The Effects of Reading Interventions on the Word-Reading Performance of English Language Learners: A Meta-Analysis

by

Caralyn Ludwig

A thesis submitted in partial fulfillment of the requirements for the degree of

Master of Education

in

Special Education

Department of Educational Psychology
University of Alberta

Abstract

Despite a large body of research on the effectiveness of reading interventions for monolingual English speakers, research on reading interventions for English language learners (ELLs) is still relatively scarce. Because the number of ELLs in schools is growing rapidly and these students often have weaker English literacy skills than their English-speaking peers, this meta-analysis examined the effectiveness of reading interventions in improving the word-reading skills of school-aged ELLs. Fourteen experimental studies with reported outcomes for pre-test and post-test were selected, and four moderator variables (group size, intensity of the intervention, student risk status, and type of intervention) were explored to explain differences in the intervention effects. The results of the random effects analysis showed that the reading interventions had a large, positive effect on ELLs' real word (g = 1.07), nonword (g = 1.00), and combined real word and nonword (g = 1.15) reading scores. Results also suggested that some reading interventions were more effective than others. We found differences in effectiveness related to the group size and the length of the intervention in the real word reading analysis, but more research into potential moderators is warranted. Overall, our findings suggest that reading interventions for ELLs produce significant effects, and should not be delayed until these students have reached a certain level of oral English proficiency. Such interventions are likely an important first step in closing the achievement gap between ELLs and their English-speaking peers.

Acknowledgments

I would first like to thank my supervisors, Dr. George Georgiou and Dr. Heather Brown, for their support over the past year. Thank you for seeing the academic potential in me, providing me with advice and encouragement, and challenging me to push myself beyond my perceived limits. I will be forever grateful for this experience. Second, to Amy, for giving me the courage I needed to step out of my comfort zone. Thank you for your unwavering support, your motivating pep talks, and your uncanny ability to always know exactly what to say. To Brandi and Andrea, for being the best group members and coffee companions I could have asked for. I am so grateful for our friendships formed through this experience. Finally, to my family: to my parents, thank you for being my biggest cheerleaders and having endless confidence in me, and to my brother, thank you for your willingness to discuss statistics. I could not have done this without you.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

English language learners (ELLs), defined as students who first learned to speak, read, and/or write a language other than English, have been identified as one of the fastest-growing student populations in North American schools (Alberta Education, 2007; Wagner, Francis, & Morris, 2005). From 2001 to 2011, the percentage of children aged 5 to 14 who were identified as language minority in Canada rose from 9.2% to 14.1% (Statistics Canada, 2001, 2011); similarly, the proportion of ELLs in Alberta elementary schools increased from 13.6% in 2007 to 20.7% in 2014 (Cowley & Easton, 2009, 2016). In Edmonton Public Schools alone, the number of ELLs has more than quintupled in just over a decade, increasing from 4,000 students in 2004 to more than 22,000 in September 2015, and now comprising nearly 25% of the student population (Robertson & Stoddard, 2016; Zabjek, 2015). These large and rapidly growing numbers of ELLs pose complex instructional challenges for educators, particularly in the field of reading.

It has been shown that ELLs often enter school with weaker English literacy skills than their native English-speaking peers and are at risk for falling behind academically (Lesaux, 2006; Li & Edwards, 2010; Vadasy & Sanders, 2011). ELLs are expected not only to learn academic content, but also to learn a new language at the same time (Gersten, 1996). These demands have contributed to a large achievement gap between ELLs and native English-speakers. On the 2015 National Assessment of Educational Progress in the United States, fourth-grade ELLs scored 37 points lower than non-ELLs in reading, where 10 points is roughly equivalent to one grade level (Li & Edwards, 2010; National Center for Education Statistics, 2015). Alarmingly, this gap has remained essentially unchanged since 1998 (Kena et al., 2014), implying that educators are unable to meet the needs of their ELL students.

One contributing factor to this large and unchanging achievement gap may be that some schools do not provide reading interventions to ELLs until these students reach a certain level of oral English proficiency (Dussling, 2016; Limbos & Geva, 2001). This may be due, in part, to a lack of funding required for early assessment and intervention. Limbos and Geva (2001) reported that some educators delay assessing ELLs for learning or reading difficulties for up to four or five years, believing the students' educational challenges are due to the acculturation process and that they simply need time to allow their oral language skills to develop. However, research indicates that not only are reading interventions important for ELLs (Dussling, 2016; Gersten et al., 2007), but that ELLs can benefit from reading interventions as much as native English-speaking students, regardless of their initial level of oral language proficiency (Gunn, Smolkowski, Biglan, Black, & Blair, 2005; Linan-Thompson & Hickman-Davis, 2002).

Research examining the effects of supplemental reading interventions for ELLs is still relatively limited, although there has been increased interest in this area as the ELL population continues to grow rapidly (e.g., Linan-Thompson, Vaughn, Hickman-Davis, & Kouzekanani, 2003; Nelson, Vadasy, & Sanders, 2011; Stuart, 1999, 2004). Many reading intervention studies have included ELLs in their samples, but have not disaggregated data for this sub-population of students (see e.g., Stuart, 1999), making it impossible to determine how well the interventions work for ELLs compared to monolingual English speakers. Students from diverse language backgrounds may differ in how they respond to intervention, so it is important to consider the instructional impact on this group of students separately from the rest of the population. This meta-analysis was therefore conducted to determine how effective English reading interventions are in improving ELLs' reading ability, as no such meta-analysis currently exists. Specifically, we wanted to look at the effects of intervention on ELLs' word-reading skills, because word reading is foundational to understanding text and has been connected to reading comprehension—the ultimate goal of reading—in both monolingual English speakers and ELLs (Dussling, 2016; Gough, 1996; Kendeou, van den Broek, White, & Lynch, 2009; Vaughn et al., 2006a). For the purpose of this study, word reading includes both real word reading (e.g., accurately reading *cat, size, horizon*) and nonword reading (e.g., accurately reading *ree, blork, wroutch*). It does not include word/text reading fluency or reading comprehension.

Before running this meta-analysis, it is helpful to look at the size of effects found in previous meta-analyses that included both monolingual English speakers and ELLs without differentiating between the two populations. The National Reading Panel conducted two meta-analyses examining the effects of alphabetic-based reading interventions for students in Kindergarten to Grade 6 (Ehri, Nunes, Stahl, & Willows, 2001a; Ehri et al., 2001b). The first meta-analysis was conducted by Ehri et al. (2001a) and looked at the effects of phonics interventions, where students were taught letter-sound correspondences in order to help them accurately decode words. Moderate, statistically significant effect sizes were reported for decoding nonwords and regularly spelled words ($d = 0.60$ and $d = 0.67$, respectively); for sight word reading, which included irregularly spelled words, the effect size was smaller ($d = 0.40$). (According to Cohen [1988], effect sizes can be interpreted using the following criteria: 0.2 is small, 0.5

is moderate, and 0.8 is large.) The second meta-analysis was conducted by Ehri et al. (2001b) and examined the effects of phonemic awareness interventions, where children were taught to focus on and manipulate the sounds in spoken words. Again, moderate, statistically significant effect sizes were reported for both real word reading ($d = 0.61$) and nonword reading ($d = 0.56$). Taken together, the findings of these meta-analyses suggest that various reading interventions are fairly effective for improving students' word-reading skills. Given that both of these meta-analyses included ELLs, but did not disaggregate data for this population, it is reasonable to hypothesize that reading interventions may also moderately improve ELLs' word-reading abilities.

Because the effectiveness of each intervention may vary as a result of different factors, we explored the role of several moderator variables that have also been identified in previous meta-analyses with native English speakers (see e.g., Cavanaugh, Kim, Wanzek, & Vaughn, 2004; Goodwin & Ahn, 2010; Suggate, 2010). First, the intervention group size, or the instructor-to-student ratio, was examined. It is generally assumed that smaller group sizes are more effective for improving the reading performance of struggling readers, as students have more opportunities to work directly with the instructor and receive specific feedback (Vaughn et al., 2003). Previous meta-analyses have indicated that there is no difference in effect sizes between students receiving intervention one-on-one with the instructor and students receiving intervention in small groups (Elbaum, Vaughn, Hughes, & Moody, 2000; Suggate, 2010, 2016); however, research has shown that both one-on-one and small-group instruction are more effective than large-group instruction (e.g., more than five students) (Vaughn et al., 2003; Wanzek & Vaughn, 2007).

The intensity of the intervention, which encompassed the length of each intervention session, the number of intervention sessions per week, and the length of the entire intervention, was also thought to be a potential moderator. It may be assumed that longer and more intense interventions result in more learning. However, Ehri et al. (2001b) found that phonemic awareness interventions that lasted 5 to 18 hours had statistically larger effect sizes than interventions that were shorter or longer than this range. In their meta-analysis on morphological awareness interventions, Goodwin and Ahn (2010) also found that interventions lasting 10 to 20 hours had the largest statistically significant effect sizes, followed closely by interventions lasting more than 20 hours. Thus, interventions with higher intensity and more instructional time may not always result in larger effect sizes.

A third moderator that was explored was student risk status. Some studies in this meta-analysis included all ELLs in their samples (with or without reading difficulties), and others only the ELLs who were at risk for reading difficulties. In a meta-analysis examining reading interventions for struggling readers in Grades 4 through 12, significantly larger effects were found for the students with learning disabilities than for the students without learning disabilities (Scammacca, Roberts, Vaughn, & Stuebing, 2015). Similarly, Goodwin and Ahn (2010) found larger effect sizes for at-risk children, including students with learning disabilities, students with reading disabilities, and ELLs, who were considered an at-risk group in this study. Ehri et al. (2001a) also reported that effect sizes were significantly greater for young (i.e., children in Kindergarten and Grade 1) at-risk children, perhaps because these students began the interventions as new readers with much to learn. Student risk status could therefore be an important moderator to consider.

Finally, the type of intervention provided to the students was examined, coded according to the primary reading skill targeted (e.g., phonics, phonological awareness, vocabulary, or other). The meta-analysis by Scammacca et al. (2015) found that effect sizes differed significantly depending on the type of intervention. Comprehension interventions and vocabulary interventions were associated with the largest effects for overall reading outcomes. Suggate (2016) suggested that some interventions may have advantages over others because the skills they explicitly target transfer better to general reading skills such as word reading. For example, according to the phonological linkage hypothesis (Hatcher, Hulme, & Ellis, 1994), phonics interventions may improve children's reading skills more effectively than phonological awareness interventions because phonics instruction directly links phonemes and words. Previous meta-analyses have also suggested that phonics and phonological awareness interventions are particularly helpful for young children (Ehri et al., 2001a, 2001b; Suggate, 2016). Consequently, it is reasonable to expect that the instructional content of the interventions might have an impact on effect sizes.

In summary, while there are some meta-analyses on the effectiveness of reading interventions for general populations of students, to our knowledge, no such studies exist for ELLs. As the ELL population is expanding rapidly, and these students often have much weaker literacy skills than their English-speaking peers, further research is needed to determine how best to improve ELLs' reading abilities in an effort to close the achievement gap. A meta-analysis will allow us to synthesize current research and examine the overall effectiveness of reading interventions in improving ELLs' word-reading skills, while taking into account moderator variables such as group size, intensity of intervention, student risk status, and type of intervention.

The Present Study

The purpose of this meta-analysis was to examine the effectiveness of reading interventions on the word-reading skills of school-aged ELLs. Specifically, we asked the following research questions:

1. What are the size, consistency, and directionality of the effect of reading interventions on ELLs' real word, nonword, and combined word reading skills?
2. Did group size, intensity of the intervention, student risk status, or type of intervention account for some of the variability in the effectiveness of the interventions from pre-test to post-test?

Based on the findings of previous meta-analyses (e.g., Ehri et al., 2001a; Goodwin & Ahn, 2010; Scammacca et al., 2015; Wanzek & Vaughn, 2007), we hypothesized that:

1. Reading interventions with school-aged ELLs will have moderate, positive effects on the students' real word, nonword, and combined word reading skills.
2. The variability in the effectiveness of the interventions may be accounted for by several moderator variables in the following ways:
    a. individual and small-group instruction (i.e., 1 to 5 students per group) will be more effective than large-group instruction;
    b. interventions of moderate intensity (e.g., 10 to 20 hours in length) will be the most effective;

c. ELLs who are classified as at risk for reading difficulties will show greater improvement in word reading than ELLs who are not at risk;
d. phonics interventions will be the most effective type of intervention, since they explicitly teach word-reading skills.

Chapter 2: Method

Search Strategy, Inclusion Criteria and Coding Procedures

Five computerized databases (PsycINFO, ERIC, Google Scholar, PubMed, and Proquest Theses and Dissertations) were searched for relevant articles using several combinations of the following keywords:

- reading intervention (i.e., reading + programs, improvement, instruction, remediation);
- English language learners (i.e., ELL, English as a second language, ESL, bilingualism);
- reading skills (i.e., phonemic awareness, phonological awareness, phonics, reading comprehension, reading fluency, vocabulary, decoding, word recognition).

This initial search yielded 1,073 abstracts for further screening. Studies found through the literature search were included in the meta-analysis if they met all of the following criteria:

- The study was published in English between January 1990 and July 2016.
- The participants were identified as being English language learners (ELLs). Studies with additional, English-speaking students were included when disaggregated data were provided for the students who were ELLs.
- The participants were enrolled in Kindergarten through Grade 12 in a school program in which the primary language of instruction was English. Studies that examined students in bilingual programs were excluded.

- The intervention: (a) targeted a reading-related skill; (b) was administered in English over a period of time (i.e., more than a single session); (c) was in addition to the general education curriculum provided to all students; and (d) occurred at school during regular school hours (not at home or in a clinic).
- At least one of the dependent variables addressed a reading outcome in real word reading accuracy, nonword reading accuracy, or an average of both.
- The research design was experimental or quasi-experimental.
- Pre-test and post-test means, standard deviations, and sample sizes were provided to calculate effect sizes.

In order to increase the number of data points, we also contacted the first authors of the selected articles to ask if they had any studies that could be included in the meta-analysis that we may have missed. We also asked several researchers if they had additional data that was required in order to include their studies in the analyses (for example, we required disaggregated ELL data for Lovett et al., 2008 and Wanzek and Roberts, 2012).

A total of 14 studies were included in the meta-analysis. When multiple independent samples were included in one study, they were treated as separate studies (see Solari & Gerber, 2008; Vadasy & Sanders, 2015; Vadasy, Sanders, & Nelson, 2015; Wanzek & Roberts, 2012). This resulted in 21 different samples comprising a total of 869 individuals who were ELLs.

The coding system was comprised of eight categories: (a) sample size; (b) pre-test and post-test means and standard deviations for the real word, nonword, or combined word reading scores of participants; (c) primary reading skill targeted in the intervention, coded as one of four categories: phonological awareness, phonics, vocabulary, or other; (d) number of times the intervention was delivered per week; (e) length of each intervention session, in minutes; (f) number of total minutes of intervention; (g) size of each group of students receiving the intervention (category 1: $n = 1$; category 2: $n = 2$ to 5; category 3: $n > 5$); and (h) student risk status, which indicated whether or not the students were identified as being at risk for reading difficulties. All studies were double-coded by the first author and a graduate student who received training on how to code the studies, and 100% interrater agreement was achieved.

Description of Studies Included in the Meta-Analysis

Most of the studies included in this meta-analysis examined the efficacy of a particular reading intervention in improving various reading skills of ELLs. Four studies (Solari & Gerber, 2008; Vadasy et al., 2015; Vadasy & Sanders, 2015; Wanzek & Roberts, 2012) compared the effects of more than one treatment condition. For example, Wanzek and Roberts (2012) had three treatment groups: the first group received a phonics intervention, the second group received a reading comprehension intervention, and the third group received either the phonics or the reading comprehension intervention, depending on students' area of deficit in pre-test measures. Several studies investigated whether the effectiveness of the intervention differed for ELLs and native English speakers (Dussling, 2016; Lovett et al., 2008; Vadasy & Sanders, 2010, 2011; Wanzek & Roberts, 2012). Participants came from Kindergarten to Grade 8, with 11 out of 14 studies examining ELLs in Kindergarten or Grade 1 (see Appendix A for participant characteristics).

The interventions provided to students varied in their content, but could be organized into four general categories according to the primary reading skill targeted in the intervention. The first group of studies, including Dussling (2016), had an explicit focus on phonics, teaching letter names and sounds and giving students an opportunity to read decodable books aloud. Phonological awareness skills (e.g., orally blending sounds in words) were also targeted in several studies, but were usually one smaller part of the whole intervention, rather than the primary focus. Nelson et al. (2011) was one of several studies that focussed on vocabulary. A different target word was taught every day through various activities such as discussing the meaning of the word, spelling the word, identifying a picture of the word, and using the word in an oral language activity. The final group of studies provided multifaceted interventions. For example, Vaughn et al. (2006a) provided instruction designed to build phonemic awareness, letter knowledge, word recognition, fluency, and reading comprehension skills. The characteristics of each intervention are provided in Appendix B.

Statistical Analysis

Standardized mean differences (SMDs) are a simple way of quantifying the difference between pre-test and post-test scores by creating a standardized numerical result that represents the effectiveness of a particular intervention (Coe, 2002). They are useful when the size of the intervention effect is not directly comparable across studies due to a difference in measurement scale (Borenstein, Hedges, Higgins, & Rothstein, 2009). Because the studies included in this meta-analysis used different tests to measure word-reading outcomes, the use of SMDs was helpful. The current meta-analysis employed a random effects model, which assumes that the SMDs vary from study to study (Borenstein et al., 2009). The metafor package for the R statistical program (Viechtbauer, 2010) was used for the analyses. We computed SMDs for each sample to compare the ELLs' pre-test and post-test scores on real word reading, nonword reading, and/or averaged real word and nonword ("combined" word) reading scores. For each of the three analyses, we then computed the overall summary effect, or Grand SMD, which is conceptually similar to the mean of all the SMDs in each analysis (Borenstein et al., 2009). Looking across all of the reading intervention studies with ELLs, we wanted to know how much these children improved their overall word-reading skills. In other words, how effective were the interventions in general? Stouffer's method, commonly referred to as the weighted Z-test, was then used to (a) compute the average effect size (i.e., the Grand SMD), and (b) determine whether the Grand SMD for each analysis was significantly different from zero (Zaykin, 2011).

It was assumed that the amount of growth that students showed in their word-reading abilities from pre-test to post-test, or the effectiveness of each study, would be different for each sample. These observed effect sizes vary from study to study due to a combination of chance differences between studies and real differences, or heterogeneity (Borenstein et al., 2009). By estimating how much of the total variation is expected to be from random error, we can conclude that any remaining variation is heterogeneity and reflects true differences in effect sizes. For this meta-analysis, the heterogeneity of the SMDs for real word, nonword, and combined word reading was assessed with a Q-test, where significance suggests that the amount of total variance is more than we would expect based on random error (Borenstein et al., 2009). The Q-test also reports a descriptive statistic called I^2. This statistic quantifies how much of the variance in SMDs is real and expresses this value as a percentage (Borenstein et al., 2009). In addition, Tau^2 (T^2, a standardized version of I^2) was used to quantify a population interval (PI) determined by the following formula: PI = $\theta \pm 1.96(\sqrt{T^2})$. The PI represents the range of effectiveness we might expect a reading intervention to have for any group of ELLs randomly chosen from the population.

Publication Bias

It is generally admitted that studies reporting statistically significant results are published more often than those that find null results (Borenstein et al., 2009). In addition, large studies are more likely to be published regardless of effect size, whereas for studies with small sample sizes, only the largest effects are likely to be published. This leads to a bias in the published literature, where relatively large effects for a given research question are reported more often than small effects for the same question. This bias may then be reflected in a meta-analysis that draws on this literature, resulting in an overestimation of the true summary effect.

In an attempt to offset publication bias by locating data in the “grey” literature, a database of unpublished theses and dissertations was searched for pertinent studies, and one study that met the inclusion criteria was retrieved. In addition, we contacted the first authors of the studies selected for the meta-analysis to ask if they had any relevant unpublished studies that should be included. The impact of potential publication bias on the summary effect was then evaluated using several analyses. First, Rosenthal’s Fail-Safe N was computed to determine the number of missing studies that would need to be incorporated into the meta-analysis in order to nullify the summary effect and make it essentially zero (Rosenthal, 1995). Furthermore, funnel plots were created to assess whether the studies were distributed asymmetrically about the mean effect size, which may indicate the presence of publication bias (Borenstein et al., 2009). The symmetry of the funnel plots was examined using Duval and Tweedie's Trim and Fill procedure, which estimates the impact of an asymmetrical distribution on the summary effect, and seeks to correct any asymmetries that are found. For this procedure, effect sizes for unpublished studies that may have been omitted are added into the funnel plot so that the distribution is symmetrical, and the effect size is re-computed so that it is no longer biased (Borenstein et al., 2009). We also assessed publication bias using the Rank Correlation Test and Egger's Regression Test. These tests examine the relationship between the size of the effects from each study and the associated standard error. Like the funnel plot, if a relationship exists, it implies the presence of publication bias (Brown, Smyth, & Ansari, in press; Viechtbauer, 2014). To assess the impact of the most extreme SMDs on the summary effect and confirm that our findings were not driven by any single study, the leave-one-out method was used. For each of the three meta-analytic models (real word, nonword, and combined word reading), samples were removed one at a time from the analysis and the Grand SMD was re-calculated without these samples. (See Results section for data from these analyses.)

Chapter 3: Results

Impact on Word-Reading Skills

The random effects model demonstrated that the Grand SMDs from pre-test to post-test were both large (Cohen, 1992) and positive for real word reading ($g = 1.07$, $z = 6.16$, $p < .0001$), nonword reading ($g = 1.00$, $z = 5.63$, $p < .0001$), and combined real word and nonword reading ($g = 1.15$, $z = 8.56$, $p < .0001$). Population interval analyses showed that for any random sample of ELLs, we would expect that the effect of a given intervention would range from 0.2 to 1.9 for real word reading, -0.1 to 2.1 for nonword reading, and 0.3 to 2.0 for combined word reading (see Tables 1-2 and Figures 1-3).

Table 1

Meta-Analytic Results: Overall Standardized Mean Differences for Real Word, Nonword, and Combined Analyses

Outcome Measures	Grade	*k*	*n*	*SMD*	*SE*	*Z*	*p*
Real Word	K-8	12	219	1.07	0.174	6.161	< .0001
	K-1	8	121	1.15	0.246	4.688	< .0001
Nonword	K-8	15	568	1.00	0.177	5.633	< .0001
	K-1	10	444	1.15	0.246	4.688	< .0001
Combined	K-8	16	477	1.15	0.134	8.559	< .0001
	K-1	12	379	1.24	0.158	7.860	< .0001

Note. K-8 = students in Kindergarten to Grade 8; K-1 = students in Kindergarten to Grade 1.

Table 2

Meta-Analytic Results: Population Intervals and Heterogeneity Analyses for Real Word, Nonword, and Combined Analyses

Outcome Measures	Grade	$PI_{lower}{}^{a}$	$PI_{upper}{}^{a}$	I^2 (%)	Q	p
Real Word	K-8	0.23	1.91	56.68	25.801	.0069
Nonword	K-8	-0.13	2.13	82.21	57.780	< .0001
Combined	K-8	0.35	1.95	66.42	44.053	.0001

[a]These population intervals are based on 95% confidence intervals.

The heterogeneity analysis demonstrated that for real word reading, 57% of the variability in effect sizes from sample to sample was systematic, $Q(11) = 25.801$, $I^2 = 56.68$, $p = .0069$. 83% and 66% of the variability was also real for nonword reading, $Q(14) = 57.780$, $I^2 = 82.21$, $p < .0001$, and combined measures of word reading, $Q(15) = 44.053$, $I^2 = 66.42$, $p = .0001$), respectively. Because the results of the heterogeneity analyses for all three word-reading outcomes were significant, there was more variability in SMDs from sample to sample than would be expected from random error, and moderator variables could be explored as potential sources of additional variance in the SMDs.

Results of Moderator Analyses

Moderator analyses were conducted on the real word, nonword, and combined word reading samples. In the nonword and combined word reading samples, none of the moderator variables (group size, intensity of the intervention, student risk status, or type of intervention) significantly explained the variability in the effect sizes between studies.

In contrast, for real word reading, group size was significant (β = -0.830, Z = -3.908, p < .0001), where intervention groups with more than five students seemed to be less effective than groups of two to five students. However, it is noteworthy that there were only two categories (n = 2 to 5 or n > 5) for this analysis, as there were no groups where the interventionist worked one-on-one with students. The number of minutes in each intervention session was also significant for the real word reading samples (β = -0.018, Z = -2.998, p < .0027), indicating that longer intervention sessions were less effective than shorter sessions. See Tables 3-5 for Q statistics and p values for each moderator variable analysis.

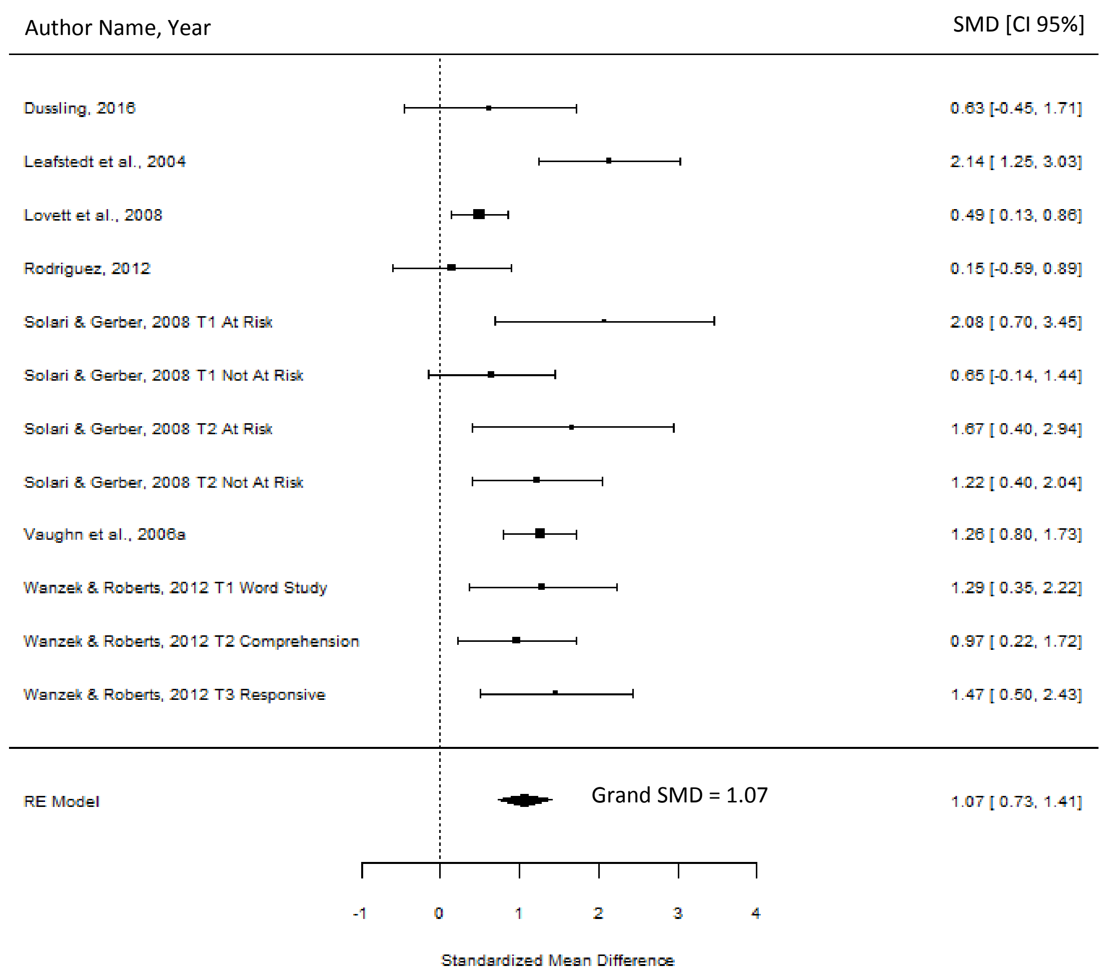

Figure 1. Forest plot: Strength of the standardized mean difference between pre-test and post-test real word reading scores. *Source:* Viechtbauer (2010)

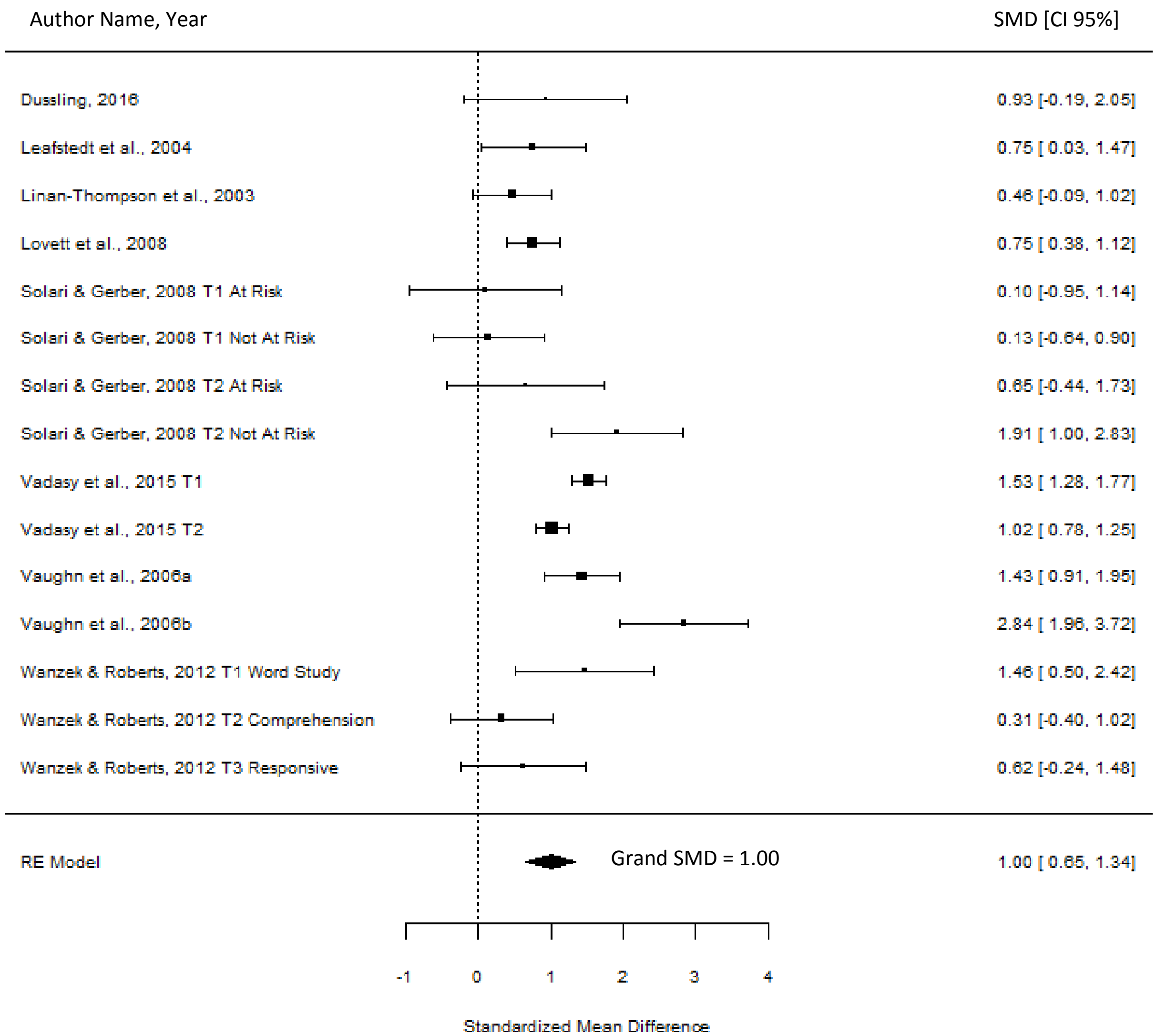

Figure 2. Forest plot: Strength of the standardized mean difference between pre-test and post-test nonword reading scores. *Source:* Viechtbauer (2010)

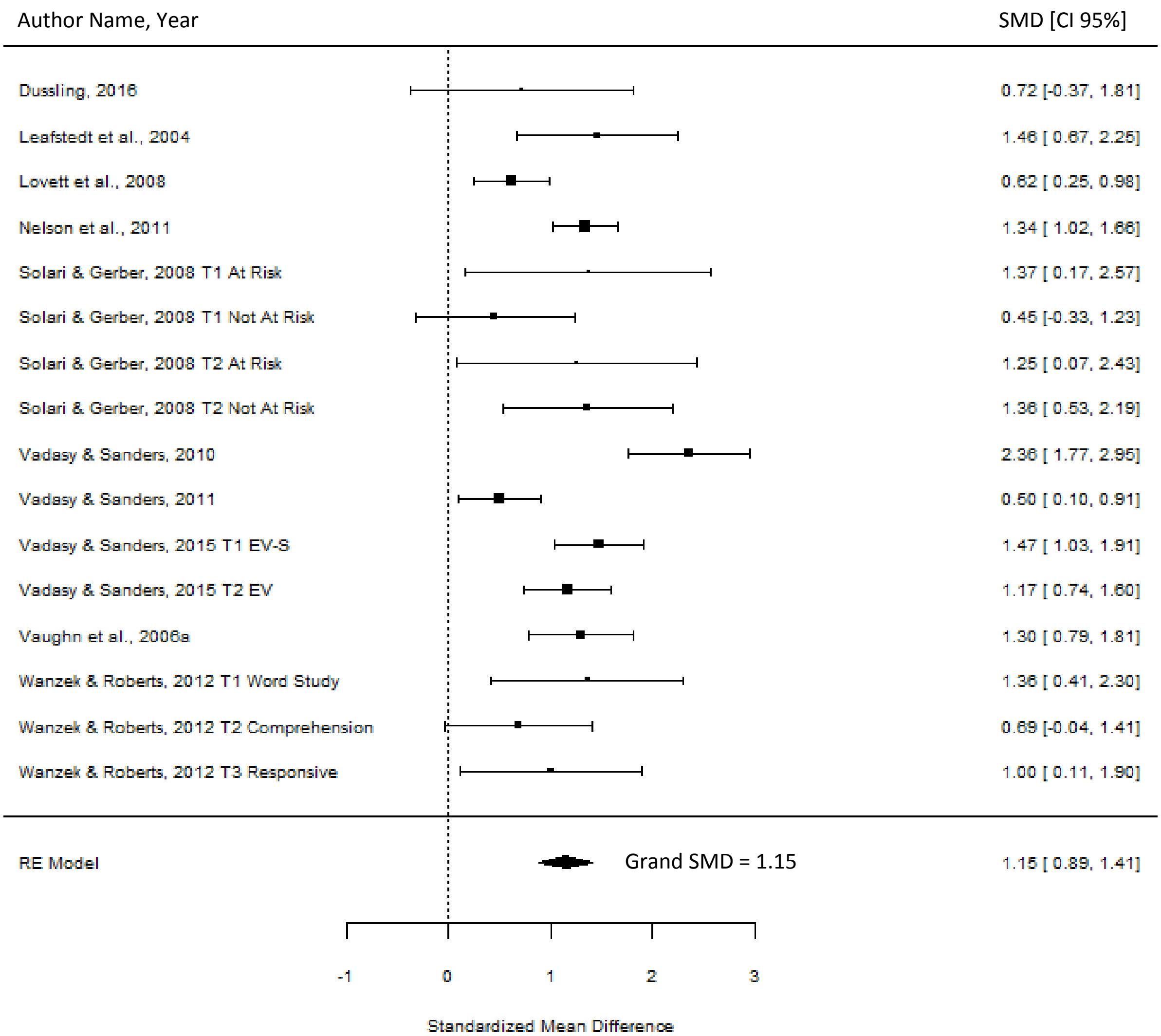

Figure 3. Forest plot: Strength of the standardized mean difference between pre-test and post-test combined word reading scores. *Source:* Viechtbauer (2010)

Table 3

Moderator Analyses for Real Word Reading

Moderator	Descriptives	Q_M	*p*
Number of sessions per week	*M* = 3.63 Range = 2 to 5	1.195	.2743
Minutes per session	*M* = 34.58 Range = 15 to 90	8.989	**.0027**
Total number of minutes of intervention	*M* = 2284.17 Range = 300 to 6300	1.073	.3003
Group size	2 to 14 students	15.275	**< .0001**
Student risk status	Number of typical samples = 3 Number of at risk samples = 9	2.073	.1499

Note. k = 12 for all moderator analyses.

Table 4

Moderator Analyses for Nonword Reading

Moderator	Descriptives	Q_M	*p*
Number of sessions per week	*M* = 3.9 Range = 2 to 5	1.470	.2253
Minutes per session	*M* = 31 Range = 15 to 60	1.906	.1674
Total number of minutes of intervention	*M* = 2600 Range = 300 to 7000	3.321	.0684
Group size	1 to 8 students	0.161	.6881
Student risk status	Number of typical samples = 4 Number of at risk samples = 11	0.294	.5876

Note. k = 15 for all moderator analyses.

Table 5

Moderator Analyses for Combined Measures of Word Reading

Moderator	Descriptives	Q_M	*p*
Number of sessions per week	*M* = 3.84 Range = 2 to 5	0.066	.7971
Minutes per session	*M* = 27.19 Range = 15 to 60	1.098	.2948
Total number of minutes of intervention	*M* = 2093.13 Range = 300 to 6300	0.720	.3960
Group size	1 to 8 students	1.959	.3755
Student risk status	Number of typical samples = 6 Number of at risk samples = 10	3.304	.0691
Primary reading skill targeted in the intervention	PA: 3 samples Phonics: 5 samples Vocabulary: 3 samples Other: 5 samples	0.536	.9109

Note. *k* = 16 for all moderator analyses; PA = phonological awareness.

Publication Bias

All of the random effects models were assessed for publication bias using the Fail-Safe N method, Egger's Method, the Rank Correlation Test, and the Trim and Fill procedure. There was little evidence of publication bias across the three meta-analytic models (real word, nonword, and combined word reading) included in this analysis. The Trim and Fill and Rank Correlation analyses showed some evidence of publication bias in the model for real word reading. Therefore, the true Grand SMD may be somewhat lower for real word reading (corrected Grand SMD = 0.88) than reported in the original analysis (see Table 6 and Figure 4; note that the x-axis in Figure 5 [Funnel plot for nonword reading analysis] is slightly different than the x-axes in Figure 4 [Funnel plot for real word reading analysis] and Figure 6 [Funnel plot for combined word reading analysis], and the funnel plots are not intended to be directly compared to each other). However, significant evidence of publication bias for the real word data was inconsistent across the four methods, and therefore should not be interpreted as conclusive. As a final check, we examined the reliability of our results using the leave-one-out method. After removing one study at a time from the analyses, the Grand SMD ranged from 0.96 to 1.15 for real word reading, 0.90 to 1.06 for nonword reading, and 1.05 to 1.21 for combined word reading. In sum, there was very little evidence of publication bias in these analyses.

Table 6

Publication Bias Analyses

Outcome Measures	Fail-Safe N	Egger's Method		Rank Correlation Test		Trim and Fill Procedure	
		z	*p*	Kendall's tau	*p*	Imputed	Corrected SMD
Real Word	379	1.758	.079	0.455	**.045**	**3***	0.88
Nonword	1014	-0.193	.847	0.029	.923	0	-
Combined	1276	0.142	.887	-0.033	.894	0	-

*Evidence of publication bias.

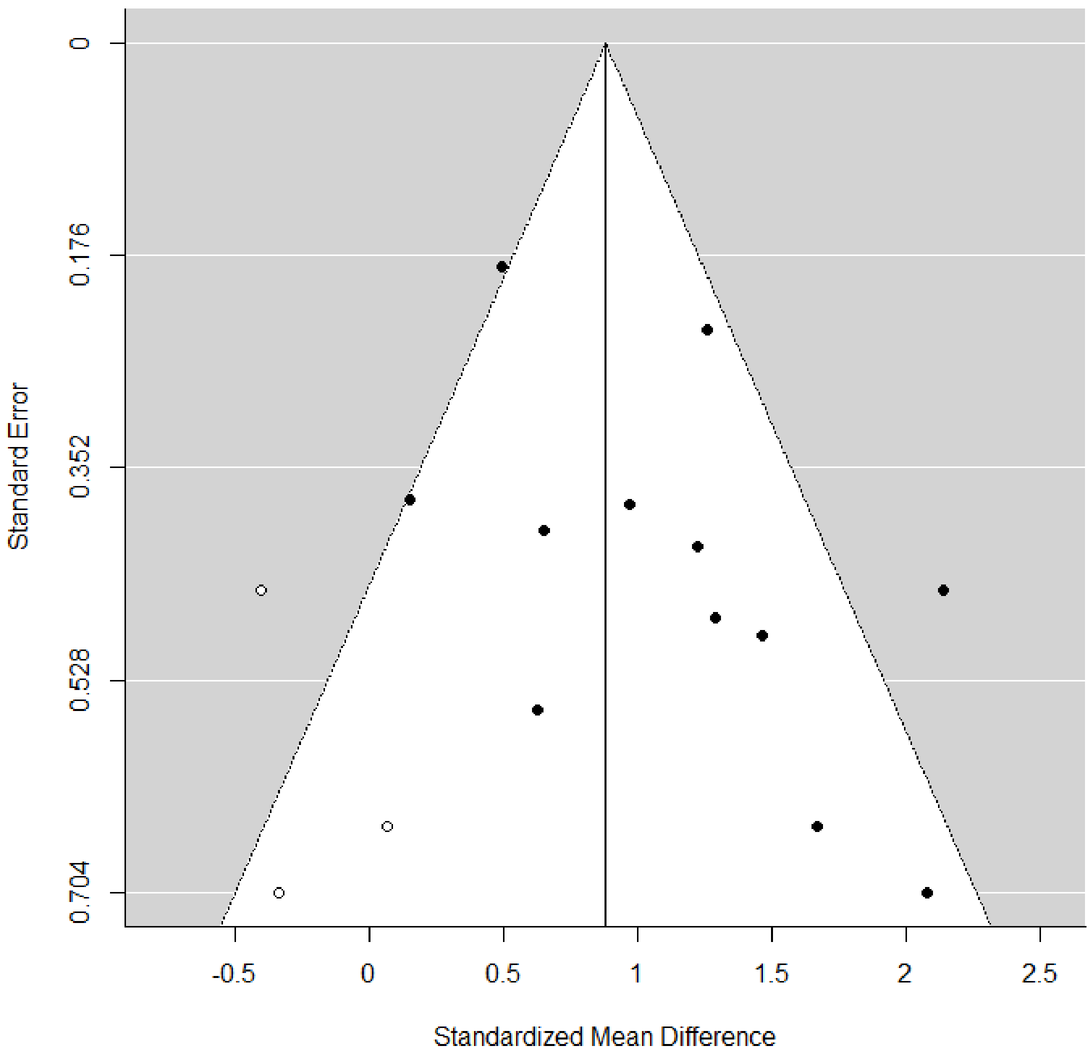

Figure 4. Funnel plot for real word reading analysis. *Source:* Viechtbauer (2010)

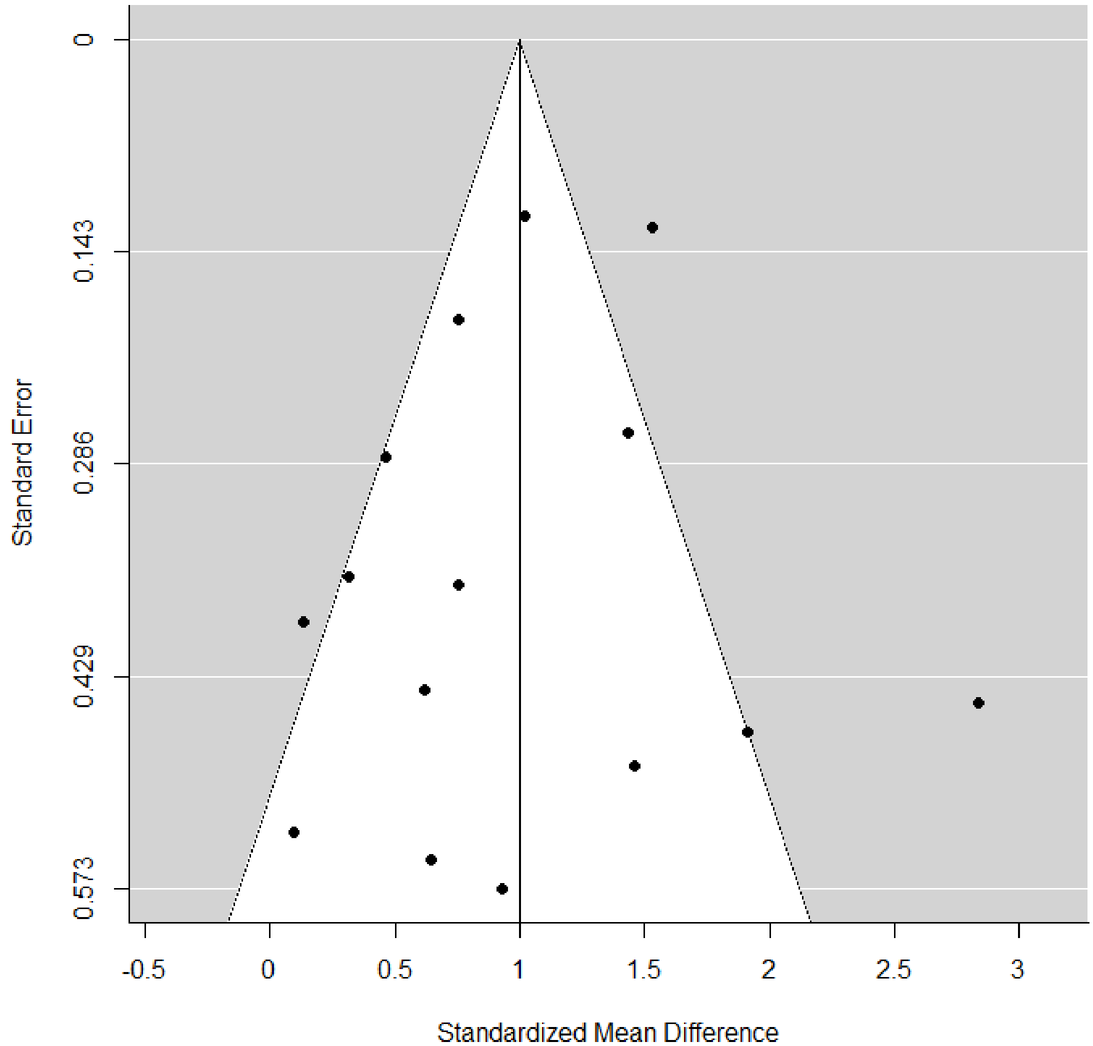

Figure 5. Funnel plot for nonword reading analysis. *Source:* Viechtbauer (2010)

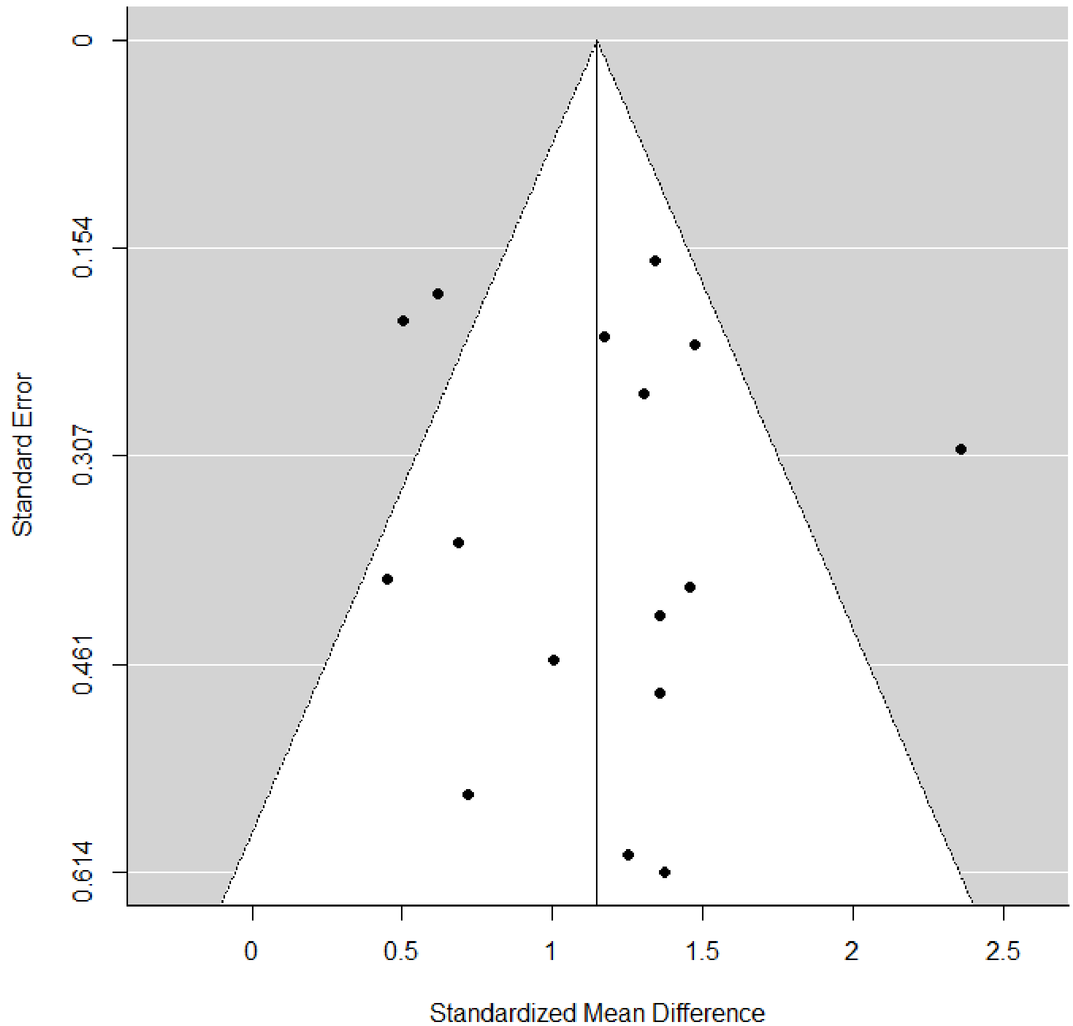

Figure 6. Funnel plot for combined word reading analysis. *Source:* Viechtbauer (2010)

Chapter 4: Discussion

The purpose of this meta-analysis was to examine the effects of reading interventions on the real word, nonword, and combined word reading skills of school-aged ELLs. For all three word-reading analyses, the Grand SMDs from pre-test to post-test were both large and positive, suggesting that the reading interventions effectively improved these students' word-reading skills overall. These effects were much larger than expected based on previous meta-analyses conducted with primarily monolingual English students (see e.g., Ehri et al., 2001a, 2001b).

For all three word-reading analyses, a considerable portion of the variability in effect sizes from sample to sample was systematic. The significant results of the heterogeneity analyses suggested that one or more moderator variables were impacting the variability in the size of the effects from pre-test to post-test. Of the moderators that were examined in this study, only two were found to be significant, and only for the real word reading samples. First, group size was significant, with intervention groups composed of more than five students being less effective than groups composed of two to five students. This aligns with our hypothesis, which predicted that large-group instruction would be less effective than individual or small-group instruction. It appears, though, that two studies, Lovett et al. (2008) and Rodriguez, Filler, and Higgins (2012), were driving this effect. These were the only studies that provided intervention to groups of more than five students, and both had small SMDs compared to the other studies in the real word reading analysis.

The second significant moderator was the length, in minutes, of each intervention session. Based on previous evidence that more instructional time does not always result in more learning (see Ehri et al., 2001b; Goodwin & Ahn, 2010), we had predicted that interventions of moderate intensity, or length, would be the most effective. Our findings generally confirmed this hypothesis, indicating that longer intervention sessions were less effective than shorter ones. However, it appears that the two studies driving the effect for group size, Lovett et al. (2008) and Rodriguez et al. (2012), were the ones driving this effect as well. These studies provided longer intervention sessions to large groups of students, rather than shorter sessions to small groups of students, as the other studies had done. The mean length of one intervention session in Lovett et al. (2008) and Rodriguez et al. (2012) was 75 minutes; in contrast, the mean length of one intervention session in all of the other samples in this analysis was 26.5 minutes. It is possible that the researchers of these two studies believed that more instructional time might counteract any negative consequences of larger group sizes. However, it appears that the longer intervention sessions, the larger group sizes, or both, contributed to the decreased effectiveness of these interventions.

The third moderator variable we examined, student risk status, was not significant in any of the word-reading analyses. One possible explanation for this may be that ELLs, as a group, could already be considered at risk. Some studies, such as the meta-analysis by Goodwin and Ahn (2010), considered limited English proficiency to be a risk factor, and subsequently regarded all ELLs as being at risk. Therefore, the ELLs who were coded as “at risk” in the present study may not be significantly different from the ELLs who were coded as “not at risk”, as they all have the risk factor of limited English proficiency. In addition, the criteria for coding ELLs as “at risk” or “not at risk” varied considerably from study to study. For example, Vaughn et al. (2006a) regarded students as at risk if they scored below the 25th percentile in word reading at pre-test; in contrast, Lovett et al. (2008) only coded students as at risk if they met the criteria for a reading disability based on three standardized reading measures and had below-average oral language skills. This variability in coding procedures across studies may have made consistent results for this moderator difficult to find.

As only two moderators in one word-reading analysis significantly accounted for some of the variability in the SMDs, there was much unexplained heterogeneity in this study. Population interval (PI) analyses indicated that for any random sample of ELLs, a given intervention may have very little effect on their real word and combined word reading skills, or it may have a very large effect. For nonword reading, the lower boundary of the PI crossed zero, suggesting that a given intervention could have no impact or even a negative impact. In contrast, the upper boundary showed that the intervention could have a very strong impact on ELLs' nonword reading ability. This wide PI quantifying the effectiveness of the interventions may seem surprising, but this lack of precision may have occurred because the type, content, and delivery of the interventions were so varied.

The diversity of the 14 intervention studies in this analysis made it challenging to group them in a meaningful way so that an analysis on our final moderator, type of intervention, could be run. Coding the interventions according to the primary reading skill that was targeted for intervention allowed us to create broad categories that could encompass all of the samples, with several samples in each group. However, coding the studies in this way did not fully reflect what the interventions entailed. For example, although vocabulary was the primary reading skill targeted in Nelson et al. (2011), other reading skills were also involved. Students worked on their decoding skills when they practiced reading target words, and blending, spelling, and comprehension activities were included to reinforce vocabulary learning. Sorting this intervention under “vocabulary” may have been the best fit for our moderator analysis, but it did not give a full picture of the intervention. This also meant that there was a great deal of diversity within each category, where one vocabulary—or phonological awareness, phonics, or other—intervention was very different from the next. Thus, it makes sense that a group of such diverse interventions would produce a wide range of effect sizes.

In addition to the type of intervention, other characteristics related to the content or delivery of the intervention may have contributed to some of the unexplained heterogeneity. For example, explicit, systematic instruction, ample opportunities for practice, and specific feedback from the instructor are all components of effective interventions (National Reading Panel, 2000), but are difficult to quantify and are often not described by researchers. Studies may vary greatly in how they structure the progression of lessons, manage student behaviour, or scaffold students when they require extra support. None of these features are directly comparable across studies as moderator variables, but they likely have some impact on the effectiveness of the interventions.

One major limitation of this research was the large amount of unexplained heterogeneity across samples, resulting in large confidence and population intervals. As previously discussed, this may have been due to instructional differences that were impossible to code, individual student differences, or other, unknown factors. The diversity among the interventions made some moderator variables, such as the type of intervention, impossible to code properly. Other moderators could not be investigated due to the small number of studies included in the analysis. For example, the first language of the ELLs was thought to be a potentially important moderator; however, the majority of samples were comprised of only Spanish-speaking ELLs, and the remaining samples included a mix of languages, making it impossible to run an effective moderator analysis. Similarly, we were unable to examine students' age as a moderator, as 11 of the 14 studies were conducted with a sample of students in Kindergarten or Grade 1. If there had been a larger group of studies in the meta-analysis, it may have been feasible to run analyses for these and other moderator variables.

Finally, because not all of the studies in our analysis included a control group, we were unable to compare the effect sizes of treatment and comparison groups. This is a particular limitation when you consider that the majority of the studies were conducted with Kindergarten and Grade 1 students. In the regular classroom setting, word-reading instruction typically begins in earnest in Kindergarten or Grade 1, when this type of instruction is most effective (Ehri et al., 2001a). Thus, we cannot be certain how much of the large improvement in students' word-reading skills could be attributed to the interventions, and how much was due to the instruction students received in their general classrooms. Although we could not examine the age of students as a moderator, we ran separate analyses on the samples with Kindergarten and Grade 1 students, and found that the interventions were only slightly more effective for this age group (the difference was $g \sim 0.1$). It seems, then, that the intensity of classroom word-reading instruction at this age had minimal effect on students' improvement from pre-test to post-test compared to the older students in the analysis. However, in all studies, other classroom supports put in place for the ELLs (e.g., oral language supports, additional time spent working with the teacher) may have also affected students' improvement over the course of the interventions. The absence of a comparison group in this meta-analysis makes it impossible to determine the effects of the interventions alone. Future reading intervention studies with ELLs should consider including a control group in order to address this issue.

As discussed previously, one reason for the large and unchanging achievement gap between ELLs and native English speakers may be that reading interventions for ELLs are often delayed until these students have developed adequate English oral skills (Dussling, 2016; Gunn, et al., 2005; Limbos & Geva, 2001). The findings of the present meta-analysis provide evidence that ELLs can benefit greatly from a wide variety of reading interventions. In fact, slightly greater improvements in word-reading skills were found for ELLs in Kindergarten and Grade 1 than for older students, suggesting that waiting for ELLs to reach a certain level of English proficiency before providing intervention is unnecessary. Previous research has indicated that even though ELLs have weaker literacy skills when entering school, they can catch up to their English-speaking peers if they receive strong instruction (Chiappe & Siegel, 2006; D'Anguilli, Siegel, & Maggi, 2004). This is critical, as a recent study reported that "one in six children who are not reading proficiently in third grade do not graduate from high school on time, a rate four times greater than that for proficient readers" (Hernandez, 2011, p. 3). Early intervention is essential for struggling readers, and ELLs are no exception. This meta-analysis demonstrates that reading interventions are highly effective for ELLs and should be implemented immediately upon their arrival at school.

References

Studies preceded by an asterisk were included in the meta-analysis.

Alberta Education. (2007). English as a second language: Guide to implementation, kindergarten to grade 9. Retrieved February 11, 2017, from https://education.alberta.ca/media/563809/esl-guide-to-implementation-k-9.pdf

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. West Sussex, UK: John Wiley & Sons, Ltd.

Cavanaugh, C. L., Kim, A., Wanzek, J., & Vaughn, S. (2004). Kindergarten reading interventions for at-risk students: Twenty years of research. *Learning Disabilities: A Contemporary Journal, 2,* 9-21.

Chiappe, P., & Siegel, L. S. (2006). A longitudinal study of reading development of Canadian children from diverse linguistic backgrounds. *The Elementary School Journal, 107*, 135-152. doi:10.1086/510652

Coe, R. (2002). It's the effect size, stupid: What effect size is and why it is important. *British Educational Research Association Annual Conference*, 1–18. Exeter, UK.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Cohen, J. (1992). Quantitative methods in psychology. *Psychological Bulletin, 112,* 115-159.

Cowley, P., & Easton, S. (2009, March). *Report card on Alberta's elementary schools 2009.* Retrieved from https://www.fraserinstitute.org/sites/default/files/70ABEEL09COMP.pdf

Cowley, P., & Easton, S. (2016, May). *Report card on Alberta's elementary schools 2016.* Retrieved from https://www.fraserinstitute.org/sites/default/files/results-alberta-elementary-school-rankings-2016.pdf

D'Angiulli, A., Siegel, L. S., & Maggi, S. (2004). Literacy instruction, SES, and word-reading achievement in English-language learners and children with English as a first language: A longitudinal study. *Learning Disabilities Research & Practice, 19*, 202-213. doi:10.1111/j.1540-5826.2004.00106.x

*Dussling, T. M. (2016). *The impact of a supplemental reading intervention on the literacy skills of English language learners and native English-speaking first grade children* (Unpublished doctoral dissertation). Syracuse University.

Ehri, L. C., Nunes, S. R., Stahl, S. A., & Willows, D. M. (2001a). Systematic phonics instruction helps students learn to read: Evidence from the National Reading Panel's meta-analysis. *Review of Educational Research, 71*, 393-447.

Ehri, L. C., Nunes, S. R., Willows, D. M., Schuster, B. V., Yaghoub-Zadeh, Z., & Shanahan, T. (2001b). Phonemic awareness instruction helps children learn to read: Evidence from the National Reading Panel's meta-analysis. *Reading Research Quarterly, 36*, 250-287.

Elbaum, B., Vaughn, S., Hughes, M. T., & Moody, S. W. (2000). How effective are one-to-one tutoring programs in reading for elementary students at risk for reading failure? A meta-analysis of the intervention research. *Journal of Educational Psychology, 92*, 605-619. doi:10.1037//00224)663.92.4.605

Gersten, R. (1996). The double demands of teaching English language learners. *Educational Leadership, 53*(5), 18-22.

Gersten, R., Baker, S., Shanahan, T., Linan-Thompson, S., Collins, P., & Scarcella, R. (2007). *Effective literacy and English language instruction for English learners in the elementary grades* (NCEE 2007-4011). Washington, DC: National Center for Education Evaluation and Regional Assistance. Retrieved from http://files.eric.ed.gov/fulltext/ED497258.pdf

Goodwin, A. P., & Ahn, S. (2010). A meta-analysis of morphological interventions: Effects on literacy achievement of children with literacy difficulties. *Annals of Dyslexia, 60*, 183-208. doi:10.1007/s11881-010-0041-x

Gough, P. B. (1996). How children learn to read and why they fail. *Annals of Dyslexia, 46*, 1-20.

Gunn, B., Smolkowski, K., Biglan, A., Black, C., & Blair, J. (2005). Fostering the development of reading skill through supplemental instruction: Results for Hispanic and non-Hispanic students. *The Journal of Special Education, 39*, 66-85. doi:10.1177/00224669050390020301

Hatcher, P. J., Hulme, C., & Ellis, A. W. (1994). Ameliorating early reading failure by integrating the teaching of reading and phonological skills: The phonological linkage hypothesis. *Child Development*, *65*, 41–57.

Hernandez, D. J. (2011). *Double jeopardy: How third-grade reading skills and poverty influence high school graduation.* Baltimore, MD: The Annie E. Casey Foundation.

Kena, G., Aug, S., Johnson, F., Wang, X., Zhang, J., Rathum, A., . . . Kristapovich, P. (2014). *The condition of education 2014 (NCES 2014-083).* U. S. Department of Education, National Center for Education Statistic. Washington, DC. Retrieved from http://nces.ed.gov/pubs2014/2014083.pdf

Kendeou, P., van den Broek, P., White, M. J., & Lynch, J. S. (2009). Predicting reading comprehension in early elementary school: The independent contributions of oral language and decoding skills. *Journal of Educational Psychology*, *101*, 765–778. doi:10.1037/a0015956

*Leafsteadt, J. M., Richards, C. R., & Gerber, M. M. (2004). Effectiveness of explicit phonological-awareness instruction for at-risk English learners. *Learning Disabilities Research & Practice, 19*, 252-261. doi:10.1111/j.1540-5826.2004.00110.x

Lesaux, N. K. (2006). Building consensus: Future directions for research on English language learners at risk for learning disabilities. *Teachers College Record, 108*, 2406–2438.

Li, G., & Edwards, P. A. (2010). *Best practices in ELL instruction*. New York, NY: Guilford Press.

Limbos, M. M., & Geva, E. (2001). Accuracy of teacher assessments of second-language students at risk for reading disability. *Journal of Learning Disabilities, 34*, 136-151. doi: 10.1177/002221940103400204

Linan-Thompson, S., & Hickman-Davis, P. (2002). Supplemental reading instruction for students at risk for reading disabilities: Improve reading 30 minutes at a time. *Learning Disabilities Research & Practice, 17*, 242–251.

*Linan-Thompson, S., Vaughn, S., Hickman-Davis, P., & Kouzekanani, K. (2003). Effectiveness of supplemental reading instruction for second-grade English language learners with reading difficulties. *The Elementary School Journal, 103*, 221-238. doi:10.1086/499724

*Lovett, M. W., De Palma, M., Frijters, J., Steinbach, K., Temple, M., Benson, N., & Lacerenza, L. (2008). Interventions for reading difficulties: A comparison of response to intervention by ELL and EFL struggling readers. *Journal of Learning Disabilities, 41*, 333-352. doi:10.1177/0022219408317859

National Center for Education Statistics. (2015). Fourth-grade reading scores increase for two demographic groups compared to 2013. Retrieved February 21, 2017, from https://www.nationsreportcard.gov/reading_math_2015/#reading/groups?grade=4

National Reading Panel. (2000). Report of the National Reading Panel: Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction: Reports of the subgroups. Washington, D.C.: National Institute of Child Health and Human Development Clearinghouse.

*Nelson, J. R., Vadasy, P. F., & Sanders, E. A. (2011). Efficacy of a tier 2 supplemental root word vocabulary and decoding intervention with kindergarten Spanish-speaking English learners. *Journal of Literacy Research, 43*, 184-211. doi:10.1177/1086296x11403088

Robertson, D., & Stoddard, S. (2016, February 16). Edmonton Public Schools: Information report. Retrieved from https://www.epsb.ca/media/epsb/ourdistrict/boardoftrustees/boardmeetings/2015-16/february16/03-SuccessforEveryStudent-EnglishLanguageLearners.pdf

*Rodríguez, C. D., Filler, J., & Higgins, K. (2012). Using primary language support via computer to improve reading comprehension skills of first-grade English language learners. *Computers in the Schools, 29*, 253-267. doi:10.1080/07380569.2012.702718

Rosenthal, R. (1995). Writing meta-analytic reviews. *Psychological Bulletin, 118*, 183-192.

Scammacca, N. K., Roberts, G., Vaughn, S., & Stuebing, K. K. (2015). A meta-analysis of interventions for struggling readers in grades 4-12: 1980-2011. *Journal of Learning Disabilities, 48*, 369-390. doi:10.1177/0022219413504995

*Solari, E. J., & Gerber, M. M. (2008). Early comprehension instruction for Spanish-speaking English language learners: Teaching text-level reading skills while maintaining effects on word-level skills. *Learning Disabilities Research & Practice, 23*, 155-168. doi:10.1111/j.1540-5826.2008.00273.x

Statistics Canada. (2001). Mother tongue (8), age groups (13) and sex (3) for population, for Canada, provinces, territories, census divisions and census subdivisions, 2001 census. Retrieved from http://www12.statcan.gc.ca/English/census01/products/standard/themes/Rp-eng.cfm?LANG=E&APATH=3&DETAIL=1&DIM=0&FL=A&FREE=1&GC=0&GID=0&GK=0&GRP=1&PID=55787&PRID=0&PTYPE=55430&S=0&SHOWALL=No&SUB=0&Temporal=2006&THEME=41&VID=0&VNAMEE=&VNAMEF=

Statistics Canada. (2011). Mother tongue (8), age groups (25) and sex (3) for the population of Canada and forward sortation areas, 2011 census. Retrieved from http://www12.statcan.gc.ca/census-recensement/2011/dp-pd/tbt-tt/Rp-eng.cfm?LANG=E&APATH=3&DETAIL=0&DIM=0&FL=A&FREE=0&GC=0&GID=0&GK=0&GRP=1&PID=102841&PRID=10&PTYPE=101955&S=0&SHOWALL=0&SUB=0&Temporal=2011&THEME=90&VID=0&VNAMEE=&VNAMEF

Stuart, M. (1999). Getting ready for reading: Early phoneme awareness and phonics teaching improves reading and spelling in inner-city second language learners. *British Journal of Educational Psychology, 69*, 587-605. doi:10.1348/000709999157914

Stuart, M. (2004). Getting ready for reading: A follow-up study of inner city second language learners at the end of Key Stage 1. *British Journal of Educational Psychology, 74*, 15-36. doi:10.1348/000709904322848806

Suggate, S. P. (2010). Why what we teach depends on when: Grade and reading intervention modality moderate effect size. *Developmental Psychology, 46*, 1556-1579. doi:10.1037/a0020612

Suggate, S. P. (2016). A meta-analysis of the long-term effects of phonemic awareness, phonics, fluency, and reading comprehension interventions. *Journal of Learning Disabilities, 49*, 77-96. doi:10.1177/0022219414528540

*Vadasy, P. F., & Sanders, E. A. (2010). Efficacy of supplemental phonics-based instruction for low-skilled kindergarteners in the context of language minority status and classroom phonics instruction. *Journal of Educational Psychology, 102*, 786-803. doi:10.1037/a0019639

*Vadasy, P. F., & Sanders, E. A. (2011). Efficacy of supplemental phonics-based instruction for low-skilled first graders: How language minority status and pretest characteristics moderate treatment response. *Scientific Studies of Reading, 15*, 471-497. doi:10.1080/10888438.2010.501091

*Vadasy, P. F., & Sanders, E. A. (2015). Attention to orthographic and phonological word forms in vocabulary instruction for kindergarten English learners. *Reading Psychology, 37*, 833-866. doi:10.1080/02702711.2015.1116477

*Vadasy, P. F., Sanders, E. A., & Nelson, J. R. (2015). Effectiveness of supplemental kindergarten vocabulary instruction for English learners: A randomized study of immediate and longer-term effects of two approaches. *Journal of Research on Educational Effectiveness, 8*, 490-529. doi:10.1080/19345747.2015.1033120

*Vaughn, S., Cirino, P. T., Linan-Thompson, S., Mathes, P. G., Carlson, C. D., Hagan, E. C., . . . Francis, D. J. (2006a). Effectiveness of a Spanish intervention and an English intervention for English-language learners at risk for reading problems. *American Educational Research Journal, 43*, 449-487. doi:10.3102/00028312043003449

Vaughn, S., Linan-Thompson, S., Kouzekanani, K., Bryant, D. P., Dickson, S., & Blozis, S. A. (2003). Reading instruction grouping for students with reading difficulties. *Remedial and Special Education, 24*, 301-315. doi:10.1177/07419325030240050501

*Vaughn, S., Mathes, P., Linan- Thompson, S., Cirino, P., Carlson, C., Pollard- Durodola, S., . . . Francis, D. (2006b). Effectiveness of an English intervention for first- grade English language learners at risk for reading problems. *The Elementary School Journal, 107*, 153-180. doi:10.1086/510653

Viechtbauer, W. (2010). Conducting meta-analyses in *R* with the metafor package. *Journal of Statistical Software*, *36*(3). https://doi.org/10.18637/jss.v036.i03

Viechtbauer, W. (2014). *Meta-Analysis Package for R*.

Wagner, R. K., Francis, D. J., & Morris, R. D. (2005). Identifying English language learners with learning disabilities: Key challenges and possible approaches. *Learning Disabilities Research and Practice, 20*, 6-15.

Wanzek, J., & Vaughn, S. (2007). Research-based implications from extensive early reading interventions. *School Psychology Review, 36*, 541-561.

*Wanzek, J., & Roberts, G. (2012). Reading interventions with varying instructional emphases for fourth graders with reading difficulties. *Learning Disability Quarterly, 35*, 90-101. doi:10.1177/0731948711434047

Zabjek, A. (2015, October 9). As many as one-quarter of all Edmonton students learning English as second language. *Edmonton Journal.* Retrieved from http://edmontonjournal.com/news/local-news/as-many-as-one-quarter-of-all-edmonton-students-learning-english-as-second-language

Zaykin, D. V. (2011). Optimally weighted Z-test is a powerful method for combining probabilities in meta-analysis. *Journal of Evolutionary Biology, 24,* 1836-1841. doi:10.1111/j.1420-9101.2011.02297.x

Appendix A

Participant Characteristics in Each Study Included in the Meta-Analysis

Study	*n*	Grade	Male:Female	First Language	Student Risk Status[a]
Dussling (2016)	7	1	4:3	Mix (Tamil, Bosnian, Turkish, Russian, French, Lingala, Tshiluba)	At risk
Leafstedt et al. (2014)	16	K	5:4	Spanish	At risk
Linan-Thompson et al. (2003)	26	2	6:7	Spanish	At risk
Lovett et al. (2008)	60	2-8	Not reported	Mix (Portuguese, Spanish, Tagalog, Italian, Arabic, Syrian, Urdu)	At risk
Nelson et al. (2011)	93	K	16:15	Spanish	Not at risk
Rodriguez et al. (2012)	14	1	9:5	Spanish	Not at risk
Solari & Gerber (2008)	T1 at risk: 7 T1 not at risk: 13 T2 at risk: 7 T2 not at risk: 14	K	21:20	Spanish	Mix of at risk and not at risk
Vadasy & Sanders (2010)	38	K	1:1	Mix (Spanish, Vietnamese, Chinese, Somali, Tagalog)	Not at risk
Vadasy & Sanders (2011)	48	1	17:8	Mix (Spanish, Vietnamese, Chinese, Somali, Tagalog)	At risk
Vadasy & Sanders (2015)	T1: 51 T2: 49	K	T1 – 8:9 T2 – 30:19	Mix of 17 languages	Not at risk

Appendix A continued

Study	N	Grade	Male:Female	First Language	Student Risk Status[a]
Vadasy et al. (2015)	T1: 163 T2: 161	K	T1 – 84:79 T2: 99:62	Mix (Spanish, African languages, Asian languages)	Not at risk
Vaughn et al. (2006a)	43	1	27:23	Spanish	At risk
Vaughn et al. (2006b)	22	1	1:1	Spanish	At risk
Wanzek & Roberts (2012)	T1: 11 T2: 15 T3: 11	4	Not reported	Spanish	At risk

Note. K = Kindergarten; T1 = treatment 1; T2 = treatment 2; T3 = treatment 3.
[a]Student risk status classifies students as either at risk or not at risk for reading difficulties.

Appendix B

Intervention Characteristics

Study	Reading Outcome(s)	Outcome Measure(s)	Primary reading skill targeted	Number of students in each group	Sessions per week	Minutes per session	Number of weeks	Total minutes of intervention
Dussling (2016)	Real word, nonword	WRMT-R/NU Word ID and Word Attack	Phonics	4 or 5	5	30	6	900
Leafstedt et al. (2014)	Real word, nonword	WJ-III Word ID and Word Attack	PA	3 to 5	2	15	10	300
Linan-Thompson et al. (2003)	Nonword	WRMT-R Word Attack	Mix (fluency, PA, phonics)	1 to 3	5	30	13	1950
Lovett et al. (2008)	Real word, nonword	WRMT-R Word ID and Word Attack, WRAT-3 Reading (Blue form)	Phonics	4 to 8	4.5	60	Not reported	6300
Nelson et al. (2011)	Combined	WRMT-R/NU Word ID and Word Attack	Vocabulary	2 to 5	5	20	20	2000
Rodriguez et al. (2012)	Real word	WMLS-R Letter Word ID	Phonics and PA	14	3	90	8	2160
Solari & Gerber (2008)	Real word, nonword	WJ-III, Word ID and Word Attack	PA and comprehension	4 or 5	3	20	8	480

Appendix B continued

Study	Reading Outcome(s)	Outcome Measure(s)	Primary reading skill targeted	Number of students in each group	Sessions per week	Minutes per session	Number of weeks	Total minutes of intervention
Vadasy & Sanders (2010)	Combined	WRMT–R/NU Word ID and Word Attack	Phonics	1	4	30	18	2160
Vadasy & Sanders (2011)	Combined	WRMT-R/NU Word ID and Word Attack	Phonics	1	4	30	20	2400
Vadasy & Sanders (2015)	Combined	WRMT-R/NU Word ID and Word Attack	Vocabulary	1	4	15	14	840
Vadasy et al. (2015)	Nonword	WRMT-R/NU Word Attack	Vocabulary	2 or 3	4	30	20	2400
Vaughn et al. (2006a)	Real word, nonword	WLPB-R Letter Word ID and Word Attack	Mix (PA, phonics, word recognition, fluency, comprehension)	3 to 5	5	50	Not reported	5750
Vaughn et al. (2006b)	Nonword	WLPB-R Word Attack	Mix (PA, phonics, word recognition, fluency, comprehension)	3 to 5	5	50	28	7000

Appendix B continued

Study	Reading Outcome(s)	Outcome Measure(s)	Primary reading skill targeted	Number of students in each group	Sessions per week	Minutes per session	Number of weeks	Total minutes of intervention
Wanzek & Roberts (2012)	Real word, nonword	WJ-III Letter Word ID and Word Attack	Phonics and comprehension	2 to 4	4	30	28	3360

Note. Letter Word ID = Letter Word Identification subtest; PA = phonological awareness; WJ-III = Woodcock-Johnson III Tests of Achievement; WLPB-R = Woodcock Language Proficiency Battery--Revised; WMLS-R = Woodcock-Muñoz Language Survey--Revised; WRAT-3 Reading (Blue form) = Wide Range Achievement Test, Third Edition; WRMT-R = Woodcock Reading Mastery Test--Revised; WRMT-R/NU = Woodcock Reading Mastery Test--Revised/Normative Update; Word Attack = Work Attack subtest; Word ID = Word Identification subtest.