-based interventions or evidence-based interventions are a core component of RTI. Research-based intervention is defined as an intervention for which data from scientific, rigorous research designs have demonstrated (or empirically validated) the efficacy of the intervention. That is, within the context of a group or single-subject experiment or a quasi-experimental study, the intervention is shown to improve the results for students who receive the intervention. (NCRTI, 2014, p. 6)

This study includes interventions implemented as part of the RTI model and classifies studies into Tier I, Tier II and Tier III. Tier I includes interventions that are implemented as part of the core curriculum with all students in a regular education setting. Tier II focuses on specialized or targeted interventions for students who are not making adequate
progress in the core program or in Tier I (Hazelkorn et al., 2011). Tier III “is considered to be the most sustained and intensive of all the levels and is focused on individual student need” (Stecker, 2007, p. 52).

**Research respondents.** Eligible studies must quantitatively examine the effects of research-based reading interventions for ELLs from kindergarten through 8th grade attending public schools in English speaking countries that are implementing the RTI model. The National Literacy Panel (2006) defines ELLs as “students who come from language backgrounds other than English and whose proficiency is not developed enough to where they can profit fully from English only instruction” (Fien et al., 2011, p. 143).

**Key variables.** Studies must report results from at least one quantitative test of reading and must include assessment of phonemic awareness, phonics, reading fluency, vocabulary, and reading comprehension. Studies that measure other components of reading may be included but “only studies from which an effect size can be computed are eligible” (Lipsey & Wilson, 2001, p. 21).

**Research methods.** Empirical research studies with quantitative findings that present descriptive and inferential statistics are eligible (Lipsey & Wilson, 2001). Because this method focuses on the aggregation and comparison of the findings, results of different studies need to present similar constructs and relationships and similar statistical forms of analyses (Lipsey & Wilson, 2001). “Experimental and quasi-experimental studies with statistical data including means and standard deviations, or significance test results necessary to calculate an estimated effect size of the impact” (Vanchu-Orosco, 2012, p. 88) of the research-based interventions under study were
included. This study examined studies that establish comparisons between treatment and control conditions. For example, studies that compare an RTI versus a “business as usual” intervention were included. Studies that calculate changes in scores from pre-interventions to post-intervention will be excluded.

**Cultural and linguistic range.** Studies must be conducted in English in English-speaking countries.

**Time frame.** The RTI model was introduced as part of the reauthorization of IDEA in 2004 as an alternative to identify a learning disability (IDEA, 2004, 614 (b) p. 6). Therefore, the time frame for the search was 2004 to 2015.

**Publication type.** Published and high-quality studies that are unpublished are eligible. This includes peer reviewed articles, non-refereed journals, and dissertations from institutions that are classified as doctoral granting. This study also included papers and proceedings from conferences and meetings that have a peer-review process and professional associations (for example, American Education Research Association (AERA) and National Association of School Psychologists (NASP).

Including studies that are not published is critical to control publication bias. The problem with publication bias is that the published literature may not be representative of the studies conducted on a topic and can produce a stronger overall effect size than if all studies were included as part of the meta-analysis (Card, 2012).

**Exclusion criteria.** Based on the following criteria some studies were excluded for the current meta-analysis:
• Studies that did not include means and standard deviations or p-values or data necessary to calculate an effect size.
• Studies that examined research-based interventions for other academic areas such as writing and mathematics.
• Qualitative studies.
• Studies published prior to implementation of RTI.

Finding Relevant Literature

This meta-analysis included both peer-reviewed articles and articles that are not peer reviewed but are from journals that have a strong reputation and editorial review. Additionally, previously mentioned, this study included dissertations from institutions that are classified as doctoral granting; papers and proceedings from conferences and meetings that have a peer-review process.

This study involved a computerized database search to find candidate studies. As recommended by Lipsey and Wilson (2001) and Vanchu-Orosco (2012) to identify a high number of potentially eligible studies for a meta-analysis, the search should be based on a set of keywords that broadly cover the topic under investigation. The researcher can identify these keywords by finding descriptors in a database related to the topic of interest and by reviewing the different terms authors use in titles and abstracts of studies in the area of interest (Lipsey & Wilson, 2001). The search criteria for the current meta-analysis included the following key words in various combinations: “English language learners,” “second language learners,” “English as a second language,” “multilingual learners,” “interventions,” “reading,” “literacy,” “response to intervention,” “Tier I,” “Tier II,”
“Tier III,” “phonological awareness,” “phonics,” “reading fluency,” “reading comprehension,” and “vocabulary.”

The search for potential eligible studies was conducted using different electronic databases including: Academic Search Complete, Education Resources Information Center (ERIC), PsycINFO, PsycArticles, ProQuest dissertations & theses, and Google Scholar. The author conducted a comprehensive search in these databases yielding 130 studies including dissertations and peer-reviewed articles. To organize and group studies by tiers of interventions, reading components, and meta-analyses/syntheses, the author used Ref Works. The author located an additional six potential eligible studies after reviewing conferences including the AERA annual meeting, NASP annual convention, and Society for Research on Educational Effectiveness. The reference lists for different meta-analyses (Klingner, Artiles, & Méndez Barletta, 2006; Hans, 2009; Bagasi, 2014) and a synthesis (Cheung & Slavin, 2012) were also reviewed. After reviewing the abstracts and methods, 63 studies were retained for further examination. The author reviewed the method and results sections for each of the retained studies and eliminated 43 studies. Some studies addressed research-based intervention with ELLs and used reliable outcome measures but did not mention RTI, others implemented the interventions in Spanish or English and Spanish and measured outcomes in both languages, others used reliable reading measures but were not specific about the intervention, and others did not provided appropriate statistical information to calculate effect sizes. Therefore, the author decided to retain pre-test-post-test studies with an experimental group but no comparison group (Healy, Vanderwood, & Edelston, 2005; Miller, 2013; Richards-Tutor et al., 2012)
and a study with a multi-baseline design (Gyovai, 2009) with an active independent variable but no random assignment. Twenty studies (with at least twenty-seven possible effect sizes) were retained for this study including peer reviewed articles (11), dissertations (8), and a paper proceeding from a conference (1).

**Coding Study Characteristics and Empirical Findings**

The coding process in a meta-analysis is used to determine what relevant information needs to be extracted from each study in order to develop a database for statistical analysis (Lipsey & Wilson, 2001). This study encoded two different types of information: one based on the study characteristics or descriptors and the other based on the empirical findings of the study (Lipsey & Wilson, 2001). Coding was guided by the research questions but also included specific aspects of studies that need to be considered such as “characteristics of the sample, measurement, design, and source” (Card, 2012, p. 65). Based on recommendations from Lipsey and Wilson (2001) and Card (2012) the coding for this study included source characteristics, sample characteristics, measurement characteristics, and design characteristics.

Source characteristics included the number of the study or ID number, author, title, year of the study, and publication type (e.g. journal article, organization report, dissertation, and conference proceeding paper). Lipsey and Wilson (2001) recommend coding papers with more than one study separately by adding a decimal to the identification number. For this meta-analysis, the studies with additional independent substudies are coded separately. The author added a decimal to the study ID (e.g. 01.1 and 01.2 is study 01 with two independent substudies).
Study retrieval, specifies the method to retrieve a study including electronic database, organizational web site search, bibliographic reference, and synthesis/meta-analysis was coded (Vanchu-Orosco, personal communication, July 2015).

Sample characteristics or demographic information coded included participant sampling method, ethnic composition of the sample, number of student receiving free-reduced lunch, gender, number of ELLs and non-ELLs, predominant language of ELLs, levels of English language proficiency (if reported), participants grade levels, mean age or age range. This allowed the researcher to analyze the effect sizes by different subgroups and to examine effects on respondents with different characteristics (Lipsey & Wilson, 2001). Sampling method included population, simple random selection, stratified random selection, systematic selection, or available (convenience sample) (Lipsey & Wilson, 2001; Vanchu-Orosco, 2012, p. 323).

Measurement characteristics included the name of the assessment (including author and version), type of scale used to measure the outcome (standardized or developed by researcher), constructs, reliability, and validity type (Vanchu-Orosco, 2012). If the study uses more than one assessment, Lipsey and Wilson (2001) suggest coding each measure separately to allow for a more comprehensive “empirical examination of the relationship between the particular ways in which a construct is operationalized and the nature of findings from different studies” (Lipsey & Wilson, 2001, p. 78). Therefore, if the study presented results using different measures, the author added a decimal to specify the measure and subtest (01.0.1.12 represented study 01, 0 =
no breakout for grade level, 1 = CTOPP, and 12= Blending Words subtest). The Codebook contains specific codes and examples (See Appendix B).

Design characteristics identified the study type (e.g., post-hoc, experimental, quasi-experimental), research design/approach (e.g., comparison, repeated measures, independent groups, and others), and the statistical method (e.g., descriptive statistics, t-test, F-test, chi-square, ANOVA, ANCOVA, multiple regressions, and among others (Vanchu-Orosco, 2012).

The second part of the coding process focused on the treatment effects of the interventions with reading ability as the main outcome to be examined in this study. This study classified reading interventions by tiers of intervention—Tier I, Tier II, and Tier III—and the researcher sorted them by reading components including phonemic awareness, phonics, fluency, vocabulary, comprehension, and oral language. It is important to note not every research study identifies what they are investigating by these names, even though they are studying the concept(s) and thus were classified with others of the same type.

The treatment and control groups’ sample size, mean, standard deviation, and effect size was coded to analyze difference across groups (Lipsey & Wilson, 2001). Coding the effects of the treatment allowed the researcher to pull study sub-group to calculate effect sizes and to assemble findings across the tiers of interventions. A proxy for study quality was included. The quality of journals, conference papers, and dissertations are considered equivalent for this study (Vanchu-Orosco, 2012). Journal articles and conference proceedings are peer reviewed and committee members review
unpublished dissertations. Published research reports, which may or may not have a peer review process, are classified as being of “lesser quality” (Vanchu-Orosco, 2012, p.106).

As recommended by Vanchu-Orosco (2012), the following steps were included to code and classify studies:

(a) create the codebook with initial set of codes (b) reading five articles with the initial codebook and revising as new information came to light; (c) coding three or more articles with the revised codebook and revising again; (d) create coding forms and coding manual to accompany the codebook (e) coding all remaining studies. (p. 105)

For the present study, a second coder coded 29.63% of random eligible studies using the codebook and coding forms. Information between coder was used to calculate the inter-rater reliability. The second coder was a researcher familiar with meta-analysis and present study. The code-book was reviewed with the second coder with additional coding materials. Additional coding materials included a table with measures and components (Appendix C) and the preliminary coding form (Appendix D). The table with measures and components (Appendix C) presents the reading measures with composites and subtests, code for each measure, and reading components assess by each measure and subtest. The preliminary coding form was created during the search process and lists how the study was retrieved, tier of intervention, grade level, and languages.

The percentage of agreement, for a random sample of 29.63% of all studies, was 91.46%. For disagreement between the author and second coder, the rationale for the difference was discussed and consensus on coding was reached. After conducting the inter-rater reliability, the author completed a second review of the coding with the revised code book and made necessary changes before entering data into Comprehensive Meta-
Analysis (CMA) Version 3.3.070 (Borenstein et al., 2014). CMA is a software specifically designed for meta-analysis (Borenstein et al., 2009).

**Statistical Analysis**

The effect size for a meta-analysis allows the researcher to estimate the strength of a relationship between the independent variable and the dependent variable(s) (Borenstein et al., 2009; Gliner, Morgan, & Leech, 2009). Effect size is defined as an index of the direction and magnitude of association between two variables and may include a correlation between two variables, differences between two groups, and contingencies between two dichotomies. An important criterion for effect size is that it must be possible to compute or approximate its standard error. The standard error allows a researcher to give more weight to studies that have small standard errors than those that provide less precise estimates. (Card, 2012, p. 87)

The summary statistic for calculating effect sizes for the studies chosen was the standardized mean difference effect size that involves a group contrast on measures that have a continuous outcome construct (Lipsey & Wilson, 2001; Wilson, 2011) in this case reading achievement. The standardized mean difference applies to “comparisons between means of outcome measures for experimental and control groups in treatment effectiveness research” (Lipsey & Wilson, 2001, p. 48). The standardized mean difference effect size is the “difference between the group means divided by the pooled standard deviation” (Lipsey & Wilson, 2001, p. 172) where d is the difference score, mean1 is the mean of the treatment group, and mean2 is the mean of the control group (equation 3.1), and $S_{pooled}$ is the pooled standard deviation (equation 3.2).
\[ d = \frac{\bar{x}_1 - \bar{x}_2}{s_{pooled}} \]  

\[ s_{pooled} = \sqrt{\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1 + n_2 - 2}} \]  

(3.1)  

(3.2)

“The sample estimate of the standardized mean difference is often called Cohen’s d in research synthesis. The symbol \( \delta \) denotes the effect size parameter and \( d \) for the sample estimate of that parameter” (Borenstein et al., 2009, p. 27). To calculate the effect size for a meta-analysis the studies should have numerical values that are comparable, must be able to compute its effect size standard error (Wilson, 2011). It is important to note when the sample is small d can overestimate the value of \( \delta \), the population parameter (equation 3.3). This bias will be fixed with the unbiased estimate converting d to Hedges’ g using J, a correction factor (equation 3.4) (Borenstein et al., 2009).

\[ \delta = \frac{\mu_1 - \mu_2}{\sigma} \]  

\[ J = 1 - \frac{3}{4 df - 1} \]  

(3.3)  

(3.4)

Effect sizes based on the standardized difference between means formed the basis of the analysis. For primary studies Hedge’s g, an unbiased estimator of \( \delta \), the standardized mean difference, based on Cohen’s d, will be used to calculate the effect size for differences between means (Vanchu-Orosco, 2012). The effect size was interpreted as ES < 0.20 small, ES = 0.50 medium, and ES > 0.80 large (Cohen, 1992).

Calculating independent effect sizes will include the following steps: “estimating the mean effect size, tests of significance for the test statistics and the size of the effect, and estimating and testing the variation between the units of analysis” (Vanchu-Orosco,
2012). If the means or standard deviations were not available, the effect sizes were calculated from reported statistics including tests of significance or t-test. If the means were not available, difference in gain score treatment dummy can be used to estimate the means (Lipsey & Wilson, 2001; Wilson, 2011). For example, the gain score, posttest value minus pre-test value of the same measure for a group, was used if the study reports the mean gain for each group and the pooled standard deviation for the posttest score is reported or can be estimated.

Two types of analyses, mean gain and mean difference, were conducted for Tier I and Tier II studies to examine the empirical evidence of reading interventions for ELLs. The mean difference analysis compared post-test data from treatment and control groups. The mean gain analysis grouped pre-post studies by treatment and control groups to compare performance of ELL participants receiving reading interventions (treatment groups) to ELLs that were not exposed to the intervention under research (control groups).

If the standard deviation, natural variability within the group on a measure (Wilson, 2011), was not reported different methods were used such using the standard error or other statistics available. The practical meta-analysis effect size calculator created by Lipsey and Wilson (2001) was helpful to calculate effect sizes when means and standard deviations were not reported. CMA (v 3.3.070, 2014) was used to compute meta-analytic statistics to answer research questions.

The present study did not include the source of heterogeneity through a moderator analysis, which allows the researcher to examine if the effect sizes vary based on the
level of the moderator (Card, 2012). Some of the moderators that may explain the
remaining variance (heterogeneity among effect sizes) includes different levels of English
language proficiency of the participants, educational experience in the US, years of
experience of the personnel providing the intervention and knowledge about ELLs, length
of the intervention, English language development services, support participants receive
at home, and exposure to literacy after school. Other variables that can explain the
variance are different levels of intellectual functioning of participants, academic abilities
in native and second language, and socio-emotional factors.

One of the limitations of meta-analysis is publication bias. This study included
published and unpublished studies such as dissertations and a conference paper to obtain
a better estimate of the true effect size of the target population of studies; however, the
file drawer problem or unpublished research with lower treatment effects is still problem
for any type of literature review (Borenstein et al., 2009). This study employed two
methods for addressing bias, the funnel plots and the Trim and Fill method. The funnel
plots suggested the effects of the meta-analyses with Tier I and Tier II studies were
symmetrically distributed indicating there was no indication of publication bias. The
Trim and Fill method also suggested no problems with publication bias (Appendix N).

The problem with dependency was another shortcoming of this study. The
majority of primary studies used multiple outcome measures for the same intervention
with the same sample. Other studies had independent samples or sub-studies; however,
the same researcher conducted the studies creating problems with dependency.
CHAPTER FOUR

Results

The purpose of this study was to: (a) determine whether there is empirical support for research-based reading interventions that produce improvement in reading for ELLs, and (b) estimate the strength of this relationship. This study addressed the following research questions:

1. Is there empirical support for effects of Tier I interventions on reading for ELL students?
2. Is there empirical support for effects of Tier II interventions on reading for ELL students?
3. Is there empirical support for effects of Tier III interventions on reading for ELL students?

Twenty-seven studies (20 articles) quantitatively examined the effects of Tier I, Tier II and Tier III research-based reading interventions for ELL students from kindergarten through 8th grade. Eversole (2010), Kamps et al. (2007), McIntosh, Graves, and Gersten (2007), and Ransford-Kaldon, Sutton Flynt, and Ross (2011) provided information to calculate effect sizes for independent sub-studies. The Eversole (2010) study contained four different studies including second, third, fourth, and fifth grade, the Kamps et al. (2007) study presented results for two independent studies for first and
second grade, McIntosh, Graves, and Gersten (2007) research contained information for
two studies, year one and year two, with two independent samples, and Ransford-Kaldon
et al. (2011) work contained three different studies including kindergarten, first, and
second grade.

Study Characteristics

Publications from 2005 through 2013 were included in this meta-analysis. Eight
studies published from 2005-2009 and twelve studies from 2010-2013. Five studies were
retrieved from the Academic Search Complete database, five from PsycINFO, two from
ERIC, two from Proquest Dissertations and Theses. Other studies were first found in
organizations’ websites (AERA, NASP, SREE) and a study bibliography. For example,
the McIntosh et al. (2007) study was initially found in the references of an article
retrieved from the NASP website. The full documents were retrieved from ERIC or
Google Scholar (Table 4).

Type of studies included was quasi-experimental (50%), experimental (30%), and
post-hoc (20%). Both experimental and quasi-experimental studies have an active
independent variable but quasi-experimental studies do not utilize random assignment of
participants to groups (Gliner, Morgan, & Leech, 2009). Experimental research included
studies with experimental-control group comparison and randomized name/student
number selection procedure (See Table 4 for study number 16); randomized experimental
design with matching (studies 10 & 17).

Quasi-experimental included pretest-posttest studies with a control group design
with matched samples (study 12) but no random assignment; studies with an
experimental-control group comparison but no random assignment or other type of assignment (studies 1, 6, 15, & 11); pre-test-post-test designs with an experimental group but no comparison or control group (studies 5, 13, &18) and a multi-baseline study (study 3), with an active independent variable but no random assignment (Gliner, Morgan & Leech, 2009). Other types of studies are comparative or post-hoc studies that used archival data to allow the comparison of groups (studies 2 & 9). Table 4 provides information regarding the year, publication type, study retrieval, and type of study for each study included in this meta-analysis.
<table>
<thead>
<tr>
<th>Study No.</th>
<th>Authors</th>
<th>Pub year</th>
<th>Pub Type</th>
<th>Retrieval</th>
<th>Type of Study</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Dougherty Stahl et al.</td>
<td>2012</td>
<td>Journal</td>
<td>Academic</td>
<td>Quasi</td>
</tr>
<tr>
<td>2</td>
<td>Eversole</td>
<td>2010</td>
<td>Dissertation</td>
<td>PsycINFO</td>
<td>Post-Hoc</td>
</tr>
<tr>
<td>3</td>
<td>Gyovai et al.</td>
<td>2009</td>
<td>Journal</td>
<td>Academic</td>
<td>Quasi</td>
</tr>
<tr>
<td>4</td>
<td>Graves et al.</td>
<td>2011</td>
<td>Journal</td>
<td>Academic</td>
<td>Experimental</td>
</tr>
<tr>
<td>5</td>
<td>Healy et al.</td>
<td>2005</td>
<td>Journal</td>
<td>Academic</td>
<td>Quasi</td>
</tr>
<tr>
<td>6</td>
<td>Kamps et al.</td>
<td>2007</td>
<td>Journal</td>
<td>Academic</td>
<td>Quasi</td>
</tr>
<tr>
<td>7</td>
<td>Keita</td>
<td>2011</td>
<td>Dissertation</td>
<td>ERIC</td>
<td>Post-Hoc</td>
</tr>
<tr>
<td>8</td>
<td>Kourea</td>
<td>2007</td>
<td>Dissertation</td>
<td>Proquest D</td>
<td>Quasi</td>
</tr>
<tr>
<td>9</td>
<td>Linan-Thompson et al.</td>
<td>2007</td>
<td>Journal</td>
<td>PsycINFO</td>
<td>Post-Hoc</td>
</tr>
<tr>
<td>10</td>
<td>Lovett et al.</td>
<td>2008</td>
<td>Journal</td>
<td>ERIC</td>
<td>Experimental</td>
</tr>
<tr>
<td>11</td>
<td>McIntosh et al.</td>
<td>2007</td>
<td>Journal</td>
<td>Bibliography*</td>
<td>Quasi</td>
</tr>
<tr>
<td>12</td>
<td>McMaster et al.</td>
<td>2008</td>
<td>Journal</td>
<td>ERIC</td>
<td>Quasi</td>
</tr>
<tr>
<td>13</td>
<td>Miller</td>
<td>2013</td>
<td>Dissertation</td>
<td>PsycINFO</td>
<td>Quasi</td>
</tr>
<tr>
<td>14</td>
<td>Nguyen-Quang</td>
<td>2012</td>
<td>Dissertation</td>
<td>AERA*</td>
<td>Experimental</td>
</tr>
<tr>
<td>15</td>
<td>O'Connor et al.</td>
<td>2014</td>
<td>Journal</td>
<td>PsycINFO</td>
<td>Experimental</td>
</tr>
<tr>
<td>16</td>
<td>Pieretti</td>
<td>2011</td>
<td>Dissertation</td>
<td>Proquest</td>
<td>Experimental</td>
</tr>
<tr>
<td>17</td>
<td>Ransford-Kaldon et al.</td>
<td>2011</td>
<td>Conference</td>
<td>SREE*</td>
<td>Experimental</td>
</tr>
<tr>
<td>18</td>
<td>Richards-Tutor et al.</td>
<td>2012</td>
<td>Journal</td>
<td>ERIC</td>
<td>Quasi</td>
</tr>
<tr>
<td>19</td>
<td>Sapienza</td>
<td>2012</td>
<td>Dissertation</td>
<td>PsycINFO</td>
<td>Quasi</td>
</tr>
<tr>
<td>20</td>
<td>Soong</td>
<td>2012</td>
<td>Dissertation</td>
<td>NASP*</td>
<td>Post-Hoc</td>
</tr>
</tbody>
</table>

Table 4. Study Retrieval and Type of Study

Participants in these studies attended public schools in English speaking countries including the United States (19 studies) and Canada (1 study). The studies in the United States were conducted in different states and regions: California, Florida, Georgia, Minnesota, New York, and Texas, as well as the Midwestern region. The majority of studies focused on lower/early elementary grades. Sixty-nine percent of the studies included students in grades kindergarten through 2nd; 22% grades 3rd through 5th grade; and 9% upper grades 6-8. The predominant native language was not identified in 20% of

---

2 Academic = Academic Search Complete; ERIC = Education Resources Information Center; AERA = American Education Research Association; SREE = Society for Research on Educational Effectiveness; NASP = National Association of School Psychologists. Quasi = Quasi-experimental
the studies; the most frequent native language was Spanish (65%), and the other identified languages were Somali (10%); Hmong (5%), Portuguese (5%). The most frequent ethnicity was coded as mixed (50% of the studies) suggesting that 60% of the participants or more were from different ethnic backgrounds including African-American, White, Hispanic, Asian, Indian, Pacific Islander, Somalian, and Multiracial. Hispanic was the most predominant ethnicity in 40% of the studies, and Hmong in 10% of the studies. Sixteen of the studies had a student population with over 80% free or reduced lunch, one study reported 50-80% free or reduced lunch and three studies did not report this information. In summary, the majority of the participants in these studies including the sub-studies attended schools in the United States, spoke Spanish, received free or reduced lunch, and was from different ethnic backgrounds. Some authors specified Spanish as the most frequent language; however, they did not specify the frequency for each ethnicity.

Table 5 presents information regarding the sample size for each study, number of ELLs, the most frequent language, ethnicity, grade level, percentage of students receiving free or reduced school lunch, and states or regions were participants attended school and where research took place. Ethnicity indicates the most frequent (greater than 60%) ethnicity in the sample. Mixed suggests the participants were from different ethnic backgrounds including African-American, White, Hispanic, Asian, Indian, Pacific Islander, Somalian, and Multiracial.
Studies that examined the effects of Tier I, II and III interventions used three different types of design: post-test data, post-test data studies comparing ELL versus non-ELL students, and pre-test post-test studies. Tables 6, 7, 8, and 9 provide information about type of intervention, outcome measure, and type of design for each tier. The majority of studies was Tier II and used multiple outcome measures with each sub-study.

---

3 NR = Not reported; Sp. = Spanish.
(Pieretti, 2011; Ransford-Kaldon et al., 2011; O’Connor et al., 2014) or the same sample of children (Healy et al., 2005; Kourea et al., 2007; Gyovai et al., 2009; Richards-Tutor et al., 2012; Miller, 2013). Studies with independent samples included Eversole (2010) with grades 4th and 5th, McIntosh et al. (2007) with independent samples for year 1 and year 2, O’Connor et al. (2014) with three treatment groups, Pieretti (2011) with three treatment groups, and Ransford-Kaldon et al. (2011) with independent samples for kindergarten, 1st and 2nd grade.

Kourea (2007) used a pre-post design with ELLs and with post-test data only to compare ELLs versus non-ELLs. Likewise, Keita (2011) used post-test data to compare ELLs in a control group versus ELLs in treatment groups and post-test data only to compare ELLs versus non-ELLs. Forty-seven percent (8 studies) studies used the Dynamic Indicators of Basic Early Literacy Skills (DIBELS) oral reading fluency (ORF) as an outcome measure, 35% (6 studies) used the Phoneme Segmentation Fluency subtest from the DIBELS or AIMSWeb, 29% (5 studies) used the DIBELS Nonsense Word Fluency and 29% used Passage Comprehension subtest from a Woodcock measure. Only two studies, Kourea (2007) and Gyovai et al. (2009), used the same intervention, Early Reading Intervention.
<table>
<thead>
<tr>
<th>Study</th>
<th>Grade</th>
<th>Intervention</th>
<th>Measure</th>
<th>Design</th>
<th>Sub-Study</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kamps et al. (2007)</td>
<td>1&amp; 2</td>
<td>Direct Instruction Approach</td>
<td>WRMT-R Letter-Word ID, WA, &amp; PC DIBELS ORF &amp;NWF</td>
<td>Post TX Vs CG</td>
<td>Y</td>
</tr>
<tr>
<td>Keita (2011)</td>
<td>3</td>
<td>Sidewalk</td>
<td>TCAP Reading Composite</td>
<td>Post TX Vs CG</td>
<td>N</td>
</tr>
<tr>
<td>Dougherty Stahl et al. (2012)</td>
<td>1</td>
<td>Wilson Foundation</td>
<td>DIBELS PSF, LSF &amp; ORF</td>
<td>ELL/NonEL L Post</td>
<td>N</td>
</tr>
<tr>
<td>Keita (2011)</td>
<td>3</td>
<td>Sidewalk</td>
<td>TCAP Reading Composite</td>
<td>ELL/NonEL L Post</td>
<td>N</td>
</tr>
<tr>
<td>Kourea (2007)</td>
<td>1</td>
<td>ERI w CFA</td>
<td>CTOPP RSN Composite, WJ-III Letter-Word ID, WA, &amp; PC</td>
<td>ELL/NonEL L Post</td>
<td>N</td>
</tr>
</tbody>
</table>

Table 6. Tier II Post Data Studies Treatment versus Control Groups & Post Data ELLs versus Non-ELLs.
<table>
<thead>
<tr>
<th>Study</th>
<th>Grade</th>
<th>Intervention</th>
<th>Measures</th>
<th>Design</th>
<th>Sub-Study</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eversole (2010)</td>
<td>4 &amp; 5</td>
<td>Reading Fluency &amp; Comprehension</td>
<td>CST AYP</td>
<td>Pre-Post</td>
<td>Y</td>
</tr>
<tr>
<td>Graves et al. (2011)</td>
<td>6</td>
<td>CR/RN</td>
<td>WRMT-R PC</td>
<td>Pre-Post</td>
<td>N</td>
</tr>
<tr>
<td>Gyovai et al. (2009)</td>
<td>K</td>
<td>ERI</td>
<td>AMSWeb PSF &amp; NWF</td>
<td>Pre-Post</td>
<td>N</td>
</tr>
<tr>
<td>Healy et al. (2005)</td>
<td>1</td>
<td>SLRS w token economy</td>
<td>DIBELS NWF, PSF &amp; ORF</td>
<td>Pre-Post</td>
<td>N</td>
</tr>
<tr>
<td>Kourea (2007)</td>
<td>2</td>
<td>ERI w CFA</td>
<td>WLPB-R OL, WA &amp; PC</td>
<td>Pre-Post</td>
<td>N</td>
</tr>
<tr>
<td>Linan-Thompson et al. (2007)</td>
<td>1</td>
<td>PR/Supplemental OL</td>
<td>DIBELS ORF</td>
<td>Pre-Post</td>
<td>Y</td>
</tr>
<tr>
<td>McIntosh et al. (2007)</td>
<td>1</td>
<td>Type Tier II Literacy</td>
<td>DIBELS ORF</td>
<td>Pre-Post</td>
<td>TXCG</td>
</tr>
<tr>
<td>Miller (2013)</td>
<td>3 &amp; 5</td>
<td>SSRW</td>
<td>DIBELS Composite</td>
<td>Pre Post</td>
<td>N</td>
</tr>
<tr>
<td>Nguyen-Quang (2012)</td>
<td>1 &amp; 2</td>
<td>Language Enrichment</td>
<td>DIBELS Composite</td>
<td>Pre Post</td>
<td>TX vs CG</td>
</tr>
<tr>
<td>O'Connor et al. (2014)</td>
<td>2</td>
<td>Sound Partners/Ladders Literacy</td>
<td>DIBELS ORF, WRMT-R Total Reading GORT</td>
<td>Pre Post</td>
<td>TX vs CG</td>
</tr>
<tr>
<td>Pieretti (2011)</td>
<td>1</td>
<td>Literacy Enhancement</td>
<td>CTOPP Elision &amp; Blending Words; WJ-III Letter-Word ID, WA, &amp; PC; ROWPVT Receptive LLI Benchmark; DIBELS NWF, ORF, LNF, &amp; PSF</td>
<td>Pre-Post</td>
<td>Y</td>
</tr>
<tr>
<td>Ransford-Kaldon et al. (2011)</td>
<td>K, 1 &amp; 2</td>
<td>Leveled Literacy Intervention (LLI)</td>
<td>DIBELS PSF &amp; NWF</td>
<td>Pre-Post</td>
<td>TXCG</td>
</tr>
<tr>
<td>Richards-Tutor et al. (2012)</td>
<td>K</td>
<td>Core Intervention Model</td>
<td>DIBELS PSF &amp; NWF</td>
<td>Pre-Post</td>
<td>N</td>
</tr>
</tbody>
</table>

Table 7. Tier II Pre-Post Studies


63
<table>
<thead>
<tr>
<th>Study</th>
<th>Grade</th>
<th>Intervention</th>
<th>Measure</th>
<th>Design</th>
<th>Sub-Study</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eversole (2010)</td>
<td>2-5</td>
<td>Reading Fluency &amp; Comprehension</td>
<td>California Standardized Testing</td>
<td>Post-Test</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>TX Vs CG</td>
<td></td>
</tr>
<tr>
<td>McMaster et al. (2008)</td>
<td>K</td>
<td>Kindergarten-Peer Assisted Learning Strategies (K-PALS)</td>
<td>PALS (PAS, PAB, RNL, RLS) WRMT-R (Letter Word ID &amp; WA)</td>
<td>Pre-Post Test CG</td>
<td>No</td>
</tr>
<tr>
<td>McMaster et al. (2008)</td>
<td>K</td>
<td>K-PALS</td>
<td>AIMS Web (PSF) KPALS ORF</td>
<td>Post TX</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Vs CG non-ELL</td>
<td></td>
</tr>
<tr>
<td>Sapienza (20120)</td>
<td>3</td>
<td>Success Marker</td>
<td>RPA</td>
<td>Pre-Post TXCG</td>
<td>No</td>
</tr>
<tr>
<td>Soong (2012)</td>
<td>K</td>
<td>English-Only ESL</td>
<td>Florida Assessment for Instruction in Reading</td>
<td>Pre-Post TX</td>
<td>No</td>
</tr>
</tbody>
</table>

Table 8. Tier I studies

---

5 ESL = English as a Second Language; PAS = ; PAB= ; RNL = ; RLS ;WRMT-R = Woodcock Reading Mastery Test-Revised; RPA = ; TX = Treatment Group; CG = Control Group
Analysis by Tier

The meta-analytic analyses of this study addressed the following research questions:

1. Is there empirical support for effects of Tier I interventions on reading for ELL students?
2. Is there empirical support for effects of Tier II interventions on reading for ELL students?
3. Is there empirical support for effects of Tier III interventions on reading for ELL students?

The studies in this meta-analysis represent a random sample of all the values in the population. “Under the random-effects model the true effects in the studies are assumed to have been sampled from the distribution of true effects” (Borenstein et al., 2009, p. 74). Therefore, for the present study the author used the random effects model to generalize these results to a different group of studies including other interventions and other ethnic groups (A. Olmos, personal communication, November 2015).

Table 9. Tier III Studies

<table>
<thead>
<tr>
<th>Study</th>
<th>Grade</th>
<th>Intervention</th>
<th>Measure</th>
<th>Design</th>
<th>Sub-Study</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eversole (2010)</td>
<td>3,4 &amp; 5</td>
<td>Reading Fluency &amp; Comprehension Reading Mastery I/II Fast Cycle, Corrective Reading, PHAST &amp; PHAB/DI</td>
<td>CST</td>
<td>Pre-Post TX Vs. CG</td>
<td>Yes</td>
</tr>
<tr>
<td>Lovett et al. (2008)</td>
<td>2-8</td>
<td>Reading Mastery I/II Fast Cycle, Corrective Reading, PHAST &amp; PHAB/DI</td>
<td>CTOPP Blending Words WRAT-3 Reading Composite WRMT-R Word Attack</td>
<td>ELL/non ELL Post- Test</td>
<td>No</td>
</tr>
</tbody>
</table>

Notes. PHAST = Phonological and Strategy Training; PHAB/DI= Phonological Analysis and Blending/Direct Instruction. WRAT = Wide Range Achievement Test, Third Edition.
Four main meta-analyses including mean gain and mean difference analyses were conducted to examine the empirical evidence of reading interventions for ELLs: (1) the mean difference analysis for Tier I studies compared post-test data from treatment and control groups; (2) the mean gain analysis with Tier I pre-post-test studies grouped by treatment and control groups to compare performance of ELL participants receiving Tier I reading interventions (treatment groups) to ELLs that were not exposed to the intervention under research (control groups); (3) the mean difference analysis for Tier II studies with post-test data from treatment and control group; and (4) the mean gain analysis with Tier II pre-post-test studies grouped by treatment and control group to compare performance of ELL participants receiving Tier II reading interventions (treatment groups) to ELLs that were not exposed to the intervention under research (control groups). As there were only four effect size estimates for Tier III interventions it was not possible to discuss overall effect sizes. Eversole (2010) with three pre-post sub-studies (3rd, 4th, and 5th) with no control group yielded medium to large effect sizes (1.45, 0.80, and 0.77). Lovett et al. (2008) provided pre-post data information for the ELL group only for the WRAT reading subtest but did not report the standard deviation for the post-treatment group.
Table 10. Summary of Analyses by Tier and Reading Components

Additionally, other analyses by reading components were conducted to examine the effectiveness of reading interventions with ELLs for different reading components including fluency, phonological awareness, and reading comprehension.

**Tier I**

**Mean difference analysis with post-test data.** To investigate empirical support for effects of Tier I interventions on reading for ELL students the author used for this analysis post-test data from two sets of studies. One set of pre-post studies with treatment and control groups. The other set included post-test data studies with treatment and control groups.
The overall standardized difference in means under the random model is 0.32, although deemed a small mean effect size, reached statistical significance, \( p<0.001 \). The forest plot (Figure 3) showed the standardized mean difference effect sizes ranged from 0.085 McMaster et al. (2008) measured by PAL\( \text{S} \) Rapid Letter Naming to 0.687 McMaster et al. (2008) as measured by PALS Phonemic Awareness Segmentation. The analysis yielded 11 positive effect sizes, the majority effect sizes are small (7 effect sizes) and 4 effect sizes deemed medium. Appendix F presents plot with \( p \)-values, weights, and standard residuals for each study. The Q-test for the distribution of observed effect sizes, \( Q_{(10)} = 7.61 \), was not statistically significant (\( p = 0.666 \)) (Table 10). The author used the random effects model in order to extrapolate these results to the general population.
<table>
<thead>
<tr>
<th>Model</th>
<th>Study name</th>
<th>Subgroup within study</th>
<th>Comparison</th>
<th>Std diff in means and 95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>McMaster</td>
<td>6.01</td>
<td>Post T I 04.0.6.01</td>
<td>Post Data</td>
<td>0.687</td>
</tr>
<tr>
<td>McMaster</td>
<td>6.02</td>
<td>Post T I 04.0.6.02</td>
<td>Post Data</td>
<td>0.648</td>
</tr>
<tr>
<td>McMaster</td>
<td>6.03</td>
<td>Post T I 04.0.6.03</td>
<td>Post Data</td>
<td>0.085</td>
</tr>
<tr>
<td>McMaster</td>
<td>6.04</td>
<td>Post T I 04.0.6.04</td>
<td>Post Data</td>
<td>0.581</td>
</tr>
<tr>
<td>McMaster</td>
<td>7.01</td>
<td>Post T I 04.0.7.01</td>
<td>Post Data</td>
<td>0.108</td>
</tr>
<tr>
<td>McMaster</td>
<td>7.02</td>
<td>Post T I 04.0.7.02</td>
<td>Post Data</td>
<td>0.223</td>
</tr>
<tr>
<td>Sapienza</td>
<td>18.10</td>
<td>Post T I 19.0.18.10</td>
<td>Post Data</td>
<td>0.352</td>
</tr>
<tr>
<td>Eversole</td>
<td>Post 2 T I</td>
<td>06.3.10.10</td>
<td>Post Data</td>
<td>0.245</td>
</tr>
<tr>
<td>Eversole</td>
<td>Post 3 T I</td>
<td>06.4.10.1</td>
<td>Post Data</td>
<td>0.338</td>
</tr>
<tr>
<td>Eversole</td>
<td>Post 4 T I</td>
<td>06.5.10.1</td>
<td>Post Data</td>
<td>0.472</td>
</tr>
<tr>
<td>Eversole</td>
<td>Post 5 T I</td>
<td>06.6.10.1</td>
<td>Post Data</td>
<td>0.110</td>
</tr>
<tr>
<td>Fixed</td>
<td></td>
<td></td>
<td></td>
<td>0.318</td>
</tr>
</tbody>
</table>

Figure 3. Forest Plot of Effect Sizes for Tier I Post-Test Data

**Mean gain analysis with tier I pre-post studies treatment versus control group.** The Tier I studies were grouped by treatment and control group to compare performance of ELL participants receiving Tier I reading interventions (treatment groups) to ELLs that were not exposed to the intervention under research (control groups). The Q-test for the distribution of observed effect sizes for students in the treatment and control groups, $Q_{(7)} = 11.73$ and $Q_{(5)} = 0.80$, respectively, was not statistically significant ($p = 0.110$ and $p = 0.97$) (Table 12).
<table>
<thead>
<tr>
<th>Group/Model</th>
<th>Number Effect Sizes</th>
<th>Effect size and 95% confidence interval</th>
<th>Heterogeneity</th>
</tr>
</thead>
<tbody>
<tr>
<td>TX</td>
<td></td>
<td>Effect size estimate</td>
<td>Standard error</td>
</tr>
<tr>
<td>Fixed Effects</td>
<td>8</td>
<td>1.26</td>
<td>0.09</td>
</tr>
<tr>
<td>Random effects</td>
<td>8</td>
<td>1.32</td>
<td>0.13</td>
</tr>
<tr>
<td>CG</td>
<td></td>
<td>0.86</td>
<td>0.11</td>
</tr>
</tbody>
</table>

Table 12. Tier I Pre-Post Studies Treatment & Control Groups- Point Estimates, Confidence Interval & Q-Statistics

For the treatment group the overall standardized difference in means under the random model is 1.32, deemed a large mean effect size and statistically significant. The standardized mean difference effect sizes ranged from 0.859 to 2.128 (see Appendix F for forest plots). For the control group the overall standardized difference in means under the random model is 0.86 statistically significant. The standardized mean difference effect sizes ranged from 0.738 to 1.097. The overall standardized difference in means for the control group is large but less substantial than the overall standardized mean difference for the treatment group.

Results from both meta-analyses for Tier I interventions suggest there was evidence of positive effects of Tier I interventions on the reading of ELL students; however, these results need to be interpreted with caution due to the limited number of studies and effect sizes. The first analysis with post data only from treatment and control group yielded 11 positive, small to medium, effect sizes. The overall standardized

Notes. TX = Treatment group; CG = Control Group
difference in means under the random model was 0.32, albeit small, reached statistical significance $p < 0.001$. The second analysis that grouped the studies by treatment and control group, resulted in statistically significant large overall effect sizes for ELLs in comparison groups, 0.86, and for ELLs receiving intervention, 1.32. Though it was larger for the students in the treatment condition this suggests the students in the control groups made similar progress without the Tier I intervention.

**Tier II**

**Mean difference analysis with post-test data.** The analysis to investigate empirical support for effects of Tier II interventions on reading for ELL students included post-test data from two sets of studies. One set of pre-post studies with treatment and control groups. The other set included post-test data studies with treatment and control groups. This analysis yielded 31 effect sizes. Appendix G presents plot with p-values, weights, and standard residuals for each study.

<table>
<thead>
<tr>
<th>Model</th>
<th>Number of Effect Sizes</th>
<th>Effect size and 95% confidence interval</th>
<th>Heterogeneity</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Point estimate</td>
<td>Standard error</td>
</tr>
<tr>
<td>Fixed effects</td>
<td>31</td>
<td>0.74</td>
<td>0.06</td>
</tr>
<tr>
<td>Random effects</td>
<td>31</td>
<td>0.67</td>
<td>0.16</td>
</tr>
</tbody>
</table>

Table 13. Tier II Post-Test Data - Point Estimates, Confidence Interval & Q-Statistics

The Q-test for the distribution of observed effect sizes, $Q_{(30)} = 208.85$ was statistically significant, $p < 0.001$, suggesting heterogeneity in conditions and differences are not related to sampling variations (Table 13). The overall standardized mean difference under the random model is 0.67, deemed medium, reached statistically
significance, p < 0.001. The standardized mean difference effect sizes ranged from -1.64 to 2.84 (Figure 4).

<table>
<thead>
<tr>
<th>Study name</th>
<th>Subgroup within study</th>
<th>Comparison</th>
<th>Std diff in means and 95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Graves 7.03 Post T II</td>
<td>08.0.7.03</td>
<td>Post Data</td>
<td>-0.398</td>
</tr>
<tr>
<td>Kourea 2.15 Post T I &amp; II</td>
<td>10.0.2.15</td>
<td>Post Data</td>
<td>-1.125</td>
</tr>
<tr>
<td>Kourea 2.14 Post T I &amp; II</td>
<td>10.0.2.14</td>
<td>Post Data</td>
<td>-0.274</td>
</tr>
<tr>
<td>Kourea 2.16 Post T I &amp; II</td>
<td>10.0.2.16</td>
<td>Post Data</td>
<td>-1.643</td>
</tr>
<tr>
<td>Linan-Thompson et al. 5.20 Post Data</td>
<td>5.20.5.20</td>
<td>Post Data</td>
<td>0.064</td>
</tr>
<tr>
<td>Linan-Thompson et al. 5.32 Post Data</td>
<td>5.32.5.32</td>
<td>Post Data</td>
<td>0.517</td>
</tr>
<tr>
<td>Linan-Thompson et al. 5.33 Post Data</td>
<td>5.33.5.33</td>
<td>Post Data</td>
<td>0.497</td>
</tr>
<tr>
<td>Linan-Thompson et al. 2.16 Post Data</td>
<td>2.16.2.16</td>
<td>Post Data</td>
<td>0.479</td>
</tr>
<tr>
<td>Linan-Thompson et al. 2.16 Post Data</td>
<td>2.16.2.16</td>
<td>Post Data</td>
<td>0.578</td>
</tr>
<tr>
<td>Linan-Thompson et al. 2.16 Post Data</td>
<td>2.16.2.16</td>
<td>Post Data</td>
<td>0.266</td>
</tr>
<tr>
<td>Linan-Thompson et al. 2.16 Post Data</td>
<td>2.16.2.16</td>
<td>Post Data</td>
<td>1.060</td>
</tr>
<tr>
<td>Nguyen-Quang Post T II</td>
<td>02.0.2.10</td>
<td>Post Data</td>
<td>0.273</td>
</tr>
<tr>
<td>O'Connor 2.16 Post Data T II</td>
<td>12.3.2.16</td>
<td>Post Data</td>
<td>0.711</td>
</tr>
<tr>
<td>O'Connor 7.10 PostData T II</td>
<td>12.3.7.10</td>
<td>Post Data</td>
<td>1.043</td>
</tr>
<tr>
<td>Ransford-Kaldon 12.10 K Post T II</td>
<td>12.10.12.10</td>
<td>Post Data</td>
<td>0.962</td>
</tr>
<tr>
<td>Ransford-Kaldon 12.10 Post T II</td>
<td>12.10.12.10</td>
<td>Post Data</td>
<td>0.157</td>
</tr>
<tr>
<td>Ransford-Kaldon 12.10 2nd Post T II</td>
<td>12.10.12.10</td>
<td>Post Data</td>
<td>0.146</td>
</tr>
<tr>
<td>Ransford-Kaldon 2.15 Post T II</td>
<td>12.15.2.15</td>
<td>Post Data</td>
<td>0.835</td>
</tr>
<tr>
<td>Ransford-Kaldon 2.15 Post T II</td>
<td>12.15.2.15</td>
<td>Post Data</td>
<td>0.524</td>
</tr>
<tr>
<td>Ransford-Kaldon 2.16 2nd Post T II</td>
<td>12.16.2.16</td>
<td>Post Data</td>
<td>0.582</td>
</tr>
<tr>
<td>Sapienza 18.10 Post T II</td>
<td>19.0.18.10</td>
<td>Post Data</td>
<td>0.352</td>
</tr>
<tr>
<td>Kamps. Post 7.02 GI T I &amp; II</td>
<td>07.2.7.02</td>
<td>Post Data</td>
<td>2.843</td>
</tr>
<tr>
<td>Kamps. Post 7.03 GI T I &amp; II</td>
<td>07.2.7.03</td>
<td>Post Data</td>
<td>1.163</td>
</tr>
<tr>
<td>Kamps. Post 2.16 GI T I &amp; II</td>
<td>07.2.2.16</td>
<td>Post Data</td>
<td>2.843</td>
</tr>
<tr>
<td>Kamps. Post 2.15 GI T I &amp; II</td>
<td>07.2.2.15</td>
<td>Post Data</td>
<td>2.843</td>
</tr>
<tr>
<td>Kamps. Post 7.03 GI T I &amp; II</td>
<td>07.3.7.03</td>
<td>Post Data</td>
<td>1.163</td>
</tr>
<tr>
<td>Kamps. Post 2.16 GI T I &amp; II</td>
<td>07.3.7.03</td>
<td>Post Data</td>
<td>1.130</td>
</tr>
<tr>
<td>Kamps. Post 2.16 GI T I &amp; II</td>
<td>07.3.7.03</td>
<td>Post Data</td>
<td>0.708</td>
</tr>
<tr>
<td>Keka Post TX Vs CG 17.10 T I 17.10</td>
<td>17.0. 17.10</td>
<td>Post Data</td>
<td>0.079</td>
</tr>
</tbody>
</table>

Figure 4. Forest Plot of Effect Sizes for Tier II Studies Post-Test Data Only

**Mean gain analysis with tier II pre-post studies treatment versus control group.** In addition, the Tier II studies were clustered by treatment and control group to compare performance of ELL participants receiving Tier II reading interventions (treatment groups) to ELLs that were not exposed to the intervention under research.
The Q-test for the distribution of observed effect sizes for students in the treatment and control groups, \( Q_{(31)} = 132.69 \) and \( Q_{(21)} = 108.84 \), respectively, was statistically significant (\( p < 0.001 \)) (Table 14).

<table>
<thead>
<tr>
<th>Group/Model</th>
<th>Number of Effect Sizes</th>
<th>Point Estimate</th>
<th>Standard Error</th>
<th>Lower Limit</th>
<th>Upper Limit</th>
<th>( p )-value</th>
<th>( Q )-value</th>
<th>( P )-value</th>
<th>df (( Q ))</th>
</tr>
</thead>
<tbody>
<tr>
<td>TX</td>
<td>Fixed effects</td>
<td>35</td>
<td>1.076</td>
<td>0.059</td>
<td>0.960</td>
<td>1.193</td>
<td>0.000</td>
<td>134.71</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>Random effects</td>
<td>35</td>
<td>1.242</td>
<td>0.128</td>
<td>0.991</td>
<td>1.492</td>
<td>0.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CG</td>
<td>Fixed effects</td>
<td>25</td>
<td>0.824</td>
<td>0.072</td>
<td>0.68</td>
<td>0.96</td>
<td>0.000</td>
<td>140.54</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>Random effects</td>
<td>25</td>
<td>1.069</td>
<td>0.184</td>
<td>0.70</td>
<td>1.43</td>
<td>0.000</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 14. Tier II Pre-Post Studies Treatment & Control Groups- Point Estimates, Confidence Interval & Q-Statistics

The overall standardized mean difference under the random model is 1.24 and 1.07 for the treatment and control groups respectively. The standardized mean difference effect sizes for the ELLs under intervention (treatment groups) ranged from -1.89 to 3.32. For the control group the effect sizes ranged from -1.30 to 4.74 (see plots Appendix H).

In summary, the meta-analyses provided evidence that Tier II interventions have positive effects on reading of ELLs. The analysis with post-test data yielded 25 positive effect sizes out of 31 effect sizes. The overall standardized difference in means under the random model is 0.67, deemed medium, and reached statistical significance, \( p < 0.001 \). The Appendix F presents a plot with \( p \)-values, weights, and standard residuals for each study.

The second analysis that grouped the studies by treatment and control group, resulted in large overall effect sizes for ELLs in comparison groups, 1.24, and for ELLs
receiving intervention, 1.07, both statistically significant. This suggests the students in the control groups made similar progress without the Tier II intervention.

As previously mentioned, a meta-analysis with Tier III studies was not feasible due to the limited number of Tier III studies. As there were only four effect size estimates for Tier III interventions it is not possible to discuss overall effect sizes. Eversole (2010) with three pre-post substudies (3rd, 4th, and 5th) with no control group yielded medium to large effect sizes (1.45, 0.80, and 0.77). Lovett et al. (2008) provided pre-post data information for the ELL group only for the WRAT reading subtest that yielded a medium effect size, .79, with a standard error of 0.19. Lovett et al., (2008) also compared performance of ELLs versus Non-ELLs. Both groups received Tier III interventions and used different instruments to measure the outcomes. The CTOPP Blending Words, WRAT Reading subtest, and the WRMT-R Word Attack yielded small effect sizes of 0.34, -0.36, and 0.45 respectively. This suggests ELLs and non-ELLs responded similarly to the intervention.

The following analyses focused on the essential reading components for the development of reading. Each analysis was grouped by outcome measures used by different studies including the Dynamic Indicators of Early Basic Literacy Skills (DIBELS), AIMSWeb, and Woodcock measures. The goal was to examine the evidence of the effects of the tiers of intervention on different components of reading for ELL students to determine what interventions are more effective for a specific skill.
Analyses by Reading Components

**Reading Fluency.** The National Reading Panel defined reading fluency as the ability to read orally with “speed, accuracy, and proper expression” (p. 3-5) facilitating reading comprehension. Four studies (Linan-Thompson et al., 2007; Kourea, 2007; Ransford Kaldon et al., 2011; O’Connor et al., 2014) and four sub-studies (McIntosh et al., 2007) were included to compare performance of ELL participants receiving reading interventions (treatment groups) and students that were not exposed to the intervention under research (control groups). These studies used the DIBELS Oral Reading Fluency (ORF) as one of their outcome measures. The DIBELS is a standardized curriculum based measure and defines ORF as the number of words read correctly per minute (Good & Kaminski, 2002). The Q-test for the distribution of observed effect sizes for students in the treatment and control groups, $Q_{(7)} = 5.054$ and $Q_{(7)} = 12.346$, respectively, was not statistically significant ($p = 0.653$ and 0.090).

![Table 15. Oral Reading Fluency Comparison of Treatment and Control Groups- Point Estimates, Confidence Interval & Q-Statistics](image)

<table>
<thead>
<tr>
<th>Group/Model</th>
<th>Point estimate</th>
<th>Number Effect Sizes</th>
<th>p-values</th>
<th>Q-value</th>
<th>df</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment Group</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fixed effects</td>
<td>1.54</td>
<td>8</td>
<td>0.000</td>
<td>5.054</td>
<td>7</td>
<td>0.653</td>
</tr>
<tr>
<td>Random effects</td>
<td>1.54</td>
<td>8</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Control Group</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fixed effects</td>
<td>1.51</td>
<td>8</td>
<td>0.000</td>
<td>12.346</td>
<td>7</td>
<td>0.090</td>
</tr>
<tr>
<td>Random effects</td>
<td>1.53</td>
<td>8</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

For ELLs in the treatment groups, the overall standardized difference in means was statistically significant under the random-effects model with a large effect size of...
1.54. The overall standardized mean difference under the random model for ELL in the control groups is 1.53. The overall effect size is large and statistically significant, p<0.001. The standardized mean difference effect sizes for the ELLs under intervention (treatment groups) ranged from 1.307 for the study conducted by Kourea (2007) to 2.64 for the year two study conducted by McIntosh et al. This study (McIntosh, Graves & Gersten, 2007) had independent samples for the first year and second year; however, dependency is still a problem. Even though the sub-studies used independent samples for first and second year, the same researchers conducted the sub-studies.

The standardized mean difference effect sizes for the ELLs in the control groups ranged from 0.818 for the study conducted by Kourea (2014) to 2.99 for the year two study conducted by McIntosh et al. This study (McIntosh, Graves & Gersten, 2007) had independent samples for year one and year two; however, dependency is still a problem. Even though the sub-studies used independent samples for year one and two, the same researchers conducted the sub-studies. In summary, the ELLs in the treatment and control group made similar progress in oral reading fluency. The effect sizes for both groups were considered large and statistically significant (see Appendix I for obtained plots).

**Alphabets: Phonological Awareness and Phonics:** Phonemic awareness is defined as the ability to manipulate, blend and segment sounds or phonemes in oral syllables and words (National Reading Panel, 2000). Phonics instruction focuses on letter-sound correspondence and spelling patterns to teach students how to read and spell (National Reading Panel, 2000). Three studies (Healy, Vanderwood & Edelston, 2005)
and four sub-studies, two by Ransford Kaldon, Sutton Flynt, and Ross (2011) and two conducted by Richards Tutor et al. (2012) were included to compare performance of ELLs receiving reading interventions (treatment/intervention groups) and students that received the usual instruction. These studies used the DIBELS or AIMSWeb Nonsense Word Fluency (NWF) as one of the outcome measures. The DIBELS and AIMSWeb are standardized curriculum based measures and define NWF as the correct number of nonsense words read per minute (Good & Kaminski, 2002; Shinn & Shinn (2002). The Q-test for the distribution of observed effect sizes for students in the treatment and control groups, $Q_{(6)} = 14.52$ and $Q_{(2)} = 1.39$, respectively, was not statistically significant.

<table>
<thead>
<tr>
<th>Model</th>
<th>Number Effect Sizes</th>
<th>Effect size and 95% confidence interval</th>
<th>Heterogeneity</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Point estimate</td>
<td>Standard error</td>
</tr>
<tr>
<td>TX</td>
<td>7</td>
<td>1.54</td>
<td>0.13</td>
</tr>
<tr>
<td>Fixed effects</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Random effects</td>
<td>7</td>
<td>1.66</td>
<td>0.24</td>
</tr>
<tr>
<td>CG</td>
<td>3</td>
<td>1.33</td>
<td>0.30</td>
</tr>
<tr>
<td>Fixed effects</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Random effects</td>
<td>3</td>
<td>1.33</td>
<td>0.30</td>
</tr>
</tbody>
</table>

Table 16. NWF Treatment & Control Groups - Point Estimates, Confidence Interval & Q-Statistics

For ELLs in the treatment groups, the overall standardized difference in means under the random-effects model is 1.66 (Table 16), deemed a large mean effect size, statistically significant. The overall standardized mean difference under the random-effects model for ELL in the control groups is 1.33 (Table 16), statistically significant. This effect size is considered large but smaller than the standardized mean difference for
the treatment groups. The standardized mean difference effect sizes for the ELLs in the control groups ranged from 0.985 for the study conducted by Kourea (2014) to 1.84 for the sub-study with first graders conducted by Ransford-Kaldon, Sutton Flynt & Ross (2011). Even though these sub-studies had independent samples for kindergarten and first grade, the same researchers conducted the sub-studies causing dependency problems. The seven effect sizes calculated for ELLs in treatment groups ranged from 1.09 for the sub-study with students initially not at-risk (Treatment group 1) conducted by Richards-Tutor et al. (2012) to 2.77 for Healy, Vanderwood & Edelton’s study (2005) (Appendix J).

Three sub-studies (Pieretti, 2011) and two studies (Linan-Thompson, Cirino & Vaughn, 2007; McMaster, Kung, Han & Cao, 2008) were included as the unit of analysis to compare performance of ELLs receiving reading interventions (treatment/intervention groups) and students that received the usual instruction. These studies used the Word Attack subtest from Woodcock Language Proficiency Battery-Revised (WLPB-R), the Woodcock Reading Mastery Test-Revised (WRMT-R), or the Woodcock Johnson-III as one of their outcome measures. The Q-test for the distribution of observed effect sizes for students in the treatment and control groups, $Q_{(4)} = 1.15$ and $Q_{(2)} = 0.78$, respectively, was not statistically significant ($p = 0.884$ and 0.676) suggesting homogeneity in conditions and differences are related to sampling variations (Table 17).
<table>
<thead>
<tr>
<th>Group/Model</th>
<th>Number Effect Sizes</th>
<th>Effect size and 95% confidence interval</th>
<th>Heterogeneity</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Point estimate</td>
<td>Standard error</td>
</tr>
<tr>
<td>Treatment</td>
<td>Fixed effects</td>
<td>5</td>
<td>1.10</td>
</tr>
<tr>
<td></td>
<td>Random effects</td>
<td>5</td>
<td>1.10</td>
</tr>
<tr>
<td>Control</td>
<td>Fixed effects</td>
<td>3</td>
<td>0.98</td>
</tr>
<tr>
<td></td>
<td>Random effects</td>
<td>3</td>
<td>0.98</td>
</tr>
</tbody>
</table>

Table 17. Word Attack Treatment & Control Groups - Point Estimates, Confidence Interval & Q-Statistics

For ELLs in the treatment groups, the overall standardized difference in means under the random model is 1.10, deemed a large mean effect size and statistically significant. The overall standardized mean difference under the random model for ELL in the control groups is 0.98, statistically significant. This effect size is considered large but smaller than the standardized mean difference for the treatment groups. The effect sizes calculated for ELLs in treatment groups ranged from 0.74 for Pieretti’s sub-study (2011) with students receiving the LEG intervention to 1.24 for the study conducted by Linan-Thompson, Cirino & Vaughn (2007). The standardized mean difference effect sizes for the ELLs in the control groups ranged from 0.73 (McMaster et al., 2008) to 1.087 for the study conducted by Linan-Thompson, Cirino & Vaughn, 2007 (see Appendix K for obtained plots).

**Reading comprehension.** The National Reading Panel referred to the definition by Durkin (1993). This author viewed comprehension as an active and intentional thinking process “during which meaning is constructed through interactions between text
and reader” (National Reading Panel, p.4-39). Besides being an interactive process, the National Reading Panel notes reading comprehension is a cognitive process that requires complex skills involving the understanding of vocabulary. Cummins (n.d.) stated comprehension involves not only vocabulary or understanding the meaning of text but also how words are organized in sentences and paragraphs to produce meaning. Three sub-studies (Pieretti, 2011) and two studies (Linan-Thompson, Cirino & Vaughn, 2007; Graves, Pyle & Brandon, 2011) were included as the unit of analysis to compare performance of ELLs receiving reading interventions (treatment/intervention groups) and students that received the usual instruction. These studies used the Passage Comprehension subtest from the Woodcock Language Proficiency Battery-Revised (WLPB-R), the Woodcock Reading Mastery Test-Revised (WRMT-R), or the Woodcock Johnson-III as one of their outcome measures. The Q-test for the distribution of observed effect sizes for students in the treatment and control groups, \( Q(4) = 5.77 \) and \( Q(2) = 0.60 \), respectively, was not statistically significant (\( p = 0.217 \) and 0.741) suggesting homogeneity in conditions and differences are related to sampling variations (Table 18).

<table>
<thead>
<tr>
<th>Group/Model</th>
<th>Number Effect Sizes</th>
<th>Effect size and 95% confidence interval</th>
<th>Heterogeneity</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Point estimate</td>
<td>Standard error</td>
</tr>
<tr>
<td>Treatment Fixed effects</td>
<td>5</td>
<td>0.58</td>
<td>0.15</td>
</tr>
<tr>
<td>Treatment Random effects</td>
<td>5</td>
<td>0.54</td>
<td>0.19</td>
</tr>
<tr>
<td>Control Fixed effects</td>
<td>3</td>
<td>0.33</td>
<td>0.17</td>
</tr>
<tr>
<td>Control Random effects</td>
<td>3</td>
<td>0.33</td>
<td>0.17</td>
</tr>
</tbody>
</table>

Table 18. Comprehension Treatment & Control Groups - Point Estimates, Confidence Interval & Q-Statistics
For ELLs in the treatment groups, the overall standardized difference in means under the random model is 0.54 (medium), reached statistically significance. The overall standardized mean difference under the random model for ELL in the control groups is 0.33, considered small. The effect sizes calculated for ELLs in treatment groups ranged from 0.035 for Pieretti’s sub-study (2011) with students receiving the LEG intervention to 0.992 for the study conducted by Linan-Thompson, Cirino & Vaughn (2007). The standardized mean difference effect sizes for the ELLs in the control groups ranged from 0.125 (Graves, Pyle & Brandon, 2011) to 0.422 for the study conducted by Linan-Thompson, Cirino & Vaughn (2007) (Appendix L).

Furthermore, the mean difference analysis to investigate empirical support for effects of interventions on reading comprehension for ELL students included post-test data from treatment and control groups from pre-post studies (Linan-Thompson, Cirino & Vaughn, 2007; Graves, Pyle & Brandon, 2011) and studies with post-test data only (Kamps et al., 2007). Pieretti was not included in this analysis. The Q-test for the

<table>
<thead>
<tr>
<th>Model</th>
<th>Effect size and 95% confidence interval</th>
<th>Heterogeneity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>Effect Sizes</td>
<td>Point estimate</td>
</tr>
<tr>
<td>Fixed Random effects</td>
<td>4</td>
<td>0.87</td>
</tr>
</tbody>
</table>
| Table 19. Post-Test Data Reading Comprehension - Point Estimates, Confidence Interval & Q-Statistics

81
distribution of observed effect sizes, $Q_{(3)} = 14.67$. The overall standardized mean
difference under the random model is large, 0.94, and statistically significant. The
standardized mean difference effect sizes ranged from 0.40 to 1.78 (Appendix L).

In summary, the analysis by reading measures indicated the overall effect sizes
for the treatment and control groups are considered large and statistically significant.
These results suggest that ELLs in the treatment and control groups made similar
progress in oral reading fluency, phonics and phonological awareness as measured by the
DIBELS (ORF and NWF), AIMSWeb (ORF and NWF), and Woodcock measures (Word
Attack subtest). Results from interventions addressing reading comprehension indicated
the overall effect size for the treatment group was medium, while the overall effect size
for the control group is small. In addition, the analysis using post data yielded a
statistically significant large effect size for reading comprehension.

**Publication Bias**

One of the major limitations of meta-analysis is publication bias. One of the
methods for addressing bias is the funnel plot.

In the absence of publication bias, the studies are distributed symmetrically about
the mean effect size, since the sampling error is random. In the presence of
publication bias the studies are expected to follow the model, with symmetry at
the top, a few studies missing in the model, and more studies near the bottom.
(Borenstein et al., 2009, p. 283)

Funnel plots were generated to assess publication bias for each analysis with Tier
I and Tier II. The funnel plots (Figure 5 and Figure 6) indicated the effects of the meta-
analyses with Tier I and Tier II studies were symmetrically distributed suggesting there is
no indication of publication bias.
Another approach to examine the impact of publication bias is the Trim and Fill method. This is an iterative method that computes the best estimate of the unbiased effect size by removing the most extreme small studies from the positive side of the funnel plot, “re-computing the effect size at each iteration until the funnel plot is symmetric about the new effect size” (Borenstein, 2009, p. 286).

For the Tier I studies mean difference analysis with post-test data, the Trim and Fill approach imputed one additional study to improve the distribution. The addition of
this study (red circle) would decrease the standardized mean difference effect size from 0.32 to 0.31 under the random effects model (Appendix N).

For the mean gain analysis with Tier I pre-post studies control groups, the Trim and Fill approach included three additional studies (red circles) and imputed two studies for the treatment groups to improve these distributions (Appendix N). The addition of these studies would decreased the standardized mean difference effect size from 0.86 to 0.77 under the random effects model for control groups and from 1.32 to 1.20 for the treatment groups. Even though the original effect size changed, the adjusted effect size is still considered large suggesting no problems with publication bias.

For the Tier II studies mean difference analysis with post-test data, the Trim and Fill approach imputed no additional studies to improve the distribution. For the mean gain analysis Tier II pre-post studies control groups, the Trim and Fill approach included six additional studies (red circles) to improve these distributions and imputed nine studies for the treatment groups (Appendix N). The addition of these studies would decreased the standardized mean difference effect size from 1.07 to 0.69 under the random effects model for control groups and from 1.24 to 0.95 for the treatment groups suggesting there is no problems with publication bias.
CHAPTER FIVE

Discussion

This study examined research on reading interventions and focused on evidence-based literacy interventions for English Language Learners (ELLs) implemented as part of the Response to Intervention (RTI) model. Meta-analysis was used to aggregate and compare findings of research studies (Lipsey & Wilson, 2001) and the effects on reading achievement of different research-based reading interventions. RTI is a multi-level system of prevention and intervention that provides more intensive instructional support during each successive tier (Stecker, 2007). RTI incorporates assessment and intervention to enhance students’ academic achievement and behavior. Within this model, schools implement evidence-based interventions, use assessment and data to identify students’ at risk, apply progress monitoring tools, and adjust the intensity and type of intervention based on the students’ response to the intervention (National Center on Response to Intervention, 2014).

This study sought to add to the understanding of the implementation and effectiveness of RTI with English language learners and reveal implications for policy and school-based leadership. To study this, the researcher classified and analyzed the research on reading interventions with English language learners since the implementation of RTI (2004) to present. After reviewing the abstracts and methods of 130 studies, 57 studies did not meet the inclusion criteria while 63 studies were retained.
for further examination. Out of 63 studies, 43 studies did not meet the inclusion criteria. Even though most of these 43 studies examined research-based interventions with ELLs and used reliable outcome measures, the studies did not mention RTI or tiers of intervention. Other studies implemented the interventions in Spanish or English and Spanish and measured outcomes in both languages, others used reliable reading measures but were not specific about the intervention, and other studies did not provide appropriate statistical information to calculate effect sizes.

Previous syntheses and meta-analyses focused on interventions for reading with ELLs; however, some of the studies included in these syntheses and meta-analyses were not implemented as part of the RTI model with the graduated levels of support or tiers of intervention. The National Literacy Panel (2006) synthesized quantitative and qualitative studies to investigate the development of literacy for language minority students (August & Shanahan, 2006). Klingner, Artilles, and Méndez Barletta (2006) conducted a synthesis to investigate the difference between ELLs with a learning disability and students who struggle with literacy due to limited proficiency in English. Based on this synthesis of research, the following factors were proposed for a successful RTI model with ELLs: a learning environment where literacy is considered a sociocultural practice (Artiles, 2002 cited by Klingner, Artilles, and Méndez Barletta, 2006), where cultural and linguistic diversity are valued (Ortiz, 1997, 2002; Nieto, 2004; Baca, 2012 cited by Klingner, Artilles, and Méndez Barletta, 2006), and where teachers know instructional practices that are tailored for ELLs.
Han (2009) conducted a meta-analysis of evidence-based reading instruction for ELLs from pre-school through sixth grade. Han’s meta-analysis included 29 studies from peer-reviewed journals. Dissertations, reports, and conference presentations were not included. The author classified the studies into Tier I and Tier II but the majority of studies in this meta-analysis do not refer to RTI as a framework. Another synthesis that focused on reading interventions with ELLs was conducted by Cheung and Slavin (2012). These authors reviewed twenty-two studies, from 1980 to 2010, to examine the effectiveness of reading programs with Spanish dominant ELL students. This synthesis identified effective reading programs for Spanish-speaking students but did not describe the intervention as part of the RTI model.

Overall, the main difference of the present study with previous meta-analyses and syntheses is the focus on research-based interventions implemented within an RTI framework. The present study included peer reviewed articles as well as dissertations and a conference paper from 2005 through 2014. Han included studies from peer-reviewed journals from 1967 through 2009. Cheung and Slavin (2012) investigated reading programs with Spanish dominant students. The present meta-analysis examined studies with speakers of other languages besides Spanish including Hmong, Portuguese, and Somali.

The purpose of the present study was to determine if there was empirical support for research-based reading interventions that produce improvement in reading for ELLs. Twenty-seven studies that quantitatively examined the effects of Tier I, Tier II and Tier III research-based reading interventions for ELLs, from kindergarten through 8th grade,
were included in this meta-analysis. The meta-analytic analyses of this study addressed the following research questions:

1. Is there empirical support for effects of Tier I interventions on reading for ELL students?
2. Is there empirical support for effects of Tier II interventions on reading for ELL students?
3. Is there empirical support for effects of Tier III interventions on reading for ELL students?

Tier I of RTI encompasses universal screening, classroom based-instruction, and assessment in the general education classroom with all students (Vaughn & Roberts, 2007; Hazelkorn et al., 2011). Two types of analyses were conducted with Tier I studies, a mean difference and a mean gain analyses. The mean difference analysis for Tier I studies compared post-test data from treatment and control groups. This analysis yielded 11 effect sizes and revealed a statistically significant overall effect size for Tier I interventions on the reading of ELL students; however, the effect size was small (ES = 0.32). The mean gain analysis with Tier I pre-post-test studies, grouped by treatment and control groups, yielded 15 effect sizes, 8 effect sizes for the treatment, and 7 effect sizes for the control group. The results of this analysis showed large effect sizes for both groups: 0.86 for the comparison group and 1.32 for ELLs receiving interventions (treatment). This suggests the students in the treatment groups made more substantial progress than students in the control groups but the effect sizes for both groups are large.
Tier II focuses on specialized or targeted interventions for students who are not making adequate progress in the core program or in Tier I (Hazelkorn et al., 2011). The students at-risk receive targeted instruction to help close the gap between their current performance and their expected performance. The specialized, scientifically based instruction can be 30 minutes or more in addition to Tier I. The present meta-analyses yielded 91 effect sizes for Tier II studies. The mean difference analysis for Tier II studies with post-test data from treatment and control group resulted in 31 effect sizes. The overall standardized difference in means under the random model was medium (ES = 0.67) and reached statistical significance. The mean gain analysis compared performance of ELL participants receiving Tier II reading interventions (treatment groups) to ELLs that were not exposed to the intervention under research (control groups). This analysis resulted in a large overall effect sizes for ELLs in comparison groups (ES = 1.07) and for ELLs receiving intervention (ES = 1.24), both statistically significant. This suggests the students in the control groups made similar progress without the Tier II intervention.

Tier III is considered to be the most sustained and intensive of all the levels and is focused on individual student need. Tier III provides intensive scientifically based instruction to students with significant difficulty in reading that did not respond sufficiently to Tier I and Tier II (Vaughn & Roberts, 2007). It is important to note a separate analysis for Tier III studies was not feasible due to the limited number of studies found for this tier.

Furthermore, the researcher conducted additional analyses grouping the studies that used similar outcome measures. These analyses indicated the overall effect sizes for
the treatment and comparison groups are large and statistically significant in the areas of oral reading fluency, phonics, and phonological awareness. The exception was reading comprehension. The mean difference analysis using post-test data from control and treatment groups yielded a statistically significant large effect size for reading comprehension. In addition to the mean difference analysis, a mean gain analysis showed the overall standardized difference in means under the random model was 0.54 (medium) and 0.33 (small) for the treatment and control group respectively, both statistically significant.

As previously mentioned, RTI provides a proactive process of early interventions and evidence-based instruction to all students with additional intensive and individualized interventions to prevent student underachievement, including students at risk for academic failure and culturally and linguistically diverse students (Vellutino et al., 1996; Donovan & Cross, 2002; Francis et al., 2006; Fuchs & Fuchs, 2006; Vellutino et al., 2006; Al Otaiba et al., 2009). Despite this evidence of RTI effectiveness with non-ELLs students as well as culturally and linguistically diverse students, these findings suggest that reading interventions as part of Tier I and Tier II have questionable effects on improving reading for ELLs. While the results of this study do not provide conclusive findings regarding the effectiveness of interventions for ELLs, several implications for further research emerge. The results of this meta-analysis raised questions about the dominance of Tier II interventions in the research, the lack of difference between treatment and control groups, and teacher’s background and context.
Dominance of tier II interventions. The majority of studies were classified as Tier II; however, some studies included detailed information about Tier I and Tier II (Kamps, 2007; Kourea, 2007; McIntosh, 2007; Eversole, 2010; Dougherty Stahl, 2012) and analyzed results for each tier (Eversole, 2010). It is important to note RTI is a continuum, in order to receive Tier II interventions students receive Tier I interventions first. The main difference between the tiers is “intervention intensity and measurement precision” (Reschly, 2005, p. 511). Therefore, it is difficult to determine the effect of Tier II studies without detailed information about Tier I. This finding suggests further research for Tier I and Tier III interventions is necessary.

Lack of difference between treatment and control groups. The four main analyses with the Tier I and Tier II studies yielded large effect sizes for treatment and control groups. One was hoping to find a difference between the students receiving the usual instruction (control groups) and treatment groups; however, the students in the control groups made similar progress without the intervention. Likewise, the analyses by reading components showed similar results for control and treatment groups except for the reading comprehension measures. The overall effect size for the treatment group was medium while the overall effect size for the control group was small, both statistically significant. The mean difference analysis using post-test data from control and treatment groups yielded a statistically significant large effect size for reading comprehension. This suggests the gains in reading comprehension made by the treatment group were more substantial than the control group.
The interventions with medium to large effect sizes for reading comprehension included: Proactive Reading (Linan-Thompson et al., 2007), Direct Instruction (Kamps et al., 2007), and the Culturally Relevant Oral Narrative Enhancement with Language Experience Approach (Pieretti, 2011). Based on the effect sizes and differences between the control and treatment groups, these interventions showed promising results to improve reading comprehension.

**Teacher’s background and context.** A key element of the culturally and linguistically responsive RTI model is the need of teachers with culturally responsive practices and knowledge about the needs of ELLs (Klingner & Edwards, 2006). Only a few studies (Dougherty Stahl, 2012; Kourea, 2007; McIntosh, 2007; McMaster, 2008) addressed the years of experience and type of education of the personnel delivering the intervention including information about whether or not teachers had an ELL certification (Doughtery Stahl, 2012). The principal investigators for the majority of studies presented evidence of their knowledge about RTI and ELLs including culturally responsive practices; however, there was limited information about what the teachers knew. All the studies conducted trainings about the implementation of the intervention and data collection with the personnel delivering the intervention and collecting data and used rigorous methods to ensure fidelity of the intervention.

Evidence-based interventions validated with diverse populations are a critical component of a culturally and linguistically responsive RTI (Klingner, Artiles, & Méndez Barletta, 2006; Klingner & Edwards, 2006; Ortiz & Klingner, 2010). All the studies in this meta-analysis focused on measuring the effectiveness of reading interventions with
ELLs and disaggregated data for ELLs. In general, the studies made an attempt to include sociocultural factors of the intervention by providing a detailed description of RTI, considerations for ELLs, and qualitative information including perceptions of teachers about RTI, as well as a description of the core program and the context (type of school, location, and population). As stated by Klingner, Artiles, and Méndez Barletta (2006), it is essential for the success of RTI to implement it in a learning environment where literacy is considered a sociocultural practice.

**Limitations**

Zehler et al. (2003) reported the outcome data for instructional programs with ELLs is not disaggregated by language proficiency level. Benchmarks and rates of progress vary within the group of ELLs with different levels of proficiency in the second language (Linan-Thompson, Cirino, & Vaughn, 2007), despite these observations from previous research, most of the studies did not disaggregate results by language proficiency of the participants. In addition, the majority of participants were Spanish-speaking students and only a few studies included other languages (Hmong, Portuguese, and Somali). Another shortcoming of this study is the limited number of studies for Tier III suggesting the need of more studies for at-risk ELL students that need intensive interventions.

As previously reported by Han (2009), there is a lack of research on vocabulary. Explicit vocabulary instruction is recommended to enhance reading comprehension in first and second language (Klingner et al., 2006). The present study located several
vocabulary studies with ELLs but these studies were not implemented as part of the RTI model and were excluded for the present meta-analysis.

Other limitations are related to the methodology, previously discussed under the methods section. The present study did not include the source of heterogeneity through a moderator analysis, which allows the researcher to examine if the effect sizes vary based on the level of the moderator (Card, 2012). Some of the moderators that may explain the remaining variance includes different levels of English language proficiency of the participants, educational experience in the US, years of experience of the personnel providing the intervention and knowledge about ELLs, length of the intervention, English language development services, support participants receive at home, and exposure to literacy after school. Another limitation of meta-analysis is publication bias. This study included published and unpublished studies such as dissertations and a conference paper to obtain a better estimate of the true effect size of the target population of studies. The funnel plots and Trim and Fill methods suggested no problems with publication bias (Appendix M and Appendix N).

The problem with dependency was another shortcoming of this study. The majority of primary studies used multiple outcome measures for the same intervention with the same sample. Other studies had independent samples or sub-studies; however, the same researcher conducted the studies creating problems with dependency.

Conclusions

In terms of policy, this study reinforced the idea that evidence-based interventions are not applicable to all students and believing that “one size fits all” can create
inappropriate referrals and misidentification of ELLs in special education (Orosco & Klingner, 2010). With the reauthorization of IDEA (2004), the RTI model is used to identify Specific Learning Disabilities (SLD) when the students show lack of response to research-based interventions. Eligibility for SLD for ELLs must go beyond data from curriculum-based measurements. To prevent the false positive identification of ELLs (Klingner, 2006) over-identification or under-identification of ELLs with learning disabilities, school personnel with knowledge about second language acquisition must rule out if the problem is related to second language acquisition, compare the response to intervention of ELLs to similar peers, and ensure interventions are validated with this population. Previous research emphasized a cultural and linguistic RTI model involves research-based interventions tailored and validated with minority and ELL students (Orosco & Klingner, 2006). School districts need to demonstrate the effectiveness of the multi-levels of support including the interventions not only with the general population but also with ELLs and other subgroups (Hank Fien et al., 2010).

Overall findings of this study revealed a lack of difference between treatment and control groups for Tier I and Tier II interventions. One was expecting to find a difference between the treatment and control groups receiving RTI interventions but instead this study showed large effect sizes for control and treatment groups across interventions and reading components except for reading comprehension. Therefore, before adopting Tier I and Tier II reading programs for ELL students, school leaders need to examine carefully results of these interventions with this subgroup and interpret with caution studies that only used pre-post intervention models with no control groups. If the interventions show
promising results for ELLs, school leaders may consider piloting the intervention to evaluate the effects of the interventions and compare effects to the interventions already implemented by the school district.

A key element of the culturally and linguistically responsive RTI model is the need of teachers with culturally responsive practices and knowledge about the needs of ELLs (Klingner & Edwards, 2006). Only a few studies addressed the context and preparation of the personnel delivering the intervention. The primary studies targeted the essential reading components proposed by the National Reading Panel, conducted trainings about the implementation of the intervention, and used rigorous methods to ensure fidelity of the intervention but there was not clear evidence of trainings addressing linguistically and culturally responsive practices. This finding suggests that future research with ELLs and RTI should address the preparation of teachers or personnel delivering the interventions and investigate possible moderators that can explain the heterogeneity among effects sizes.

This research attempted to add information about RTI with ELLs and demonstrated that further research is necessary to meet the linguistic and cultural needs of ELL students. The low number of RTI studies with ELLs noted does indicate that language is often not considered as a variable in the implementation of research-based intervention as part of RTI. Benchmarks and rates of progress vary within the group of ELLs with different levels of proficiency in the second language (Linan-Thompson, Cirino, & Vaughn, 2007). A few studies provided the language proficiency of the participants and educational experience but in general the studies did not differentiate
how ELLs with different levels of language proficiency responded to the interventions. Therefore, future research with ELLs and RTI should disaggregate results for ELLs by language (if different groups of ELLs are included) and language proficiency of the participants.

Additional research is warranted in the areas of oral language and vocabulary, these are essential components for the development of reading and Cognitive Academic Language Proficiency (CALP) of ELLs. Further research is also needed for Tier III interventions for at-risk students that need intensive interventions as well as studies with upper grade students and speakers of other languages besides Spanish.

Before investing resources and implementing interventions for English language learners, school leaders need to demonstrate if the tiers of intervention are structurally sound and implemented with fidelity, as well as if the general population and specific subgroups are achieving successful outcomes (Hank Fien et al., 2010). Leaders need to support teachers and provide specialized training to develop culturally responsive practices and knowledge about the needs of ELLs. The capacity to infuse language-based interventions might enhance the effectiveness of interventions with ELL students. As stated by Cheung and Slavin (2012) the most effective interventions provide substantial professional development and coaching for teachers and cooperative learning, which provides opportunities for ELL students to practice English in a meaningful context.

This study added information to the existing syntheses and meta-analyses about reading interventions with ELLs by focusing on evidence-based interventions implemented within an RTI framework. When trying to provide empirical support for
effects of Tier I, Tier II and Tier III interventions, this study identified the following issues for researchers to evaluate the effectiveness of interventions with ELLs: identification of the composition of the sample, method of disaggregating data for ELLs, quality of outcome measures, and difference between how the participants responded compared to student receiving the usual program.

Results from this study revealed that RTI is working with ELLs especially in the area of reading comprehension, but it could be accelerated its impact by ensuring that Tiers I, II and III teachers understand English language acquisition and culturally responsive practices, the context in which children learn and live are critical to framing the supports they receive and instructional literacy practices need to go beyond those recommended by the National Reading Panel and the National Literacy Panel in order to appropriately include the literacy needs of ELLs.
REFERENCES


---


World-Class Instructional Design and Assessment. (2013). *Developing a culturally and linguistically responsive approach to response to instruction and intervention for English language learners.* Wisconsin, MD: Board of Regents of the University of Wisconsin.


APPENDIX A

A Culturally and Linguistically Responsive RTI Model

Adapted from “Considerations when implementing RTI with English language Learners,” by Klingner, J., 2010.
APPENDIX B

Interventions with English Language Learners Meta-Analysis Codebook

Note: 0 = N/A or Not Reported or No for all coding categories

Report Identification and Citation
STUDYID  Study ID number starts with 01.
Studies with additional independent substudies are coded separately. Add a decimal to the study ID.
For example: 01.1 and 01.2 is study 01 with two independent substudies
If the study presents results using different measures, create a row for each measure:
The first number indicates the study number
The second number indicates grade level (enter 0 if no breakout for grade level)
The third number indicates measure used (Use table from Appendix C for codes)
The fourth number indicates composite or subtest used (Use table from Appendix C for codes)
For example:
01.0.0.0 would simply be coded as 01 – no breakout for grade level, only one measure, no independent substudies
01.0.1.12 (study 01, no breakout for grade level, CTOPP, Blending Words)
01.2.1.12 (study 01, first grade, CTOPP, Blending Words)
01.3.1.22 (study 01, second grade, CTOPP, Non-word Repetition)

NOTES  Use Same Students for additional measures
(0) No
(1) Yes
(2) Different students for each treatment group, the same control group

AUTHOR  Enter last name and initial for first name (e.g., Nguyen, F)
YEAR  Year of publication
STATE/PROVINCE  Enter state or province (move before sampling)
COUNTRY  Enter country (move before sampling)

Retrieval Information
PUBTYPE  (Publication type)
(1) Journal
PUBNAME (enter the code)
(A) American Educational Research Journal
(B) Journal of Learning Disabilities
(C) Urban Education
(D) Council for Exceptional Children
(E) Learning Disability Quarterly
(F) Child Development
(G) The Elementary School Journal
(H) Perspectives
(I) Journal of Educational Psychology

115
| (J) The California School Psychologist          |
| (K) Behavioral Education                       |
| (L) Assessment for Effective Intervention      |
| (2) Thesis or doctoral dissertation            |
| (3) Organization report                        |
| (4) Conference                                 |
|   (A) Workshop                                 |
|   (B) Paper Presentation                       |
|   (C) Poster                                   |
| (5) Unpublished manuscript                     |
| (6) Other (Specify source)                     |

**STUDYRET**  (Study Retrieval) Use Appendix B Preliminary Coding for retrieval information

(0) n/a

**1** Electronic database

   (A) Academic Search Complete
   (B) Education Resources Information Center (ERIC)
   (C) PsycINFO
   (D) PsycARTICLES
   (E) ProQuest dissertations & theses
   (F) JSTOR
   (G) Google Scholar
   (H) Sociological Abstracts

**2** Organizational web site search

   (A) American Education Research Association (AERA)
   (B) National Association for Bilingual Education (NABE)
   (C) International Reading Association (IRA)
   (D) National Council of Teachers of English (NCTE)
   (E) Gates Foundation
   (F) Annie E. Casey
   (G) National Center for Education Statistics (NCES)
   (H) National Center on Educational Outcomes (NCEO)
   (I) American Institutes for Research (AIR)
   (J) National Association of School Psychologist

**3** Bibliography

**4** Synthesis/meta-analysis

**QUALITY**  (Research quality)

(0) Not reported
(1) Peer reviewed
(2) Published dissertation
Characteristics of the sample or Demographic Information

SAMPLING (Participant Sampling Method)

(1) Population
   (A) state population
   (B) school district population
   (C) local population

(2) Simple random selection
   (A) state population
   (B) school district population
   (C) local population

(3) Stratified random selection
   (A) state population
   (B) school district population
   (C) local population

(4) Systematic selection
   (A) state population
   (B) school district population
   (C) local population

(5) Available (convenience sample)
   (A) state population
   (B) school district population
   (C) local population

(6) Purposive sampling
   (A) state population
   (B) school district population
   (C) local population

ETHNICITY

(0) Not reported

(1) Greater than 60% White

(2) Greater than 60% African-American

(3) Greater than 60% Hispanic

(4) Greater than 60% Asian

(5) Mixed (different ethnicities)

(6) Mixed, cannot estimate the proportion

(7) Greater than 60% African

For the following ethnicities enter number of participants for each category (ONLY PARTICIPANTS IN THE STUDY)

AFRICAN AMERICAN enter number of participants
CAUCASIAN / EUROPEAN-AMERICAN enter number of participants
HISPANIC enter number of participants
ASIAN enter number of participants
AMERICAN / INDIAN enter number of participants
SOMALI enter number of participants
PACIFIC ISLANDER enter number of participants
MULTIRACIAL enter number of participants
FREE REDUCED LUNCH
  1. Majority of participants 80% or more
  2. From 50%-80% receive free reduced lunch
  3. Less than 50%
TITLE ONE SCHOOL
  (0) Not reported
  (1) YES
  (2) No
GENDER
FEMALE enter number of females
MALES enter number of males
Number of English Language Learners
(0) not reported
ELLs Number of English Language Learners
NONELLs Number of non-ELL students
LANG Predominant language(s) of ELLs (enter the most predominant language(s))
  1 Spanish
  2 Spanish Dialect
  3 Vietnamese
  4 Somali
  5 Sudanese
  6 Cambodian
  7 Laotian
  8 Italian
  9 Hmong
 10 Tagalog
 11 Native American languages
 12 Japanese
 13 Gujarati
 14 Arabic
 15 Bangladesh
 16 Portuguese
 17 Polish
 18 Syrian
19 Urdu
Enter number of ELLs for each language
0  Not disaggregated by language
1  Spanish
   Total  enter total number of students
   FEMALE  enter number of females
   MALE  enter number of males
2  Spanish Dialect
   Total  enter total number of students
   FEMALE  enter number of females
   MALE  enter number of males
3  Vietnamese
   Total  enter total number of students
   FEMALE  enter number of females
   MALE  enter number of males
4  Somalian
   Total  enter total number of students
   FEMALE  enter number of females
   MALE  enter number of males
5  Sudanese
   Total  enter total number of students
   FEMALE  enter number of females
   MALE  enter number of males
6  Cambodian
   Total  enter total number of students
   FEMALE  enter number of females
   MALE  enter number of males
7  Laotian
   Total  enter total number of students
   FEMALE  enter number of females
   MALE  enter number of males
8  Italian
   Total  enter total number of students
   FEMALE  enter number of females
   MALE  enter number of males
9  Hmong
   Total  enter total number of students
   FEMALE  enter number of females
   MALE  enter number of males
10 Tagalog
   Total  enter total number of students
   FEMALE  enter number of females
   MALE  enter number of males
11  Native American languages
Total  enter total number of students
FEMALE  enter number of females
MALE  enter number of males
12  Japanese
Total  enter total number of students
FEMALE  enter number of females
MALE  enter number of males
13  Gujarati
Total  enter total number of students
FEMALE  enter number of females
MALE  enter number of males
14  Arabic
Total  enter total number of students
FEMALE  enter number of females
MALE  enter number of males
15  Bangladeshi
Total  enter total number of students
FEMALE  enter number of females
MALE  enter number of males
16  Portuguese
Total  enter total number of students
FEMALE  enter number of females
MALE  enter number of males
17  Polish
Total  enter total number of students
FEMALE  enter number of females
MALE  enter number of males
18  Syrian
Total  enter total number of students
FEMALE  enter number of females
MALE  enter number of males
19  Urdu
Total  enter total number of students
FEMALE  enter number of females
MALE  enter number of males

LANGPROFTEST  (Language Proficiency Test)
1. California English Learner's Diagnostic Test (CELDT)
2. Comprehensive English Language Learning Assessment (CELLA)
3. Colorado English Language Assessment (CELA)
4. IDEA Proficiency Oral Language Test
5. World-Class Instructional Design and Assessment  ACCESS Placement Test
6. Other

Levels of Proficiency

120
Beginning enter number of students with this level of proficiency
Early Intermediate enter number of students with this level of proficiency
Intermediate enter number of students with this level of proficiency
Proficient enter number of students with this level of proficiency
Advanced enter number of students with this level of proficiency
GRADE  (Participants Grade Levels)
(0) not reported
(1) kindergarten
(2) Grade 1
(3) Grade 2
(4) Grade 3
(5) Grade 4
(6) Grade 5
(7) Grade 6
(8) Grade 7
(9) Grade 8
MEANAGE  enter mean age of students, if reported
AGERANGE  enter age range of students (e.g., 5-7), if reported
MEASURES
(1) Comprehensive Test of Phonological Processing (CTOPP)
(1.10) CTOPP Phonological Processing Composite Score
(1.11) CTOPP Elision subtest
(1.12) CTOPP Blending Words
(1.13) CTOPP Sound Matching
(1.14) CTOPP Phoneme Isolation
(1.15) CTOPP Blending Nonwords
(1.16) CTOPP Segmenting Nonwords
(1.20) CTOPP Phonological Memory Composite Score
(1.21) CTOPP Memory for Digits
(1.22) CTOPP Nonword Repetition
(1.30) CTOPP Rapid Symbolic Naming Composite Score
(1.31) CTOPP Rapid Digit Naming
(1.32) CTOPP Rapid Letter Naming

(1.40) CTOPP Rapid Non-Symbolic Naming Composite Score

(1.41) CTOPP Rapid Color Naming

(1.42) CTOPP Rapid Object Naming
(1.43) CTOPP Letter Naming
(1.44) CTOPP Letter Sound Identification

(2) **Dynamic Indicators of Basic Early Literacy Skills (DIBELS)**
(2.10) DIBELS Composite
(2.11) DIBELS Letter-Naming Fluency
(2.12) DIBELS Initial Sound Fluency

(2.13) DIBELS Correct Letter Sound Fluency

(2.14) DIBELS Phoneme Segmentation Fluency

(2.15) DIBELS Nonsense Word Fluency

(2.16) DIBELS Oral Reading Fluency (ORF)

(2.17) DIBELS Retell Fluency (RTF)

(2.18) DIBELS Daze
(2.19) DIBELS Word Use Fluency

(3) **Woodcock-Munoz Language Survey-Revised (WMLS-R English)**
(3.10) WMLS-R Broad English Ability
(3.20) WMLS-R Oral Language Cluster
(3.21) WMLS-R Picture Vocabulary
(3.22) WMLS-R Verbal Analogies
(3.30) WMLS-R Reading-Writing Cluster
(3.31) WMLS-R Letter Word Identification
(3.32) WMLS-R Dictation
(3.33) WMLS-R Passage Comprehension
(3.40) WMLS-R Reading Cluster
(3.50) WMLS-R Language Comprehension Cluster
(3.51) WMLS-R Understanding Directions
(3.52) WMLS-R Story Recall

(4) **Woodcock Johnson-III (WJ-III)**
(4.01) WJ-III Word Attack
(4.02) WJ-III Passage Comprehension
(4.03) WJ-III Nonsense Word Fluency
(4.04) WJ-III Listening Comprehension
(4.05) WJ-III Letter-Word Identification
(5) **Woodcock Language Proficiency Battery -Revised (WLPB-R)**
(5.10) WLPB-R Broad English Ability
(5.20) WLPB-R Oral Language Cluster
(5.21) WLPB-R Picture Vocabulary
(5.22) WLPB-R Memory for Sentences
(5.23) WLPB-R Listening Comprehension
(5.24) WLPB-R Oral Vocabulary
(5.25) WLPB-R Verbal Analogies
(5.30) WLPB-R Reading Cluster
(5.31) Letter-Word Identification
(5.32) Word Attack
(5.33) Passage Comprehension
(5.34) Reading Vocabulary

(6) **PALS**
(6.01) PALS Phonemic Awareness Segmentation
(6.02) PALS Phonemic Awareness Blending
(6.03) PALS Rapid Letter Naming (RNL)
(6.04) PALS Rapid Letter Sound (RLS)

(7) **Woodcock Reading Mastery Test Revised (WRMT-R)**
(7.10) WRMT-R Total Reading
(7.01) WRMT-R Word Identification
(7.02) WRMT-R Word Attack
(7.03) WRMT-R Passage Comprehension

(8.0) **AIMS WEB**
(8.01) Phoneme Segmentation Fluency (PSF)
(8.02) Nonsense Word Fluency (NWF)

(9.0) **KPALS**
(9.01) Oral Reading Fluency A
(9.02) Oral Reading Fluency B

(10) **California Standardized Testing (CST)**
10.10 AYP Scale Composite

(11) **Gray Oral Reading Test-4 (GORT-4)**
11.10 GORT-4 Composite

(12) **Leveled Literacy Intervention Benchmark (LLI)**
(12.10) LLI Benchmark Composite

(13) **STAR Reading assessment**
(13.10) STAR Reading composite

(14) **Criterion-Referenced Competency Tests (Georgia Performance Standards)**
(14.10) CRCT Reading Composite

(15) **Receptive One-Word Picture Vocabulary Test**
(15.10) ROWPVT Receptive Composite

(16) **Wide Range Achievement Test, Third Edition (WRAT-3)**
16.10 WRAT-3 Reading Composite

123
(17) Tennessee Comprehensive Assessment Program (TCAP)
(17.10) TCAP Reading Composite

(18) Reading Proficiency Assessment (RPA) (in lieu of the DRA scores)
18.10 RPA Composite

(19) Florida Assessment for Instruction in Reading (FAIR)
(19.10) FAIR Composite

COMPONENT Enter reading component(s) measured by reading measure composite, cluster, and/or subtest (Use Appendix C Measures and Reading Components to code the component measured by reading measures)

   (1) Phonological Awareness
   (2) Phonics
   (3) Fluency
   (4) Comprehension
   (5) Vocabulary
   (6) Oral Language
   (7) Phonological Memory

TYPEMEAS (Type of Measure)
(0) Not reported
(1) Standardized
   (A) norm-referenced
   (B) criterion-referenced
   (C) domain-referenced
   (D) standards-based
   (E) curriculum-based
(2) Researcher or Professionally Developed
   (A) not reported
   (B) not based on state or standardized assessment
   (C) based on state or standardized assessment

(3) District Created Assessment
(4) Formative Assessment

RELIABILITY (Assessment reliability reported?)
(0) not reported
(1) yes
(2) published test/can find online

Reliability Type
(0) not reported
(1) coefficient stability (test-retest)
(2) coefficient of equivalence (alternate form)
(3) coefficient of stability and equivalence
(4) internal consistency
   (A) Cronbach’s alpha
   (B) Spearman rho
   (C) KR20
(5) criterion reliability
(6) Inter-rater
Reliability Index (list value)
Assessment Validity reported
   (0) no
   (1) yes
   (2) published test/can find online
Validity Type
   (0) not reported
   (1) Cronbach’s alpha
   (2) Spearman rho
   (3) Split-half
   (4) Factor Analysis
   (5) Correlational
   (6) Criterion Related
   (7) Predictive Validity
   (8) Content
Validity Coefficient (enter value)

**Intervention**
Tier(s) of Intervention (TIERS) (Is the reading intervention Tier I, Tier II or/and Tier III?) If the researcher does not specify see definitions.
   (1) Tier I
   (2) Tier II
   (3) Tier III
CORE Curriculum (CORE) (Enter the type of reading curriculum implemented by school or school district)
   (1) Houghton Mifflin’s Language Arts Curriculum
   (2) Moving into English
   (3) Balanced Literacy Instruction
   (4) Open Court
   (5) California Treasures
   (6) Literacy Across Columbus Elementary Schools (LACES)
   (7) Trophies
   (8) Tennessee Reading Curriculum (Tennessee Department of Education, 2008)
(9) Language Enrichment (Carreker)
(10) McGraw Hill Reading
(11) 90-minute literacy block (does not specify core curriculum)

Intervention (INTERV) Enter code(s) for interventions
Tier I intervention (TIER1INT) Enter code for Tier 1 intervention
Tier II intervention (TIER2INT) Enter code for Tier 2 intervention
Tier III intervention (TIER3INT) Enter code for Tier 3 intervention

For example, enter 1 for Tier I intervention if the study used Proactive Reading and 8 for Tier II intervention if the study used Supplemental Reading Intervention.

(1) Proactive reading

(2) Peer Assisted Learning Strategies (PALS)

(3) Early Reading Intervention (ERI)

(4) Balanced literacy intervention

(5) Direct instruction approach (Reading Mastery, Early Interventions in Reading, Read Well and Read Naturally)

(6) Leveled Literacy Intervention (LLI)

(7) Targeted Reading Intervention

(8) Supplemental Reading Intervention (Oral Language)

(9) Burst Early Literacy Intervention

(10) Wilson Fundation

(11) Houghton Mifflin’s Kindergarten Curriculum

(12) Moving into English

(13) Kindergarten Peer Assisted Learning Strategies (KPALS)

(14) Tier II type literacy practices

(15) Tier II-Reading Fluency and Comprehension (Does not specify)

(16) Tier I Language Arts (does not specified)

(17) Tier I Open Court Curriculum
(18) Tier II-Corrective Reading or Rewards

(19) Tier II Read Naturally

(20) Tier II Daybook

(21) Sounds and Letters for Readers and Spellers with token economy

(22) Tier I Harcourt Trophies

(23) Tier II Ladders to Literacy-Kindergarten

(24) Tier II Sound Partners-First Grade

(25) Constructed fluency activity

(26) Sing, Spell, Read, Write phonics curriculum (SSRW)

(27) Literacy Enhancement Hierarchical Phonological Awareness/Word Recognition programs

(28) Oral Narrative Enhancement Hierarchical Phonological Awareness/Word Recognition programs

(29) Culturally Relevant Oral Narrative Enhancement Group with Language Experience Approach

(30) Tier III- Reading Mastery I/II Fast Cycle

(31) Corrective Reading by Engelmann

(32) Phonological and Strategy Training (PHAST)

(33) Phonological Analysis and Blending/Direct Instruction (PHAB/DI)

(34) Phonological and Strategy Training (PHAST) Decoding Program

(35) Sidewalk (Scott Foresman)

(36) Core Intervention Model (CIM)
(37) Tier 3 Listening Comprehension intervention (Solari and Gerber)

(38) Tier I Success Marker by Pearson Learning

(39) Tier I English-Only ESL

(40) Tier I Bilingual ESL model

Essential Reading Components indicate what components the intervention(s) addressed

<table>
<thead>
<tr>
<th>Component</th>
<th>Addressed</th>
</tr>
</thead>
<tbody>
<tr>
<td>INTERVPA</td>
<td>phonological awareness</td>
</tr>
<tr>
<td>(0) No</td>
<td>(1) Yes</td>
</tr>
<tr>
<td>INTERVPH</td>
<td>phonics</td>
</tr>
<tr>
<td>(0) No</td>
<td>(1) Yes</td>
</tr>
<tr>
<td>INTERVFL</td>
<td>fluency</td>
</tr>
<tr>
<td>(0) No</td>
<td>(1) Yes</td>
</tr>
<tr>
<td>INTERVCOMP.</td>
<td>comprehension</td>
</tr>
<tr>
<td>(0) No</td>
<td>(1) Yes</td>
</tr>
<tr>
<td>INTERVVOCAB</td>
<td>vocabulary</td>
</tr>
<tr>
<td>(0) No</td>
<td>(1) Yes</td>
</tr>
<tr>
<td>INTERVOL</td>
<td>oral language</td>
</tr>
<tr>
<td>(0) No</td>
<td>(1) Yes</td>
</tr>
</tbody>
</table>

Delivered by (DELIVEREDBY)
(1) Teachers/school personnel
(2) Researchers
(3) Graduate students
(4) University Staff
(5) Instructional assistants
(6) Undergraduate students
(7) Research assistants
(8) Paraprofessionals
(9) Special education teachers

Number of Personnel Delivering Intervention (NUMBER) Enter number

Level of Education of Personnel Delivering Intervention (LEVELED)

High School (LEVELEDHS) Enter number of individuals with this degree
Associate degree (LEVELEDAA) Enter number of individuals with this degree
Bachelors (LEVELEDBA) Enter number of individuals with this degree
Masters (LEVELEDMA) Enter number of individuals with this degree
Masters plus 30, Ed.S., Doctorate (LEVELEDMAPLUS30) Enter number of individuals with this degree
Years of Experience of Personnel Delivering the Intervention (YEARSEXPAVERAGE) Enter average of years of experience
Years of Experience of Personnel Delivering the Intervention range (YEARSEXPRANGE) Enter range of years of experience
Training Specify if staff received training to deliver the intervention
(0) not reported
(1) Yes
Fidelity of the Intervention (FIDELITY) (Does the study included observations and other procedures to ensure the personnel delivered the intervention with fidelity?)
(0) not reported
(1) Yes
Length of the Intervention (How many school days did the intervention last?) If the study reports one school year enter 180 days (1 school year = 180 instructional) If the study reports September – April enter 135 instructional days If the study reports number of months, multiply number of months by 20 instructional days
LENGTHINDAYSTIER1 Enter number of days
LENGTHINDAYSTIER2 Enter number of days
LENGTHINDAYSTIER3 Enter number of days
Duration of Intervention by Tier
TIER1DURATION Enter minutes of intervention daily
TIER2DURATION Enter minutes of intervention daily
TIER3DURATION Enter minutes of intervention daily
Type of group used to deliver the intervention for each tier
Type of group for Tier I (TYPEGROUPT1)
(1) Small group (2-7 students)
(2) One-on-one
(3) Whole classroom
Type of group for Tier II (TYPEGROUPT2)
(1) Small group (2-7 students)
(2) One on One
(3) Whole classroom
Type of group for Tier III (TYPEGROUPT3)
(1) Small group (2-6 students)
(2) One on One
(3) Whole class
Research Methodology
Type of study (TYPESTUDY)
(0) Not reported
(1) Experimental
(2) Quasi-experimental
(3) Post hoc (e.g., causal comparative design)

ANALYSIS (Type of Statistical Analysis)
(0) Descriptive Statistics (mean, s.d., n)

(1) t-test
(2) F-test
(3) Chi-square
(4) ANOVA
(5) ANCOVA (use adjusted means)
(6) Multiple Regression (use unstandardized regression coefficient, $\beta$)
(7) Effect Size (Cohen’s $d$)
(8) MANOVA

ASSIGN (Type of assignment to conditions)
(1) Random after matching, stratification, blocking, etc.
(2) Random simple (includes systematic sampling)
(3) Nonrandom (post hoc, matching)
(4) Nonrandom (other)
(5) Other

Research Results
TOTALN
PRETXN
PRETXMEN
PRETXSD
PRETXN2
PRETXMEN2
PRETXSD2
PRETXN3
PRETXMEN3
PRETXSD3
PRECGN
PRECGMEN
PRECGSD

enter sample size
Pretest Treatment Sample Size
Pretest Treatment Group Mean
Pretest Treatment Group Standard Deviation
Pretest Treatment Sample Size Group 2 (studies with two treatment groups)
Pretest Treatment Group 2 Mean
Pretest Treatment Group 2 Standard Deviation
Pretest Treatment Sample Size Group 3 (studies with three treatment groups)
Pretest Treatment Group 3 Mean
Pretest Treatment Group 3 Standard Deviation
Pretest Control Group Sample Size
Pretest Control Group Mean
Pretest Control Group Standard Deviation
PRECGN2  Pretest Control Group 2 Sample Size
PRECGMEAN2 Pretest Control Group 2 Mean
PRECGSD2  Pretest Control Group 2 Standard Deviation
PRENONELLTXN Pretest Non-ELL Treatment Group Sample Size
PRENONELLTXMEAN Pretest Non-ELL Treatment Group Mean
PRENONELLTXSD  Pretest Non-ELL Treatment Group Standard Deviation
PRENONELLCGN Pretest Non-ELL Control Group Sample Size
PRENONELLCGMEAN Pretest Non-ELL Control Group Mean
PRENONELLCGSD  Pretest Non-ELL Control Group Standard Deviation
POSTTXN  Posttest Treatment Sample Size
POSTTXXMEAN  Posttest Treatment Group Mean
POSTTXSD  Posttest Treatment Group Standard Deviation
POSTTXN2   Posttest Treatment Group 2 Sample Size
POSTTXXMEAN2 Posttest Treatment Group 2 Mean
POSTTXXSD2 Posttest Treatment Group 2 Standard Deviation
POSTTXN3   Posttest Treatment Group 3 Sample Size
POSTTXXMEAN3 Posttest Treatment Group 3 Mean
POSTTXSD3  Posttest Treatment Group 3 Standard Deviation
POSTCGN  Posttest Control Group Sample Size
POSTCGMEAN Posttest Control Group Mean
POSTCGSD  Posttest Control Group Standard Deviation
POSTCGN2 Posttest Control Group 2 Sample Size
POSTCGMEAN2 Posttest Control Group 2 Mean
POSTCGSD2 Posttest Control Group 2 Standard Deviation
POSTNONELLTXN Posttest Non-ELL Treatment Group Sample Size
POSTNONELLTXMEAN Posttest Non-ELL Treatment Group Mean
POSTNONELLTXSD Posttest Non-ELL Treatment Group Standard Deviation
POSTNONELLCGN Posttest Non-ELL Control Group Sample Size
POSTNONELLCGMEAN Posttest Non-ELL Control Group Mean
POSTNONELLCGSD Posttest Non-ELL Control Group Standard Deviation
CONTRASTCOEFFICIENTTXNONELLSELLs Enter contrast coefficient for treatment groups ELLs versus non-ELLs
SDpTXNONELLSSELLS Pooled Standard Deviation for Treatment Non-ELLs versus ELLs
dTXXNonELLS-ELLS Effect size Treatment Groups ELLs versus Non-ELLs
ELLsN Sample Size for ELL group
NON-ELLSN Sample size for Non-ELL group
P-VALUE Enter p value
F-VALUE Enter F value
DF Enter degrees of freedom
<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ESCohen’s $d$</td>
<td>Effect size Cohen’s $d$</td>
</tr>
<tr>
<td>Std.ERROR</td>
<td>Enter standard error</td>
</tr>
<tr>
<td>TXStandardScoreChangeN</td>
<td>enter treatment group sample size for standard score change</td>
</tr>
<tr>
<td>TXMEANStandardScoreChange</td>
<td>enter treatment group mean for standard score change</td>
</tr>
<tr>
<td>TXSDStandardScoreChange</td>
<td>enter treatment group standard deviation for standard score change</td>
</tr>
<tr>
<td>CGStandardScoreChangeN</td>
<td>enter control group sample size for standard score change</td>
</tr>
<tr>
<td>CGMEANStandardScoreChange</td>
<td>enter control group mean for standard score change</td>
</tr>
<tr>
<td>CGSDStandardScoreChange</td>
<td>enter control group standard deviation for standard score change</td>
</tr>
<tr>
<td>TXNt-test</td>
<td>enter treatment group sample size for t-test</td>
</tr>
<tr>
<td>NONELLStxNt-test</td>
<td>enter non-ELL treatment group sample size for t-test</td>
</tr>
<tr>
<td>tscore</td>
<td>enter t-score</td>
</tr>
<tr>
<td>df</td>
<td>enter degrees of freedom</td>
</tr>
<tr>
<td>p-value</td>
<td>enter p-value</td>
</tr>
<tr>
<td>exact p-value</td>
<td>enter exact p-value (use statistical calculator to calculate the exact p value)</td>
</tr>
<tr>
<td>TXSUCCESSN</td>
<td>enter treatment group sample size for percentage of success</td>
</tr>
<tr>
<td>TXSUCCESS</td>
<td>% of treatment group with successful outcome (enter percentage of students that met the grade level benchmark for a specific measure)</td>
</tr>
<tr>
<td>TX2SUCCESSN</td>
<td>enter treatment group2 sample size for percentage of success</td>
</tr>
<tr>
<td>TX2SUCCESS</td>
<td>% of treatment group 2 with successful outcome (enter percentage of students that met the grade level benchmark for a specific measure)</td>
</tr>
<tr>
<td>CGSUCCESSN</td>
<td>enter control group sample size for percentage of success</td>
</tr>
<tr>
<td>CGSUCCESS</td>
<td>% of control group with successful outcome (enter percentage of students that met the grade level benchmark for a specific measure)</td>
</tr>
<tr>
<td>CG2SUCCESSN</td>
<td>enter control group 2 sample size for percentage of success</td>
</tr>
<tr>
<td>CG2SUCCESS</td>
<td>% of control group 2 with successful outcome (enter percentage of students that met the grade level benchmark for a specific measure)</td>
</tr>
<tr>
<td>TXNPROP</td>
<td>Treatment group sample size for proportion of students eligible for special education</td>
</tr>
<tr>
<td>TXPROP</td>
<td>Proportion (number) of students eligible for special education for the treatment group</td>
</tr>
<tr>
<td>TX2NPROP</td>
<td>Treatment Group 2 Sample size for Proportion of students eligible for special education</td>
</tr>
<tr>
<td>TX2PROP</td>
<td>Proportion (number) of students eligible for special education for the treatment group2</td>
</tr>
<tr>
<td>CGNPROP</td>
<td>Control Group Sample size for Proportion of students eligible for special education</td>
</tr>
<tr>
<td>CGPROP</td>
<td>Proportion (number) of students eligible for special education for the control group</td>
</tr>
</tbody>
</table>
Notes: Enter relevant information regarding the demographics, groups, measures, intervention, and results including page number and tables used to code means, standard deviations or other statistical data.
### APPENDIX C

**Measures and Reading Components**

<table>
<thead>
<tr>
<th>Comprehensive Test of Phonological Processing (CTOPP)</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Phonological Processing Composite</strong></td>
<td>PA, PH</td>
</tr>
<tr>
<td>Elision Subtest</td>
<td>PA</td>
</tr>
<tr>
<td>Blending Words</td>
<td>PA</td>
</tr>
<tr>
<td>Sound Matching</td>
<td>PH</td>
</tr>
<tr>
<td>Phoneme Isolation</td>
<td>PH</td>
</tr>
<tr>
<td>Blending Nonwords</td>
<td>PA</td>
</tr>
<tr>
<td>Segmenting Nonwords</td>
<td>PA</td>
</tr>
<tr>
<td><strong>Phonological Memory Composite</strong></td>
<td>Phonological Memory</td>
</tr>
<tr>
<td>Memory for Digits</td>
<td>Phonological Memory</td>
</tr>
<tr>
<td>Nonword Repetition</td>
<td>Phonological Memory</td>
</tr>
<tr>
<td><strong>Rapid Symbolic Naming Composite</strong></td>
<td>PH</td>
</tr>
<tr>
<td>Rapid Digit Naming</td>
<td>31</td>
</tr>
<tr>
<td>Rapid Letter Naming</td>
<td>PH</td>
</tr>
<tr>
<td><strong>Rapid Non-Symbolic Naming Composite</strong></td>
<td>Phonological Memory</td>
</tr>
<tr>
<td>Rapid Color Naming</td>
<td>Phonological Memory</td>
</tr>
<tr>
<td>Rapid Object Naming</td>
<td>Phonological Memory</td>
</tr>
<tr>
<td>Letter Naming</td>
<td>PH</td>
</tr>
<tr>
<td>Letter Sound Identification</td>
<td>PH</td>
</tr>
<tr>
<td><strong>Dynamic Indicators of Basic Early Literacy Skills (DIBELS)</strong></td>
<td>2</td>
</tr>
<tr>
<td>Composite</td>
<td>PA, PH, F</td>
</tr>
<tr>
<td>Letter Naming Fluency</td>
<td>PH</td>
</tr>
<tr>
<td>Initial Sound Fluency</td>
<td>PA</td>
</tr>
<tr>
<td>First Sound Fluency</td>
<td>PH</td>
</tr>
<tr>
<td>Phoneme Segmentation Fluency</td>
<td>PA</td>
</tr>
<tr>
<td>Nonsense Word</td>
<td>PH, Alphabetic</td>
</tr>
<tr>
<td>Fluency</td>
<td>Principle</td>
</tr>
<tr>
<td>---------------------------------</td>
<td>-----------</td>
</tr>
<tr>
<td>Oral Reading Fluency</td>
<td>F</td>
</tr>
<tr>
<td>Retell Fluency</td>
<td>C</td>
</tr>
<tr>
<td>DAZE</td>
<td>C</td>
</tr>
<tr>
<td>Word Use Fluency</td>
<td>V, OL</td>
</tr>
<tr>
<td><strong>Woodcock-Munoz Language Survey-Revised (WMLS-R English)</strong></td>
<td></td>
</tr>
<tr>
<td>Broad English Ability</td>
<td>PH, V, C, OL</td>
</tr>
<tr>
<td>WMLS-R Oral Language Cluster</td>
<td>V, OL</td>
</tr>
<tr>
<td>Picture Vocabulary</td>
<td>V, OL</td>
</tr>
<tr>
<td>Verbal Analogies</td>
<td>V, OL</td>
</tr>
<tr>
<td><strong>Reading-Writing Cluster</strong></td>
<td></td>
</tr>
<tr>
<td>Letter Word Identification</td>
<td>PH</td>
</tr>
<tr>
<td>Dictation</td>
<td>PH</td>
</tr>
<tr>
<td>Passage Comprehension</td>
<td>C</td>
</tr>
<tr>
<td><strong>Reading Cluster</strong></td>
<td></td>
</tr>
<tr>
<td>Language Comprehension Cluster</td>
<td>OL</td>
</tr>
<tr>
<td>Understanding Directions</td>
<td>OL</td>
</tr>
<tr>
<td>Story Recall</td>
<td>OL</td>
</tr>
<tr>
<td><strong>Woodcock Johnson-III (WJ-III)</strong></td>
<td></td>
</tr>
<tr>
<td>Word Attack</td>
<td>PH</td>
</tr>
<tr>
<td>Passage Comprehension</td>
<td>C</td>
</tr>
<tr>
<td>Nonsense Word Fluency</td>
<td>PH, F</td>
</tr>
<tr>
<td>Listening Comprehension</td>
<td>OL</td>
</tr>
<tr>
<td><strong>Woodcock Language Proficiency Battery (WLPB-R)</strong></td>
<td></td>
</tr>
<tr>
<td>Broad English Ability</td>
<td></td>
</tr>
<tr>
<td>Oral Language Cluster</td>
<td>OL</td>
</tr>
<tr>
<td>Picture Vocabulary</td>
<td>OL</td>
</tr>
<tr>
<td>Memory for</td>
<td>OL</td>
</tr>
<tr>
<td>Test</td>
<td>Code(s)</td>
</tr>
<tr>
<td>-------------------------------------------</td>
<td>---------</td>
</tr>
<tr>
<td>Sentences</td>
<td></td>
</tr>
<tr>
<td>Listening Comprehension</td>
<td>OL</td>
</tr>
<tr>
<td>Oral Vocabulary</td>
<td>V, OL</td>
</tr>
<tr>
<td>Verbal Analogies</td>
<td>V, OL</td>
</tr>
<tr>
<td><strong>Reading Cluster</strong></td>
<td>PH, PA, C, V</td>
</tr>
<tr>
<td>Letter-Word Identification</td>
<td>PH</td>
</tr>
<tr>
<td>Word Attack</td>
<td>PH, PA</td>
</tr>
<tr>
<td>Passage Comprehension</td>
<td>C</td>
</tr>
<tr>
<td>Reading Vocabulary</td>
<td>V</td>
</tr>
<tr>
<td><strong>Peer Assisted Learning Strategy (PALS)</strong></td>
<td></td>
</tr>
<tr>
<td>Phonemic Awareness Segmentation</td>
<td>PA</td>
</tr>
<tr>
<td>Phonemic Awareness Blending</td>
<td>PA</td>
</tr>
<tr>
<td>Rapid Letter Naming (RNLS)</td>
<td>PH</td>
</tr>
<tr>
<td>Rapid Letter Sound (RLS)</td>
<td>PH</td>
</tr>
<tr>
<td><strong>Woodcock Reading Mastery Test-Revised (WRMT-R)</strong></td>
<td></td>
</tr>
<tr>
<td>Total Reading</td>
<td>PA, PH, C</td>
</tr>
<tr>
<td>Word Identification</td>
<td>PH</td>
</tr>
<tr>
<td>Word Attack</td>
<td>PA</td>
</tr>
<tr>
<td>Passage Comprehension</td>
<td>C</td>
</tr>
<tr>
<td><strong>AIMSWeb</strong></td>
<td></td>
</tr>
<tr>
<td>Phoneme Segmentation Fluency</td>
<td>PA</td>
</tr>
<tr>
<td>Nonsense Word Fluency (NWF)</td>
<td>PH, Alphabetic Principle</td>
</tr>
<tr>
<td><strong>Kindergarten Peer Assisted Learning Strategy (K-PALS)</strong></td>
<td></td>
</tr>
<tr>
<td>Oral Reading Fluency A</td>
<td>F</td>
</tr>
<tr>
<td>Oral Reading Fluency B</td>
<td>F</td>
</tr>
<tr>
<td><strong>California Standardized Testing (CST)</strong></td>
<td></td>
</tr>
<tr>
<td>AYP Scale Composite</td>
<td>PH, F, C, V</td>
</tr>
<tr>
<td>Test</td>
<td>Composite -</td>
</tr>
<tr>
<td>----------------------------------------------------------------------</td>
<td>-------------</td>
</tr>
<tr>
<td>Gray Oral Reading Test-4 (GORT-4)</td>
<td>11</td>
</tr>
<tr>
<td>GORT-4 Composite</td>
<td>10</td>
</tr>
<tr>
<td>Leveled Literacy Intervention Benchmark (LLI)</td>
<td>12</td>
</tr>
<tr>
<td>LLI Benchmark Composite</td>
<td>10</td>
</tr>
<tr>
<td>Star Reading Benchmark</td>
<td>13</td>
</tr>
<tr>
<td>STAR Reading Composite</td>
<td>10</td>
</tr>
<tr>
<td>Criterion Reference Competency Tests (CRCT) Georgia Performance Test</td>
<td>14</td>
</tr>
<tr>
<td>CRCT Reading Composite</td>
<td>10</td>
</tr>
<tr>
<td>Receptive One-Word Picture Vocabulary Test</td>
<td>15</td>
</tr>
<tr>
<td>ROWPVT Receptive Composite</td>
<td>10</td>
</tr>
<tr>
<td>Wide Range Achievement Test (WRAT-3)</td>
<td>16</td>
</tr>
<tr>
<td>WRAT-3 Composite</td>
<td>10</td>
</tr>
<tr>
<td>Tennessee Comprehensive Assessment Program (TCAP)</td>
<td>17</td>
</tr>
<tr>
<td>TCAP Reading Composite</td>
<td>10</td>
</tr>
<tr>
<td>Reading Proficiency Assessment (RPA)</td>
<td>18</td>
</tr>
<tr>
<td>RPA Composite</td>
<td>10</td>
</tr>
<tr>
<td>Florida Assessment for Instruction in Reading (FAIR)</td>
<td>19</td>
</tr>
<tr>
<td>FAIR Composite</td>
<td>10</td>
</tr>
</tbody>
</table>

**KEY**
- PA = Phonological Awareness
- PH = Phonics
- F = Fluency
- C = Comprehension
- V = Vocabulary
- OL = Oral Language
## APPENDIX D

### Preliminary Coding

<table>
<thead>
<tr>
<th>Total of Studies n= 20</th>
<th>Tier I = 6</th>
<th>Tier II = 11</th>
<th>Tier III = 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tier I</td>
<td>Tier II</td>
<td>Tier III</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

#### K-2

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Grade</td>
<td>Tier I &amp; Tier II Grade 1; Ph; S, E</td>
<td>Grade 1; PA; S, V</td>
</tr>
<tr>
<td>Note:</td>
<td>Retrieved from: Academic Search Complete</td>
<td>Retrieved from Academic Search Complete</td>
</tr>
<tr>
<td>Asian/Indian, African American</td>
<td>Retrieved from: ERIC</td>
<td>Retrieved from ERIC</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>McMaster (2008). Grade K; PA; S, O</th>
<th>Dougherty, Keane &amp; Sismic (2012) Grade 1; Ph; S, E</th>
</tr>
</thead>
<tbody>
<tr>
<td>Note:</td>
<td>Retrieved from: ERIC</td>
<td>Retrieved from: Academic Search Complete</td>
</tr>
<tr>
<td>Asian/Indian, African American</td>
<td>Retrieved from ERIC</td>
<td>Retrieved from ERIC</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Kamps, et al., (2007) Grade 1, 2; PA, F, C; S, E, So, Su, V</th>
<th>Ransford-Kaldon, Sutton Flynt, &amp; Ross (2011) K-2; PA, Ph, V, CO, OL; E, S</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Retrieved from: ERIC</td>
<td>Retrieved from SREE</td>
</tr>
</tbody>
</table>

Eversole (2011). Tier I, II & III Grade 2; S; Note: all Latino Intervention: State-approved language arts w/classroom intervention
| Retrieved from | Pieretti (2011)  
Grade; PA, ON, WR; H  
Retrieved from Proquest Dissertations & Theses |
|----------------|----------------------------------------------------------------------------------|
Tier I, II & III  
Grade 2; F, C; S;  
Note: all Latino  
Retrieved from AERA & PsycINFO |
|  
Richards-Tutor. (2012)  
Grade K; PA, Ph; S  
Pre-Post  
No control group  
Retrieved from Summons  
Cited by Hans |
<table>
<thead>
<tr>
<th>Meta-analysis (2009) DIBELS</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>O’Connor (2014). Grades K-2; PA, PH, F, V, C; S</td>
<td>Retrieved from PsycINFO</td>
</tr>
<tr>
<td>Koureau. 2007. Grade 1; PA, F; S, So, E</td>
<td>Retrieved from Proquest Dissertations</td>
</tr>
<tr>
<td>Linan-Thompson (2007). Grades 1, 2; PA, F, C; S</td>
<td>PsycINFO</td>
</tr>
<tr>
<td>Year</td>
<td>Authors</td>
</tr>
<tr>
<td>------</td>
<td>---------</td>
</tr>
<tr>
<td>2012</td>
<td>Nguyen-Quang</td>
</tr>
<tr>
<td>2007</td>
<td>McIntosh</td>
</tr>
<tr>
<td>2011</td>
<td>Eversole</td>
</tr>
<tr>
<td>2011</td>
<td>Keita</td>
</tr>
<tr>
<td>2013</td>
<td>Sapienza</td>
</tr>
<tr>
<td>2014</td>
<td>Miller</td>
</tr>
<tr>
<td>2011</td>
<td>Eversole</td>
</tr>
<tr>
<td>Multi-Racial Retrieved from PsycINFO</td>
<td>Retrieved from PsycINFO</td>
</tr>
<tr>
<td>-------------------------------------</td>
<td>-------------------------</td>
</tr>
<tr>
<td>O’Connor. (2013). Grades 3,4; Ph, F, V, C; S Retrieved from PsycINFO</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Lovett. 2008. Grades 2-8; PA, C, F; S, P, Po, T, I, A, U Intervention Reading Mastery I/II Fast Cycle or Corrective Reading Toronto, Canada Peer reviewed Retrieved from ERIC</td>
</tr>
</tbody>
</table>

142
APPENDIX E

Tier I studies Post Data Only

Effect Sizes, Weight, Standard Error, & p values for each study

<table>
<thead>
<tr>
<th>Study name</th>
<th>Subgroup within study</th>
<th>Comparison</th>
<th>StdDiff Increase and 95% CI</th>
<th>Weight (Random)</th>
<th>StdDiff</th>
<th>Relative weight</th>
<th>Std Residual</th>
<th>Std Residual</th>
<th>P-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>McMaster 6.01 Post T</td>
<td>04.0.6.01</td>
<td>Post Data</td>
<td>0.687</td>
<td>3.17</td>
<td>1.15</td>
<td>0.25</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>McMaster 6.02 Post T</td>
<td>04.0.6.02</td>
<td>Post Data</td>
<td>0.648</td>
<td>3.18</td>
<td>1.03</td>
<td>0.30</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>McMaster 6.03 Post T</td>
<td>04.0.6.03</td>
<td>Post Data</td>
<td>0.085</td>
<td>3.35</td>
<td>0.75</td>
<td>0.45</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>McMaster 6.04 Post T</td>
<td>04.0.6.04</td>
<td>Post Data</td>
<td>0.082</td>
<td>3.22</td>
<td>0.67</td>
<td>0.41</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>McMaster 7.01 Post T</td>
<td>04.0.7.01</td>
<td>Post Data</td>
<td>0.108</td>
<td>3.35</td>
<td>-0.75</td>
<td>0.50</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>McMaster 7.02 Post T</td>
<td>04.0.7.02</td>
<td>Post Data</td>
<td>0.220</td>
<td>3.33</td>
<td>0.31</td>
<td>0.78</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sapienza 18.10 Post T</td>
<td>19.0.18.10</td>
<td>Post Data</td>
<td>0.532</td>
<td>13.30</td>
<td>0.21</td>
<td>0.03</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eversole Post 2 T I</td>
<td>06.3.10.10</td>
<td>Post Data</td>
<td>0.249</td>
<td>10.98</td>
<td>0.41</td>
<td>0.54</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eversole Post 3 T I</td>
<td>06.4.10.1</td>
<td>Post Data</td>
<td>0.330</td>
<td>17.76</td>
<td>0.16</td>
<td>0.87</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eversole Post 4 T I</td>
<td>06.5.10.1</td>
<td>Post Data</td>
<td>0.472</td>
<td>16.95</td>
<td>1.20</td>
<td>0.23</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eversole Post 5 T I</td>
<td>06.6.10.1</td>
<td>Post Data</td>
<td>0.110</td>
<td>15.44</td>
<td>1.54</td>
<td>0.12</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

-2.00 -1.00 0.00 1.00 2.00
APPENDIX F

Forest Plots of Effect Sizes for Students in Treatment and Control Groups Tier I

Studies

<table>
<thead>
<tr>
<th>Study name</th>
<th>Subgroup within study</th>
<th>Comparison</th>
<th>Std diff in means and 95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>McMaster 6.01 TX T I</td>
<td>04.0.6.01</td>
<td>Pre-Post TX</td>
<td>1.696</td>
</tr>
<tr>
<td>McMaster 6.02 TX T I</td>
<td>04.0.6.02</td>
<td>Pre-Post TX</td>
<td>1.619</td>
</tr>
<tr>
<td>McMaster 6.03 TX T I</td>
<td>04.0.6.03</td>
<td>Pre-Post TX</td>
<td>0.859</td>
</tr>
<tr>
<td>McMaster 6.04 TX T I</td>
<td>04.0.6.04</td>
<td>Pre-Post TX</td>
<td>2.128</td>
</tr>
<tr>
<td>McMaster 7.01 TX T I</td>
<td>04.0.7.01</td>
<td>Pre-Post TX</td>
<td>1.218</td>
</tr>
<tr>
<td>McMaster 7.02 TX T I</td>
<td>04.0.7.02</td>
<td>Pre-Post TX</td>
<td>1.185</td>
</tr>
<tr>
<td>Sapienza 18.10 TX T I</td>
<td>19.0.18.10</td>
<td>Pre-Post TX</td>
<td>1.404</td>
</tr>
<tr>
<td>Soong 19.10 TX T I</td>
<td>20.0.19.10</td>
<td>Pre-Post TX</td>
<td>1.054</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1.259</td>
</tr>
</tbody>
</table>

Forest Plot of Effect Sizes for Students in Treatment Groups Tier I Pre-Post Studies

<table>
<thead>
<tr>
<th>Study name</th>
<th>Subgroup within study</th>
<th>Comparison</th>
<th>Std diff in means and 95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>McMaster 6.01 CG T I</td>
<td>04.0.6.01</td>
<td>Pre-Post Cg</td>
<td>1.053</td>
</tr>
<tr>
<td>McMaster 6.02 CG T I</td>
<td>04.0.6.02</td>
<td>Pre-Post Cg</td>
<td>0.926</td>
</tr>
<tr>
<td>McMaster 6.03 CG T I</td>
<td>04.0.6.03</td>
<td>Pre-Post Cg</td>
<td>0.877</td>
</tr>
<tr>
<td>McMaster 6.04 CG T I</td>
<td>04.0.6.04</td>
<td>Pre-Post Cg</td>
<td>1.097</td>
</tr>
<tr>
<td>McMaster 7.01 CG T I</td>
<td>04.0.7.01</td>
<td>Pre-Post Cg</td>
<td>1.024</td>
</tr>
<tr>
<td>McMaster 7.02 CG T I</td>
<td>04.0.7.02</td>
<td>Pre-Post Cg</td>
<td>0.738</td>
</tr>
<tr>
<td>Sapienza 18.10 CG T I</td>
<td>19.0.18.10</td>
<td>Pre-Post Cg</td>
<td>0.694</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.860</td>
</tr>
</tbody>
</table>

Forest Plot of Effect Sizes for Students in Control Groups Tier I Pre-Post Studies

144
## APPENDIX G

### Tier II studies Post-Test Data Only

<table>
<thead>
<tr>
<th>Study name</th>
<th>Subgroup within study</th>
<th>Comparison</th>
<th>Statistics for each study</th>
<th>Weight (Random)</th>
<th>Std Residual</th>
<th>P-Val</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Std diff in means</td>
<td>Relative weight</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ganesh 7.03</td>
<td>00.0.7.00</td>
<td>Post Data</td>
<td>-0.390 (2.57)</td>
<td>-1.31</td>
<td></td>
<td>0.19</td>
</tr>
<tr>
<td>Koree 2.15</td>
<td>19.0.2.15</td>
<td>Post Data</td>
<td>-1.361 (2.40)</td>
<td>-0.62</td>
<td>-1.02</td>
<td>0.07</td>
</tr>
<tr>
<td>Koree 2.14</td>
<td>18.0.2.14</td>
<td>Post Data</td>
<td>-0.374 (2.59)</td>
<td>-0.09</td>
<td>-0.56</td>
<td>0.23</td>
</tr>
<tr>
<td>Koree 2.16</td>
<td>18.0.2.16</td>
<td>Post Data</td>
<td>-1.684 (2.32)</td>
<td>-0.27</td>
<td>-1.07</td>
<td>0.02</td>
</tr>
<tr>
<td>Liner-Thom</td>
<td>01.0.5.32</td>
<td>Post Data</td>
<td>0.501 (3.77)</td>
<td>0.17</td>
<td>0.31</td>
<td>0.76</td>
</tr>
<tr>
<td>Liner-Thom</td>
<td>01.0.5.32</td>
<td>Post Data</td>
<td>0.517 (3.67)</td>
<td>0.17</td>
<td>0.31</td>
<td>0.76</td>
</tr>
<tr>
<td>Liner-Thom</td>
<td>01.0.5.33</td>
<td>Post Data</td>
<td>0.507 (3.76)</td>
<td>0.16</td>
<td>0.31</td>
<td>0.76</td>
</tr>
<tr>
<td>Marinho</td>
<td>05.2.2.15</td>
<td>Post Data</td>
<td>0.595 (3.27)</td>
<td>0.31</td>
<td>0.31</td>
<td>0.76</td>
</tr>
<tr>
<td>Marinho</td>
<td>05.2.2.15</td>
<td>Post Data</td>
<td>0.598 (3.27)</td>
<td>0.31</td>
<td>0.31</td>
<td>0.76</td>
</tr>
<tr>
<td>Marinho</td>
<td>05.2.2.16</td>
<td>Post Data</td>
<td>0.598 (3.27)</td>
<td>0.31</td>
<td>0.31</td>
<td>0.76</td>
</tr>
<tr>
<td>Marinho</td>
<td>05.2.2.19</td>
<td>Post Data</td>
<td>1.000 (0.14)</td>
<td>0.45</td>
<td>1.14</td>
<td>0.05</td>
</tr>
<tr>
<td>Ngunou</td>
<td>00.0.2.10</td>
<td>Post Data</td>
<td>0.273 (3.07)</td>
<td>-0.49</td>
<td>-0.49</td>
<td>0.62</td>
</tr>
<tr>
<td>O'Corcor</td>
<td>12.2.2.16</td>
<td>Post Data</td>
<td>0.271 (3.07)</td>
<td>0.45</td>
<td>0.45</td>
<td>0.64</td>
</tr>
<tr>
<td>O'Conor</td>
<td>12.2.1.19</td>
<td>Post Data</td>
<td>1.043 (3.07)</td>
<td>0.46</td>
<td>0.46</td>
<td>0.64</td>
</tr>
<tr>
<td>Pineford</td>
<td>11.1.11.10</td>
<td>Post Data</td>
<td>0.562 (3.00)</td>
<td>0.33</td>
<td>0.33</td>
<td>0.74</td>
</tr>
<tr>
<td>Pineford</td>
<td>11.1.1.10</td>
<td>Post Data</td>
<td>1.137 (2.37)</td>
<td>0.31</td>
<td>0.31</td>
<td>0.74</td>
</tr>
<tr>
<td>Pineford</td>
<td>11.1.2.10</td>
<td>Post Data</td>
<td>0.146 (3.00)</td>
<td>0.18</td>
<td>0.18</td>
<td>0.74</td>
</tr>
<tr>
<td>Pineford</td>
<td>11.1.2.15</td>
<td>Post Data</td>
<td>0.835 (3.10)</td>
<td>0.18</td>
<td>0.18</td>
<td>0.74</td>
</tr>
<tr>
<td>Pineford</td>
<td>11.1.2.15</td>
<td>Post Data</td>
<td>0.524 (3.35)</td>
<td>0.18</td>
<td>0.18</td>
<td>0.74</td>
</tr>
<tr>
<td>Pineford</td>
<td>11.1.2.15</td>
<td>Post Data</td>
<td>0.582 (3.24)</td>
<td>0.18</td>
<td>0.18</td>
<td>0.74</td>
</tr>
<tr>
<td>Sapurnoa</td>
<td>19.2.19.13</td>
<td>Post Data</td>
<td>0.952 (3.00)</td>
<td>0.47</td>
<td>0.47</td>
<td>0.66</td>
</tr>
<tr>
<td>Kamis</td>
<td>07.2.7.01</td>
<td>Post Data</td>
<td>2.064 (3.52)</td>
<td>1.75</td>
<td>1.75</td>
<td>0.05</td>
</tr>
<tr>
<td>Kamis</td>
<td>07.2.7.02</td>
<td>Post Data</td>
<td>2.064 (3.48)</td>
<td>2.62</td>
<td>2.62</td>
<td>0.05</td>
</tr>
<tr>
<td>Kamis</td>
<td>07.2.7.03</td>
<td>Post Data</td>
<td>1.163 (3.68)</td>
<td>0.64</td>
<td>0.64</td>
<td>0.54</td>
</tr>
<tr>
<td>Kamis</td>
<td>07.2.2.15</td>
<td>Post Data</td>
<td>2.064 (3.57)</td>
<td>1.75</td>
<td>1.75</td>
<td>0.05</td>
</tr>
<tr>
<td>Kamis</td>
<td>07.2.2.15</td>
<td>Post Data</td>
<td>2.064 (3.45)</td>
<td>2.62</td>
<td>2.62</td>
<td>0.05</td>
</tr>
<tr>
<td>Kamis</td>
<td>07.2.7.11</td>
<td>Post Data</td>
<td>1.163 (3.68)</td>
<td>0.64</td>
<td>0.64</td>
<td>0.54</td>
</tr>
<tr>
<td>Kamis</td>
<td>07.3.2.15</td>
<td>Post Data</td>
<td>1.116 (3.57)</td>
<td>0.64</td>
<td>0.64</td>
<td>0.54</td>
</tr>
<tr>
<td>Kamis</td>
<td>07.3.2.15</td>
<td>Post Data</td>
<td>0.706 (3.57)</td>
<td>0.64</td>
<td>0.64</td>
<td>0.54</td>
</tr>
</tbody>
</table>

Effect Sizes, Weight, Standard Error, & p values
### APPENDIX H

#### Forest Plots of Effect Sizes for Students in Treatment and Control Groups Tier II Studies

<table>
<thead>
<tr>
<th>Study name</th>
<th>Subgroup within study</th>
<th>Comparison</th>
<th>Std diff in means and 95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eversole 4th TX T II</td>
<td>06.5.10.1</td>
<td>Pre-Post TX</td>
<td>1.289</td>
</tr>
<tr>
<td>Eversole 5th TX T II</td>
<td>06.6.10.1</td>
<td>Pre-Post TX</td>
<td>0.405</td>
</tr>
<tr>
<td>Graves 7.03 TX T II</td>
<td>08.0.7.03</td>
<td>Pre-Post TX</td>
<td>0.313</td>
</tr>
<tr>
<td>Govai et al. 2.14 TX T II</td>
<td>13.0.2.14</td>
<td>Pre-Post TX (ES)</td>
<td>1.900</td>
</tr>
<tr>
<td>Govai et al. 2.15 TX T II</td>
<td>13.0.2.15</td>
<td>Pre-Post TX (ES)</td>
<td>1.400</td>
</tr>
<tr>
<td>Healy et al. 8.01 TX T II</td>
<td>09.0.8.01</td>
<td>Pre-Post TX</td>
<td>3.329</td>
</tr>
<tr>
<td>Healy et al. 8.02 TX T II</td>
<td>09.0.8.02</td>
<td>Pre-Post TX</td>
<td>2.771</td>
</tr>
<tr>
<td>Kourea 2.15 TX T I &amp; II</td>
<td>10.0.2.15</td>
<td>Pre-Post TX</td>
<td>1.315</td>
</tr>
<tr>
<td>Kourea 2.14 TX T I &amp; II</td>
<td>10.0.2.14</td>
<td>Pre-Post TX</td>
<td>1.269</td>
</tr>
<tr>
<td>Kourea 2.16 TX T I &amp; II</td>
<td>10.0.2.16</td>
<td>Pre-Post TX</td>
<td>1.307</td>
</tr>
<tr>
<td>Linan-Thompson et al. 5.20 TX T II</td>
<td>01.0.5.20</td>
<td>Pre-Post TX</td>
<td>0.481</td>
</tr>
<tr>
<td>Linan-Thompson et al. 5.32 TX T II</td>
<td>01.0.5.32</td>
<td>Pre-Post TX</td>
<td>1.240</td>
</tr>
<tr>
<td>Linan-Thompson et al. 5.33 TX T II</td>
<td>01.0.5.33</td>
<td>Pre-Post TX</td>
<td>0.992</td>
</tr>
<tr>
<td>McIntosh 2.16 Y1 TX T II</td>
<td>05.2.2.16</td>
<td>Pre-Post TX</td>
<td>1.367</td>
</tr>
<tr>
<td>McIntosh 2.16 Y2 TX T II</td>
<td>05.2.2.16</td>
<td>Pre-Post TX</td>
<td>1.478</td>
</tr>
<tr>
<td>Miller 13.10 T X T II</td>
<td>14.0.13.10</td>
<td>Pre-Post TX</td>
<td>0.562</td>
</tr>
<tr>
<td>Miller 14.10 T X T II</td>
<td>14.0.14.10</td>
<td>Pre-Post TX</td>
<td>1.388</td>
</tr>
<tr>
<td>Nguyen-Quang TX T II</td>
<td>02.0.2.10</td>
<td>Pre-Post TX</td>
<td>0.710</td>
</tr>
<tr>
<td>O'Connor 2.16 TX T II</td>
<td>12.3.2.16</td>
<td>Pre-Post TX</td>
<td>1.331</td>
</tr>
<tr>
<td>O'Connor 7.10 TX T II</td>
<td>12.3.7.10</td>
<td>Pre-Post TX</td>
<td>-0.384</td>
</tr>
<tr>
<td>Pieretti 1.11 LEG TX T II</td>
<td>16.0.1.11</td>
<td>St. Sc Change TX</td>
<td>1.642</td>
</tr>
<tr>
<td>Pieretti 1.12 ONLEG TX T II</td>
<td>16.0.1.12</td>
<td>St. Sc Change TX</td>
<td>1.522</td>
</tr>
<tr>
<td>Ransford-Kaldon 12.10 K TX T II</td>
<td>11.1.2.10</td>
<td>Pre-Post TX</td>
<td>1.641</td>
</tr>
<tr>
<td>Ransford-Kaldon 12.10 1st TX T II</td>
<td>11.2.2.10</td>
<td>Pre-Post TX</td>
<td>1.457</td>
</tr>
<tr>
<td>Ransford-Kaldon 12.10 2nd TX T II</td>
<td>11.3.2.10</td>
<td>Pre-Post TX</td>
<td>1.916</td>
</tr>
<tr>
<td>Ransford-Kaldon 2.15 K TX T II</td>
<td>11.1.2.15</td>
<td>Pre-Post TX</td>
<td>1.119</td>
</tr>
<tr>
<td>Ransford-Kaldon 2.15 1st TX T II</td>
<td>11.2.2.15</td>
<td>Pre-Post TX</td>
<td>1.093</td>
</tr>
<tr>
<td>Ransford-Kaldon 2.16 2nd TX T II</td>
<td>11.3.2.16</td>
<td>Pre-Post TX</td>
<td>-1.997</td>
</tr>
<tr>
<td>Ransford-Kaldon 2.11 1st TX T II</td>
<td>11.2.2.11</td>
<td>Pre-Post TX</td>
<td>0.995</td>
</tr>
<tr>
<td>Richards-Tutor et al. 2.14 TX T II</td>
<td>18.0.2.14</td>
<td>Pre-Post TX</td>
<td>1.596</td>
</tr>
<tr>
<td>Richards-Tutor et al. 2.13 TX T II</td>
<td>18.0.2.13</td>
<td>Pre-Post TX</td>
<td>1.878</td>
</tr>
</tbody>
</table>

#### Forest Plot of Effect Sizes for Students in Treatment Groups: Tier II Pre-Post Studies
Forest Plot of Effect Sizes for Students in Control Groups: Tier II Pre-Post Studies
### APPENDIX I

**Forest Plots of Effect Sizes for Students in Treatment and Control Groups Oral Reading Fluency**

<table>
<thead>
<tr>
<th>Study name</th>
<th>Group</th>
<th>Comparison</th>
<th>Std diff in means</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linan-Thompson et al. (TX T2)</td>
<td>ORF DIBELS Treatment Tier 2</td>
<td>1.679</td>
<td></td>
</tr>
<tr>
<td>McIntosh et al. (Year 1 TX T2)</td>
<td>ORF DIBELS Treatment Tier 2</td>
<td>1.367</td>
<td></td>
</tr>
<tr>
<td>McIntosh et al. (Year 2 TX T2)</td>
<td>ORF DIBELS Treatment Tier 2</td>
<td>1.478</td>
<td></td>
</tr>
<tr>
<td>Kourea (TX T 1 &amp; 2)</td>
<td>ORF DIBELS Treatment T1 &amp; 2</td>
<td>1.307</td>
<td></td>
</tr>
<tr>
<td>Ransford-Kaldon et al. (TX T2)</td>
<td>ORF DIBELS Treatment Tier 2</td>
<td>1.897</td>
<td></td>
</tr>
<tr>
<td>O'Connor et al. (TX T2)</td>
<td>ORF DIBELS Treatment Tier 2</td>
<td>1.331</td>
<td></td>
</tr>
<tr>
<td>McIntosh et al. (TX2 Y1 Tier 2)</td>
<td>ORF DIBELS Treatment 2 Tier 2</td>
<td>1.419</td>
<td></td>
</tr>
<tr>
<td>McIntosh et al. (TX2Y2 Tier 2)</td>
<td>ORF DIBELS Treatment 2 Tier 2</td>
<td>2.648</td>
<td></td>
</tr>
<tr>
<td>Fixed</td>
<td></td>
<td>1.544</td>
<td></td>
</tr>
<tr>
<td>Random</td>
<td></td>
<td>1.544</td>
<td></td>
</tr>
</tbody>
</table>

**Forest Plot of Effect Sizes for Students in Treatment Groups Oral Reading Fluency**

**Forest Plot of Effect Sizes for Students in Control Groups Oral Reading Fluency**

148
APPENDIX J

Forest Plots of Effect Sizes for Students in Treatment and Control Groups Non-Sense Word Fluency

<table>
<thead>
<tr>
<th>Model</th>
<th>Study name</th>
<th>Subgroup within study</th>
<th>Comparison</th>
<th>Std diff in means and 95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Healy, K. TX</td>
<td>AIMSWeb NWF</td>
<td>Treatment</td>
<td>Std diff in means</td>
</tr>
<tr>
<td></td>
<td>Koura, L. TX</td>
<td>DIBELS NWF</td>
<td>Treatment</td>
<td>2.771</td>
</tr>
<tr>
<td></td>
<td>Ransford-Kaldon et al. Kdg</td>
<td>DIBELS NWF</td>
<td>Treatment</td>
<td>1.315</td>
</tr>
<tr>
<td></td>
<td>Ransford-Kaldon et al. 1st</td>
<td>DIBELS NWF</td>
<td>Treatment</td>
<td>1.159</td>
</tr>
<tr>
<td></td>
<td>Gyovai, L. et al. TX</td>
<td>DIBELS NWF</td>
<td>Treatment</td>
<td>1.400</td>
</tr>
<tr>
<td></td>
<td>Richards-Tutor et al. TX1</td>
<td>DIBELS NWF</td>
<td>Treatment</td>
<td>1.091</td>
</tr>
<tr>
<td></td>
<td>Richards-Tutor et al. TX2</td>
<td>DIBELS NWF</td>
<td>Treatment</td>
<td>1.871</td>
</tr>
<tr>
<td></td>
<td>Fixed</td>
<td></td>
<td></td>
<td>Fixed</td>
</tr>
<tr>
<td></td>
<td>Random</td>
<td></td>
<td></td>
<td>Random</td>
</tr>
<tr>
<td></td>
<td>1.539</td>
<td></td>
<td></td>
<td>1.662</td>
</tr>
</tbody>
</table>

Forest Plot of Effect Sizes for ELLs in Treatment Groups Nonsense Word Fluency

<table>
<thead>
<tr>
<th>Model</th>
<th>Study name</th>
<th>Subgroup within study</th>
<th>Comparison</th>
<th>Std diff in means and 95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Koura CG</td>
<td>DIBELS NWF</td>
<td>Control</td>
<td>Std diff in means</td>
</tr>
<tr>
<td></td>
<td>Ransford-Kaldon et al. Kdg</td>
<td>DIBELS NWF</td>
<td>Control</td>
<td>0.985</td>
</tr>
<tr>
<td></td>
<td>Ransford-Kaldon et al. 1st</td>
<td>DIBELS NWF</td>
<td>Control</td>
<td>1.127</td>
</tr>
<tr>
<td></td>
<td>Fixed</td>
<td></td>
<td></td>
<td>Fixed</td>
</tr>
<tr>
<td></td>
<td>Random</td>
<td></td>
<td></td>
<td>Random</td>
</tr>
<tr>
<td></td>
<td>1.328</td>
<td></td>
<td></td>
<td>1.328</td>
</tr>
</tbody>
</table>

Effect Sizes for ELLs in Control Groups Nonsense Word Fluency
## APPENDIX K

### Forest Plots of Effect Sizes for Students in Treatment and Control Groups Word Attack

<table>
<thead>
<tr>
<th>Model</th>
<th>Study name</th>
<th>Std diff in means</th>
<th>Std diff in means and 95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Linan-Thompson et al. TX T2</td>
<td>1.240</td>
<td>-3.00 -1.50 0.00 1.50 3.00</td>
</tr>
<tr>
<td></td>
<td>McMaster et al. TX T1</td>
<td>1.185</td>
<td>Favours A</td>
</tr>
<tr>
<td></td>
<td>Pieretti TX LEG</td>
<td>0.742</td>
<td>Favours B</td>
</tr>
<tr>
<td></td>
<td>Pieretti TXONLEG</td>
<td>1.000</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Pieretti TX CRONLEG</td>
<td>0.928</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Fixed</td>
<td>1.108</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Random</td>
<td>1.108</td>
<td></td>
</tr>
</tbody>
</table>

### Effect Sizes for ELL in TX groups Word Attack

<table>
<thead>
<tr>
<th>Model</th>
<th>Study name</th>
<th>Std diff in means</th>
<th>Std diff in means and 95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Linan-Thompson et a. CG T2</td>
<td>1.087</td>
<td>-3.00 -1.50 0.00 1.50 3.00</td>
</tr>
<tr>
<td></td>
<td>McMaster et al. CG T1</td>
<td>0.738</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Pieretti. CG</td>
<td>1.069</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Fixed</td>
<td>0.980</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Random</td>
<td>0.980</td>
<td></td>
</tr>
</tbody>
</table>

### Effect Sizes for ELL in CG groups Word Attack
APPENDIX L

Forest Plots of Effect Sizes for Mean Gain and Mean Difference Analyses Reading Comprehension

Effect Sizes for Mean Gain Analysis Students in Treatment Groups Reading Comprehension

Effect sizes for Mean Difference Analysis Post-Test Data Reading Comprehension
APPENDIX M

Publication Bias Funnel Plots Method

Tier I Post-Test Data Mean Difference Analysis - Funnel Plot

Tier I Pre-Post Studies Treatment Mean Gain Analysis - Funnel Plot
Tier I Pre-Post Studies Control Mean Gain Analysis - Funnel Plot

Tier II Post-Data Only Mean Difference Analysis - Funnel Plot
Tier II Pre-Post Studies Treatment Mean Gain Analysis - Funnel Plot

Tier II Pre-Post Studies Control Mean Gain Analysis - Funnel Plot
APPENDIX N

Publication Bias Trim and Fill Method

Tier I Post-Data Mean Difference Analysis – Funnel Plot with Imputed Studies after Trim & Fill Method

Tier I treatment group mean gain analysis-funnel plot with imputed studies after Trim and Fill method
Tier I control group mean gain analysis-funnel plot with imputed studies after Trim and Fill method

Tier II Control group Mean Gain Analysis-Funnel Plot with Imputed Studies after Trim and Fill method
Tier II treatment group mean gain analysis-funnel plot with imputed studies after Trim and Fill