

# SUSTAINABILITY OF REMEDIAL READING INTERVENTION EFFECTS

## A SYNTHESIS OF THE SUSTAINABILITY OF REMEDIAL READING INTERVENTION EFFECTS FOR STRUGGLING ADOLESCENT READERS

Johny Daniel, Ph.D.<sup>1</sup>

Philip Capin, Ph.D.<sup>1</sup>

Paul Steinle, M.Ed.<sup>1</sup>

1. University of Texas at Austin

### **Corresponding author:**

Johny Daniel,

University of Texas at Austin, Department of Special Education,

1912 Speedway D4900, Austin, TX – 78712

[Johny.daniel@utexas.edu](mailto:Johny.daniel@utexas.edu)

### **Other authors' email addresses:**

Dr. Philip Capin – [pcapin@utexas.edu](mailto:pcapin@utexas.edu)

Paul Steinle – [paulksteinle@utexas.edu](mailto:paulksteinle@utexas.edu)

### **Authors' Note:**

This research was supported by grant P50 HD052117 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development at the National Institutes of Health. The content is solely the responsibility of the authors and does not necessarily represent the official views of the Eunice Kennedy Shriver National Institute of Child Health and Human Development or the National Institutes of Health.

A Synthesis of the Sustainability of Remedial Reading Intervention Effects for Struggling  
Adolescent Readers

### Abstract

A majority of reading-related intervention studies aiming to remediate struggling readers' reading outcomes assess student performance immediately following the conclusion of an intervention to determine intervention effects. Few studies collect follow-up data to measure the long-term sustainability of treatment effects. Hence, the aim of the current synthesis was to examine follow-up intervention effects of reading interventions involving adolescent struggling readers in Grades 6–12. Our literature search yielded only ten studies that reported follow-up data for intervention participants, which highlights the dearth of intervention research that examines sustainability of intervention effects. Of the ten included studies, the weighted mean effect size for all reading outcome measures was  $g_w = .78$  at immediate posttest and  $g_w = .27$  at follow-up, in favor of treatment group students. Although the magnitude of difference between treatment and control groups diminished at follow-up time, a comparison of treatment group students' immediate posttest and follow-up scores showed that students mostly maintained gains made during intervention at follow-up time points.

*Keywords:* Adolescent struggling readers, learning disabilities, follow-up, maintenance, reading interventions.

## A Synthesis of the Sustainability of Remedial Reading Intervention Effects for Struggling Adolescent Readers

*“...socially significant behavior changes are those that last over time, are used by the learner in all relevant settings and situations, and are accompanied by changes in other relevant responses...to perform below this standard is more than just regrettable; it is a clear indication that the initial instruction was not entirely successful.”* (Cooper et al., 2008, p. 623).

Reading intervention studies that aim to improve adolescent struggling readers’ reading outcomes generally measure and report the efficacy of an intervention based on students’ immediate posttest reading performance. However, there is inadequate research addressing the sustainability of intervention effects as a measure of intervention effectiveness (Suggate, 2016). In other words, little is known about adolescent struggling readers’ ability to maintain gains made due to interventions. Evaluating student performance at follow-up time points can further demonstrate a program’s effectiveness or, alternatively, detect program’s that lead to only short-term gains (Keogh, 2004; Suggate, 2016). More importantly, performance on follow-up tests can add substantial scientific value to the evaluation of reading interventions for adolescent struggling readers.

### **Effectiveness of Reading Interventions for Adolescent Struggling Readers**

A substantial body of research exists on examining the effects of instructional methods for students who struggle to read and comprehend grade-level text in middle and high school (e.g., Boardman et al., 2008; Robert et al., 2008; Scammacca et al., 2007; Vaughn et al. 2019; Vaughn et al., 2015). However, these studies have generally yielded mixed findings. For instance, the Striving Readers Project (Boulay et al., 2015), funded by the Institute of Education

Sciences (IES), summarized findings from 17 randomized controlled trials that evaluated the effects of ten different reading interventions for Grades 6–10 struggling adolescent readers (Cantrell et al., 2011; Cantrell et al., 2012; Cantrell et al., 2010; Deussen et al., 2012; Dimitrov et al., 2012; Faddis et al., 2012; Feldman et al., 2011; Hofstetter et al., 2011; Loadman et al., 2012; Meisch et al.; 2011; Newman et al., 2012; Schenck et al., 2012; Schiller et al., 2012; Swanlund et al., 2012; The Education Alliance at Brown University, 2012; Tunik et al., 2011; Vaden-Kiernan et al., 2012). The IES report summarizing the findings rated each intervention's effect on student reading outcomes. Of the ten intervention studies, IES summarized that six reported no discernable effects, three reported positive or potentially positive effects, and one study reported mixed effects of intervention on students' reading outcomes.

Similarly, meta-analyses and syntheses that aggregate the results of multiple studies also provide a mixed picture of the effects of secondary reading interventions. Some past meta-analyses have reported moderate effects of reading interventions for struggling readers in upper elementary and later grades:  $g = 0.41$  (Flynn et al., 2012);  $g = 0.47$  (Edmonds et al., 2009);  $g = 0.49$  (Scammacca et al., 2015). In contrast, Wanzek et al. (2013), measuring the effects of extensive (i.e., comprising of 100 or more sessions) reading interventions, reported small effects of interventions on various reading outcomes ( $g = 0.10$  to  $0.16$ ). Additionally, Scammacca et al. (2015) disaggregated results of interventions and reported much smaller effects for standardized reading outcome measures ( $g = 0.21$ ) with multicomponent reading interventions demonstrating the largest positive effect on standardized reading comprehension measures ( $g = 0.46$ ).

Past systematic reviews and meta-analyses also report on the type of interventions that are most effective in improving students' reading outcomes for struggling adolescent readers. For instance, Scammacca et al. (2007) reported large effects of comprehension strategy

instruction ( $d = 1.23$ ), vocabulary instruction ( $d = 1.62$ ), and word study instruction ( $d = 1.60$ ) on various researcher-developed and standardized reading measures. In a subsequent meta-analysis, Scammacca and colleagues (2015) reported large effects of vocabulary interventions ( $d = 1.58$ ) and reading comprehension interventions ( $d = 0.74$ ) on adolescent struggling readers' reading outcomes. However, it is important to note that across these meta-analyses, researchers have generally reported substantial differences in effects between researcher-developed and standardized measures with greater effects observed on researcher-developed measures (e.g., Edmonds et al., 2009; Scammacca et al., 2015). For instance, Scammacca et al., (2015) reported that while the overall effect size across all included studies and measures was 0.49, the average reported effect size was 0.21 on standardized measures.

In a more recent systematic review, Berkeley and Larsen (2018) reviewed the extent to which self-regulation of reading strategies benefited adolescent students with learning disabilities (LD). Researchers (Berkeley & Larsen, 2018) reported the average effect across 18 studies, on predominantly researcher-developed reading measures, was large at posttest ( $ES = 1.35$ ). Additionally, eight of the eighteen included studies reported follow-up data that showed treatment group students continued to exhibit improved performance on reading measures compared to their control group peers. The average follow-up effect was also large ( $ES = 0.95$ ); however, most studies assessed maintenance effects using researcher developed measures. This finding is vital in evaluating the benefits of embedding self-regulation elements to make a long-lasting impact on students' reading performance.

In summary, several past studies have implemented a variety of reading interventions to improve reading outcomes for adolescent struggling readers. One challenge with interpreting the effects of past reading intervention studies is that a majority of interventions, and almost all past

systematic reviews of these studies, focus on student performance at the end of the intervention period. Rarely do researchers follow study participants to analyze the long-term effects of interventions. Thus, the goal of this review is to examine the sustainability of reading intervention effects observed at immediate posttest compared to follow-up time points. The current review is also not limited to any one type of reading intervention (e.g., self-regulation strategy; See Berkeley & Larsen, 2018) but aims to evaluate the sustainable effects of a variety of reading interventions that target different components of reading (i.e., comprehension, vocabulary, word reading, and fluency). Follow-up is defined in this review as any data point collected two or more weeks after the end of the original intervention.

### **Importance of Follow-up Data**

Researchers have advocated for the collection of follow-up data to better assess the effectiveness of educational interventions (Keogh, 2004; Suggate, 2016). Those who collected follow-up data for early reading intervention studies generally reported positive maintenance effects of phonological awareness and phonics instruction on reading outcomes for low performing students in Grades K–3 (e.g., Blachman et al., 2004; Blachman et al., 2014; Ryder et al., 2008; Vadasy & Sanders, 2013). These studies have contributed to the growing body of evidence emphasizing the importance of code-oriented instruction in early elementary education, especially for low performing students. Results indicate that benefits of instruction extended from one to ten years after the intervention concluded (e.g., Blachman et al., 2014; Byrne & Fielding-Barnsley, 1993, 1995; Ryder et al., 2008; Vadasy et al., 2006).

Indeed, a recent meta-analysis (Suggate, 2016) of the follow-up effects of reading interventions targeting students in Grades PK–6 corroborates the effectiveness of phonological awareness instruction for typical and low performing students. Suggate (2016) identified 71

reading intervention studies with an average follow-up time of approximately 11 months. Among the 17 phonemic awareness intervention studies identified, the average effect across all reading measures at immediate posttest ( $d_w = .43$ ) was mostly sustained at follow-up time points ( $d_w = .36$ ). Similarly, averaging across all reading measures, positive effects of comprehension interventions were also sustained from immediate posttest ( $d_w = .38$ ) to follow-up ( $d_w = .46$ ) for treatment group students. In contrast, effects of fluency interventions on students' overall reading outcomes diminished from posttest ( $d_w = .47$ ) to follow-up ( $d_w = .28$ ). More surprisingly, effects of phonics interventions on reading outcomes were significant at immediate posttest ( $d_w = .29$ ) but were trivial at follow-up time point ( $d_w = .07$ ); Suggate (2016) hypothesized that the diminished performance at follow-up may be due to a stronger counterfactual rather than loss of learning for the treatment group students.

While the primary focus of early-elementary reading instruction is to develop students' word reading skills, in the upper elementary and later grades the focus of instruction shifts to extracting and constructing meaning from text (Chall & Jacobs, 2003). Results from previous early-elementary intervention studies and Suggate's (2016) meta-analysis establish the long-term benefits of implementing early reading interventions, especially for at-risk student populations. However, while considerable evidence supports the effectiveness and extended benefits of reading interventions in early elementary grades, there is much less evidence confirming the long-term benefits of effective middle and high school reading instructional practices for struggling adolescent readers.

### **Current Study**

Although our understanding of the effects of reading interventions for adolescents is growing, no previous synthesis has examined the long-term effects of these interventions. This



paper will serve as an upward extension of the Suggate (2016) meta-analysis, which examined the long-term effects of elementary reading interventions. However, unlike Suggate's (2016) meta-analysis, which focused on the ways struggling readers' response to interventions varied from typical peers, this synthesis focuses solely on the reading outcomes of struggling readers. Thus, the aim of this synthesis is to address the following research question: What are the effects of reading interventions provided in small-group settings on reading outcomes for struggling readers in Grades 6–12 at immediate posttest and follow-up time points?

## Method

### Data Collection

A comprehensive search of the literature was conducted. First, an online search of four educational literature databases was conducted on Education source, ERIC, PsycINFO, and ProQuest Dissertation and Theses Global to locate unpublished and published studies between 1996 and August 2019. We searched abstracts using search terms for *reading* (reading OR vocabulary OR phon\* OR fluency OR decod\* OR comprehen\*), *study type* (intervention OR strateg\* OR curricul\* OR approach\* OR treatment OR teaching method\* OR instruction\* OR teaching aids OR program;), *sample* (disabilit\* OR disorder OR delay\* OR struggling OR "reading problem\*" OR dyslexi\* OR "learning problem\*" OR "special education" OR "special need\*" OR "at risk" OR "high risk" OR "mild handicap\*" OR reading difficult\*), and *follow-up data* ("long-term" OR "medium-term" OR "follow-up" OR posttest OR "post test" OR longitudinal OR period OR maint\*). Compared to this study's screening process, Suggate's (2016) literature search was limited to two databases (i.e., ERIC and PsycINFO). Our search terms for the follow-up data were the same, and we added 'vocabulary' to the reading search terms. There was no overlap for the *study type* and *sample* search terms.

The second step in identifying articles relevant to the research question involved a hand search of 14 prominent educational journals spanning from January, 2017 through August, 2019. This two-year window ensured that the electronic search captured all relevant articles. The hand searched journals included: *Annals of Dyslexia*, *Cognition and Instruction*, *Exceptional Children*, *Journal of Educational Psychology*, *Journal of Learning Disabilities*, *Journal of Research on Educational Effectiveness*, *Journal of Research in Reading*, *Journal of Special Education*, *Learning Disability Quarterly*, *Learning Disabilities Research and Practice*, *Reading and Writing Quarterly*, *Reading Psychology*, *Reading Research Quarterly*, *Remedial and Special Education*, and *Scientific Studies of Reading*. Additionally, relevant articles were sourced via an ancestry search of articles that fit the inclusion criteria. Finally, we did an ancestry search of existing reviews that synthesized the effects of reading interventions for adolescent struggling readers (Berkeley & Larsen, 2018; Edmonds et al., 2009; Scammacca et al., 2016; Scammacca et al., 2007; Scammacca et al., 2015).

Figure 1 shows the process for including studies for this systematic review. The online database search revealed 22,770 potential articles. The first author screened abstracts and included any abstract related to reading interventions for full-text screening ( $n = 904$ ). Table 1 provides examples of studies that were not included in this synthesis. A total of ten studies (six peer reviewed articles and four dissertations) met all inclusion criteria:

(a) interventions involving participants identified with LD, dyslexia, or struggling readers in Grades 6-12;

(b) studies that were randomized controlled trials or quasi-experimental designs;

(c) intervention studies targeting English language reading related skills such as decoding, fluency, vocabulary, reading comprehension, or multicomponent reading interventions;

(d) studies that reported immediate posttest and maintenance data for at least one dependent measure that assessed reading-related outcomes;

(e) reading related interventions conducted in school settings (i.e. no summer school or home-based literacy program);

(f) reading related interventions conducted in school settings outside the general education classroom;

(g) studies published between January, 1996 and August, 2019;

(h) studies published in a peer reviewed journal or unpublished dissertations;

(i) studies available in English.

The target sample of studies were experimental or quasi-experimental reading interventions that reported reading-related outcome data for immediate posttest and for at least one reading measure at a follow-up time point. The authors made an *a priori* decision to exclude single case design studies due to inconsistencies in the number of data points reported in the intervention and maintenance (i.e., follow-up) phases; typically, several data points are reported in the intervention phase while one/two data points are reported for the maintenance/follow-up phase. This imbalance would lead to skewed Tau-*U* effect sizes as the magnitude of difference is dependent on the length of the phases for effect size calculation (J. E. Pustejovsky, personal communication, July 08, 2018). Thus, following Suggate's (2016) lead, only group design studies were considered for this review.

### **Study Coding**

All included studies were coded using the Guide for Education-Related Intervention Study Syntheses (Vaughn et al., 2014). This codesheet has been used in numerous previous syntheses (e.g., Daniel & Williams, 2019; Hall et al., 2017; Scammacca et al., 2016) and includes all critical components identified in the systematic review process of the What Works Clearinghouse (WWC) Study Review Guide (Clearinghouse, 2017). Critical components included in the codesheet include design information, sample description; sample sizes; baseline measures; measures' description including validity, reliability, and internal consistency information of each measure; data used for analysis; attrition information; description of treatment and control groups; and description of treatment and control group procedures. Furthermore, the codesheet was updated to include follow-up data outcome measures and scores, and times when follow-up data were collected.

### **Data Analysis**

Standardized mean difference effect sizes were calculated using Hedges's  $g$  to adjust for the possibility of small sample bias. Treatment and comparison groups' immediate posttest and follow-up means, standard deviations, and sample sizes were used to calculate Hedges's  $g$ . Additionally, we sought to examine the sustainability of effects for treatment group students by calculating an effect size comparing treatment group students' immediate posttest and follow up mean outcome scores.

All eligible effect sizes in each study that provided mean and standard deviation or other relevant statistics such as  $F$ -test scores were considered in calculating the weighted mean effect size. Group design studies contributed multiple effect sizes when the sample for each effect size was independent. For studies that reported multiple effect sizes from the same sample (e.g., two effect sizes based on two reading comprehension measures were calculated for treatment versus

control in one study), analysis also accounted for the statistical dependencies using the random effects robust standard error estimation technique developed by Hedges et al. (2010). This analysis allows for clustered data (i.e., effect sizes nested within samples) by correcting the study standard errors to take into account the correlations between effect sizes from the same sample. The robust standard error technique requires that an estimate of the mean correlation ( $\rho$ ) between all the pairs of effect sizes within a cluster be estimated for calculating the between-study sampling variance estimate,  $\tau^2$ . In all analyses, we estimated  $\tau^2$  with  $\rho = .80$ . Because this review included studies conducted in Grades 6-12, it was hypothesized that the research body was reporting a distribution of effect sizes with significant between-studies variance, as opposed to a group of studies attempting to estimate one true effect size (Lipsey & Wilson, 2001). Thus, a random-effect model was used for the current study. Robust variance estimation analysis was conducted in R using the robumeta package (Fisher & Tipton, 2015; See Appendix A).

The WWC recommends interpreting effect sizes of 0.25 and larger as “substantially important” in educational research settings (Clearinghouse, 2017). This recommendation was considered when interpreting the magnitude and importance of the effects. Finally, descriptive statistical data were used to calculate 95% confidence intervals to determine if each individual effect size was significant; that is, if a statistic is significantly different from zero at the 0.05 level, then the 95% confidence interval will not contain zero.

## **Results**

Table 2 provides a description of the ten studies that met all inclusion criteria and were included in this synthesis. Of the ten studies, six were peer reviewed journal articles and four were unpublished doctoral dissertations. Across the ten studies, immediate posttest and follow-up data were collected on 856 and 693 adolescent struggling readers respectively. Of these, at

posttest, 263 students were identified as learning disabled while 593 were identified as struggling readers.

On average, follow-up data collection took place 21.2 weeks after posttesting (range = 2 weeks to 2 years). The analysis showed that the estimated average weighted effect size on all reading outcome measures between treatment and control groups at immediate posttest was  $g_w = 0.78$ , 95% CI [0.25, 1.31], ( $\tau^2 = .55$ ) and at follow-up was  $g_w = 0.27$ , 95% CI [-0.23, 0.77], ( $\tau^2 = .20$ ). For researcher-developed reading measures, weighted effect size between the treatment and control groups at immediate posttest was  $g_w = 0.86$ , 95% CI [0.30, 1.42], ( $\tau^2 = .59$ ), and at the follow-up time point, it was  $g_w = 0.35$ , 95% CI [-0.20, 0.91], ( $\tau^2 = .20$ ). For standardized measures, weighted effect size at immediate posttest was  $g_w = 0.05$ , 95% CI [-0.15, 0.25], ( $\tau^2 = .00$ ). No mean effect on standardized measures at follow-up was calculated because only two of the four studies that administered standardized measures collected follow-up data on control group students. Additionally, we conducted statistical significance tests for each treatment and control group comparison. Of the 44 immediate posttest effect sizes measured across ten studies, 17 were significant and positive in favor of the treatment group; 27 effect sizes were not significant. Similarly, of the 26 follow-up effect sizes measured, 13 were positive and significant; 13 effect sizes were not significant.

An effect size was also calculated to measure the magnitude of difference between each treatment group's immediate posttest and follow-up reading scores. Of the 35 immediate posttest and follow-up reading outcome comparisons, the difference in scores between treatment groups' immediate posttest and follow-up were no different from zero for 31 effect sizes as shown in Figure 2 for the studies which the confidence interval contains zero. In one study (Haines et al., 2018), treatment group students performed significantly higher on a standardized reading

measure at the two-year follow-up compared to posttest. In contrast, the data from three studies showed treatment group students performed significantly lower on certain reading measures at follow-up than at the immediate posttest (Clarke et al., 2017; Esser, 2001; Jitendra et al., 2000).

### **Study Information**

**Study participants.** Participants in nine studies included in this synthesis were sixth, seventh, eighth and/or ninth graders (Berkeley et al., 2011; Clarke et al., 2017; Esser, 2001; Haines et al., 2018; Jitendra et al., 2000; Lane, 1997; Newbern, 1998; O'Connor et al., 2019; Vachon, 1998). Participants in one study (Kennedy et al., 2015) were tenth graders. Participants in six studies were selected due to their school/district identification of LD (Berkeley et al., 2011; Esser, 2001; Jitendra et al., 2000; Kennedy et al., 2015; Newbern, 1998; O'Connor et al., 2019). Four studies included students who did not have an LD identification but were below grade-level on a standardized reading measure (Clarke et al., 2017; Haines et al., 2018; Lane, 1997; Vachon, 1998). Clarke and colleagues (2017) included students from Grades 7 and 8 who scored below 90 on the Single Word Reading Test (SWRT; Foster & National Foundation for Educational Research, 2008). Haines et al. (2018) selected participants in high poverty schools who failed to pass the state test. Lane (1997) included sixth grade students who scored between the 9th and 39th percentile on a standardized reading measure (authors did not report which standardized measure was used). Similarly, Vachon (1998) included students from Grades 6 to 8 who scored between third and fifth grade equivalencies on the Woodcock Johnson Reading Mastery Test--Word Identification subtest (Woodcock, 1987), read 60 to 90 words correct per minute on a grade-level text, and scored at or below third grade level equivalency on the Peabody Picture Vocabulary test (Dunn et al., 1965).

**Intervention type.** As shown in Table 4, five of the ten studies made use of metacognitive strategy instruction to improve reading outcomes for struggling readers. However, results of the intervention at posttest and follow-up time points varied across several factors such as type of strategy, measurement instrument, and duration of the intervention. Two studies used a multicomponent framework to provide instruction in multiple areas of reading (Clarke et al., 2017; Haines et al., 2018). Across both studies, there was no clear trend on the benefits of intervention for treatment group students. Similarly, the effects of vocabulary and word reading instruction for adolescent struggling readers also did not depict a clear trend of benefits for treatment group students at follow-up time points (Kennedy et al., 2015; O'Connor et al., 2019; Vachon, 1998).

**Comprehension.** The estimated average weighted effect size between treatment and control groups on all reading comprehension measures at immediate posttest was  $g_w = 0.67$ , 95% CI [0.10, 1.25], ( $\tau^2 = .43$ ) and at follow-up was  $g_w = 0.33$ , 95% CI [-0.19, 0.85], ( $\tau^2 = .10$ ). A majority of studies ( $n = 5$ ) included in this synthesis focused on improving students' comprehension of expository (Berkeley et al., 2011; Esser, 2001; Lane, 1997) or narrative texts (Jitendra et al., 2000; Newbern, 1998). All five studies taught students to use various comprehension strategies, however, only three studies (Berkeley et al., 2011; Esser, 2001; Jitendra et al., 2000) reported employing a combination of direct (i.e., modeling, guided, and independent practice) and strategy instruction.

As shown in Table 2, Esser (2001) and Berkeley et al. (2011) provided very similar reading interventions to adolescent struggling readers. These researchers provided a combination of direct instruction and reading comprehension strategy instruction (i.e., activating background knowledge, setting a purpose for reading, previewing text, generating questions, &



summarization) to two treatment groups. In both studies, treatment group two received additional instruction after each session in attribution retraining to improve their self-belief. Berkeley et al. (2011) found positive effects of intervention, on a researcher-developed summarization outcome measure for both treatment groups at immediate posttest ( $g = 1.39$  and  $0.92$ ) and 6-week follow-up ( $g = 1.12$  and  $0.67$ ). However, both treatment and control groups did not differ significantly at immediate posttest and follow-up time points on another researcher-developed measure—the passage test measure comprising multiple-choice and open-ended questions. Conversely, Esser (2001) administered only one researcher-developed reading measure and found positive effects of intervention at immediate posttest for both treatment groups ( $g = 0.58$  and  $1.23$ ). Nonetheless, treatment and control groups did not differ significantly on the same test at the 6-week follow-up.

In Jitendra et al. (2000), tutors provided a combination of direct instruction and main idea generation instruction to treatment group students. Of the six researcher-developed measures administered at immediate posttest and follow-up time points, treatment group students outperformed control group participants on five of six measures at immediate posttest with significant effect sizes ranging from  $g = 0.93$  to  $2.65$ . However, on the six-week follow-up test, the treatment group outperformed control group participants on only three of the six measures with significant effect sizes ranging from  $g = 0.75$  to  $1.26$ .

Studies that did not provide direct instruction also reported mixed maintenance effects. Newbern (1997) taught students the mnemonic RAP (Read-Ask-Paraphrase) to generate the main idea of the passage. On a researcher-developed measure of reading comprehension, large positive effects of intervention in favor of the treatment group were reported at immediate

posttest ( $g = 1.45$ ). However, this positive intervention effect was not maintained at the two-week follow-up time point and the  $F$ -test score was not significant.

Lane (1997) taught students to activate background knowledge, think about the most important who/what, and write a sentence describing the main idea after reading. A greater magnitude of difference at follow-up was reported compared to the immediate posttest. At posttest, the treatment group outperformed control group students on a researcher-developed main idea generation measure ( $g = 0.55$ ). However, treatment and control groups were not significantly different on another researcher-developed multiple-choice comprehension measure and the Gates-MacGinitie reading tests (GMRT; Gates & MacGinitie, 1964). At the two-week follow-up, the treatment group outperformed control group participants on all three measures: Multiple choice ( $g = 0.34$ ), main idea generation ( $g = 0.49$ ), and GMRT ( $g = 0.33$ ).

The type of instruction control group students received in all five studies varied slightly. In three studies (Esser, 2001; Lane, 1997; Newbern, 1998) control group students received no comprehension strategy instruction; students were required to read text and answer comprehension questions. In one study (Jitendra et al., 2000), control group students continued their business-as-usual activities that included decoding and comprehension activities. Finally, the control group students in Berkeley et al. (2011) practiced repeated reading, graphed their fluency scores, and made predictions before reading the text.

**Vocabulary.** Our search located two vocabulary related interventions that involved struggling readers and collected follow-up data. Kennedy et al. (2015) taught 10<sup>th</sup> grade students vocabulary words from a grade-level history lesson on World War I using multimedia-based instructional videos. Of the three different treatment groups, treatment group one watched videos containing explicit instruction incorporating text and images, treatment group two watched

videos on the usage of a mnemonic strategy, and treatment group three watched videos that combined explicit instruction with the mnemonic strategy. Control group participants were also taught the same set of vocabulary words through vocabulary videos that contained only text (in the absence of images, keyword mnemonic strategy, and direct instruction).

All three treatment groups outperformed control group students at immediate posttest (Range  $g = 1.57$  to  $2.81$ ) and at the three-week follow-up (Range  $g = 1.67$  to  $2.88$ ) on a researcher-developed, open-ended vocabulary measure that asked students to write student-friendly definitions. However, on another researcher-developed multiple-choice vocabulary measure, only the participants in treatment group three (explicit instruction + mnemonic strategy) outperformed control group students at immediate posttest ( $g = 1.57$ ). At the follow-up time point, both, treatment groups two (mnemonic strategy only) and three outperformed control group students ( $T2 = 1.41$ ;  $T3 = 1.33$ ).

In the O'Connor et al. (2019) study, researchers provided daily supplemental vocabulary lessons spanning 15 minutes. These sessions were in addition to the school provided instruction students were receiving in special education classrooms. In each session, students were taught four new vocabulary words. Lessons included word synonyms, student-friendly definitions, discussions about the words, and writing sentences with learned words. Treatment group students significantly outperformed control group students at immediate posttest on both researcher-developed measures ( $g = 1.88$  and  $g = 2.31$ ). Only treatment group students were administered the follow-up vocabulary measure. A comparison between the treatment group's immediate posttest and follow-up scores showed that participants maintained gains made during the intervention and performed similarly on the researcher-developed measure at the four-week follow-up test ( $g = 0.06$ ).

**Word reading.** The authors were unable to locate any studies, for this student population, that provided a reading fluency intervention and collected follow-up data. One study (Vachon, 1998) taught students to read multisyllabic words and collected follow-up data on their decoding and fluency outcomes. Although the study was a randomized controlled trial, it is important to note that the control group in this study did receive very similar word reading instruction. The difference between the treatment and control conditions was related to the criteria that students had to meet during instruction to receive the next set of words. The researcher compared groups of students who had to achieve 90% mastery in word reading to students who did not have to achieve mastery before new sets of words were introduced. No differences were found at immediate posttest or follow-up between the mastery and non-mastery groups on standardized measures of decoding and a researcher-developed fluency measure.

**Multicomponent reading interventions.** Two studies implemented multicomponent reading interventions and collected follow-up data for treatment group students. Clarke et al. (2017) randomized study participants to three groups. In treatment group one, students: read on- and below-grade level passages to improve reading fluency, worked on improving their decoding skills through phonics instruction, and wrote sentences. In addition to receiving instruction in reading fluency, phonics, and writing, treatment group two also received instruction in new vocabulary, listening comprehension, and strategy use. The control group received business-as-usual instruction and was wait-listed to receive treatment. At posttest, on almost all reading measures, there was no significant difference between treatment and control group participants. Due to the study design, control group students received the 20-week treatment after posttesting. Follow-up data were only available for treatment group students. Treatment participants in both

groups maintained their immediate posttest performance on all reading measures at the 20-week follow-up.

Haines et al. (2017) collected data on students who participated in the Read 180 program (Scholastic, 2015). Study participants attended daily 90-minute sessions for one academic year. The program included instruction in phonemic awareness, phonics, fluency, comprehension, vocabulary, spelling, and writing. At the end of the intervention, treatment group students were matched to students who did not receive the Read 180 program instruction. Students were matched on their baseline Scholastic Reading Inventory scores (SRI; Scholastic, 2014). Treatment and control group students did not differ significantly on the SRI measure at immediate posttest, one- and two-year follow-up tests.

**Treatment dosage.** On average, researchers provided 15.6 hours (range = 1 to 55 hours) of reading-related interventions across the nine studies; it was not possible to estimate the total hours of instruction for one study (Haines et al., 2017). Two studies collected data on both treatment and control group students at immediate posttest and follow-up time points; Newbern (1997) provided three hours, and Esser (2001) provided five hours, of comprehension instruction. Both studies reported no significant difference between treatment and control groups at a follow-up testing time point. On the other hand, Berkeley et al. (2011), Lane (1997) and Jitendra et al. (2000) provided six to ten hours of comprehension related interventions, and follow-up results varied for different measures. Berkeley et al. (2011) reported stable maintenance effects at the follow-up time point on a non-standardized measure of main idea summarization. No significant differences were observed for students in treatment and control groups at immediate posttest and follow-up time points on another non-standardized measure of explicit and implicit questions related to the test passage. Jitendra et al. (2000) reported stable

positive maintenance effects at follow-up on a researcher-developed near transfer measure of comprehension, but the magnitude of difference on a researcher-developed far transfer measure was only significant at immediate posttest and not at the follow-up time point. Conversely, Lane (1997) reported moderate positive effects of intervention on researcher-developed and standardized measures at the follow-up time point.

**Outcome measures.** Of the 25 different reading measures students were assessed on, across the ten included studies, ten were standardized norm-referenced reading measures (see Table 3). These included standardized measures of reading comprehension, reading fluency, word reading, and vocabulary. The 15 researcher-developed reading comprehension measures required students to read text and either generate a main idea statement, or answer open-ended or multiple-choice questions.

**Clarity of causal inference and study quality.** In all studies except two (Haines et al., 2018; Newbern, 1998), participants were randomly assigned to treatment or comparison conditions. Haines et al. (2018) matched treatment group students to control group students who had similar pretest scores on the Scholastic Reading Inventory. The matching was done after treatment group students had completed the intervention. Pretest data were not available to establish baseline equivalency between the two groups. Newbern (1998) selected participants based on students' LD identification at their school/district and a reading score, on a standardized reading measure, indicating the participant's reading ability was one or more years below grade level. Due to scheduling issues, 13 students were assigned to the control group. The remaining 36 students were randomly assigned to one of two treatment groups. It is equally important to note that on the pretest measure, the treatment and control groups were not comparable. According to WWC (Clearinghouse, 2017), the absolute effect size between the

treatment and control groups should be  $\leq .05$ , or between  $.05$  and  $.25$  with statistical adjustments required to satisfy baseline equivalence. In Newbern's (1998) study, the absolute effect size on the pretest measure between treatment and control groups was  $d = 0.65$ . Hence, both studies do not satisfy the baseline equivalence requirement.

For all ten studies included in this synthesis, no differential attrition was reported that exceeded the acceptable level (Clearinghouse, 2017). Group sizes remained similar at the start of the study, during posttest, and at follow-up testing time points. Based on the WWC study ratings, eight of the ten studies met WWC standards for group design studies without reservations and were rated high. Two studies (Haines et al., 2018; Newbern, 1998) did not meet the WWC group design standards and were rated low.

### **Discussion**

The objective of this synthesis was to understand how well effects of reading interventions were sustained at follow-up time points for struggling adolescent readers in Grades 6–12. Ten studies met inclusion criteria, and analyses of data showed a large significant intervention effect of reading interventions at posttest, which on average reduced to a moderate effect at follow-up. Of the 25 reading measures students were assessed on, 15 were researcher-developed reading measures.

Across all studies, the effect of treatment was  $g_w = 0.78$  at posttest and  $g_w = 0.27$  at follow-up. The estimated posttest effect size in this study ( $g_w = 0.78$ ) was high relative to past reviews of reading interventions for adolescent struggling readers, which yielded effect sizes ranging from  $g = 0.41$  to  $0.47$  (Flynn et al., 2012; Edmonds et al., 2008; Scammacca et al., 2015). One explanation for the heightened posttest effect size in this study may be the high number of researcher-developed measures in the included studies. Past studies have consistently

reported that researcher-developed measures yield greater effect sizes compared to standardized reading measures (Cheung & Slavin, 2016; Slavin & Madden, 2011).

Of the 26 effect sizes measured to compare treatment and control groups' performance at immediate posttest and follow-up, 14 were positive and significant in favor of the treatment group. It was observed that the confidence intervals of these 14 effect sizes overlapped with 12 follow-up effect size confidence intervals, denoting sustainability and stability of intervention effects. Although there is no general consensus on the appropriate time to collect follow-up data, it is relevant to note that the average follow-up timeframe was approximately 21 weeks. It is also important to note that the average follow-up time was skewed due to one study collecting follow-up data one and two years after posttest. In contrast to the mean follow-up time, the median follow-up time was six weeks. Hence, it could be argued that additional research with greater time between immediate posttest and follow-up data collection is needed to build on the current review's findings which indicate that adolescent struggling readers generally maintain their reading-related gains over time.

### **Follow-up Effects of Reading Interventions for Adolescent Struggling Readers**

Interventions targeted at improving students' reading comprehension reported that middle and high school struggling readers, in general, performed better at immediate posttest and follow-up on measures of summarizing text and identifying the main idea compared to answering multiple-choice questions. However, it should be noted that almost all of these were researcher-developed measures, and it is difficult to estimate if these tests were overaligned with learning acquisitions that favored treatment group students unfairly. Only two comprehension-related intervention studies (Haines et al., 2018; Lane, 1997) administered a standardized



measure of reading comprehension to both treatment and control groups at immediate posttest and follow-up timepoints.

Lane (1997) reported that while treatment and control group students did not differ significantly at immediate posttest on the GMRT (Gates & MacGinitie, 1964), treatment group students outperformed controls at the two-week follow-up test. This finding could imply sleeper effects of intervention indicating that students may take time to adopt strategies learned during the intervention, and positive effects may be documented if follow-up data are collected and analyzed. Haines et al. (2018) observed that treatment and control group students did not differ significantly at posttest, one- and two-year follow-ups but both groups' reading performance continued to improve over time. On the other hand, two studies (Esser, 2001; Newbern, 1998) reported moderate to large positive effects of intervention on treatment group students' reading outcomes compared to control group participants at posttest. However, treatment group students in both studies were not significantly different from control group students on reading measures at follow-up time points. In contrast to Lane's (1997) findings, results from these two studies suggest fading effects once treatment is over. These examples provide preliminary evidence of the importance of collecting follow-up data to assess students' response to intervention in a more nuanced manner.

Comprehension related intervention studies that delivered instruction for a total of <6 hours (Esser, 2001; Newbern, 1998) reported no significant differences between treatment and control groups at follow-up time points. In contrast, studies that delivered reading comprehension related interventions for 6 or more hours generally reported positive effects of intervention on reading comprehension measures at immediate posttest and follow-up time points. These findings align with current recommendations in the field that advocate for

interventions spanning longer durations to allow students who struggle with reading to make substantial gains in their targeted area of reading difficulty (Vaughn et al., 2012).

Finally, Suggate's (2016) analysis of the follow-up effects of early elementary reading intervention studies showed that providing reading interventions for at-risk student populations was beneficial in the long-term. Treatment group students not only outperformed their peers at the end of treatment but continued to show sustainable positive effects of phonological awareness and comprehension related interventions months, and sometimes years, after the intervention. In an attempt to understand the long-term effects of reading interventions for middle and high school students at-risk of reading failure, we found that not many studies follow adolescent struggling readers post-intervention. Over the last two decades, a handful of studies collected follow-up data for this student population. Our analyses showed that, in general, providing intensive reading comprehension strategy instruction, either with or without direct instruction, was beneficial to students' progress in reading. This finding indicates that when provided with targeted reading instruction in small group settings, middle and high school students who struggle to read can still make gains and improve their reading outcomes. While a majority of the studies included in this synthesis aimed to improve comprehension outcomes for struggling adolescent readers, a few studies focused on improving students' vocabulary and word reading skills. Considering the paucity of such intervention studies reporting follow-up data, it is unclear how effective word reading and vocabulary interventions are in sustainably improving students' word reading ability and lexicon post-intervention.

### **Limitations**

A key constraint of this synthesis is the limited number of studies included in this review. Although an exhaustive search process was utilized to access studies with follow-up data, only

ten studies were located that provided the data needed to measure effects. Even though a previous synthesis (Berkeley & Larsen, 2018) found other reading intervention studies with follow-up data, some of these studies did not meet our inclusion criteria due to the publication date, language of instruction, and/or lack of access to disaggregated data (see Table 1).

Additionally, hand search of relevant journals, to locate studies that fit the inclusion criteria, was limited to 2017–2019, and it is possible that additional articles may have been missed. It is also likely, we missed out on potential studies due to indexing issues highlighted in previous reviews (e.g., Lemons et al., 2016) that could lead to different search results depending on the search interface, vendor retrieval algorithms, and article indexing (Burns et al., 2019). However, it is worth noting that we followed Cooper's (2017) recommendations—an online database search, ancestry, and hand searches, in addition to contacting primary investigators for disaggregated data—to locate relevant articles.

Furthermore, due to the small number of studies included in this synthesis, it was not possible to conduct moderator analyses to analyze intervention elements that influenced the strength of association between treatment and follow-up effects. Additionally, although this study reports findings for middle and high school students' reading outcomes, the study's findings are limited to struggling readers in Grades 6–10 because no studies were found involving 11<sup>th</sup> and 12<sup>th</sup> grade participants.

Another important limitation of this synthesis is that a majority of measures administered in the included studies were researcher-developed measures. Past reviews of literature have generally reported great effects of treatment when measured on researcher-developed measures compared to distal or standardized reading measures (Cheung & Slavin, 2016; Edmonds et al., 2009; Scammacca et al., 2015). One potential implication of over reliance on making inferences

related to study effectiveness based on researcher-developed measures is the potential of inflating program effectiveness. Additionally, multiple exposures to the same researcher-developed reading measure could lead to testing effects and fatigue.

### **Implications**

Findings from the current synthesis on measuring student data at follow-up time points suggests that reading interventions can still be effective methods to improve reading outcomes for struggling readers in middle and high school. For instance, two studies included in this synthesis delivered instruction to high school students. In one vocabulary intervention study (Kennedy et al., 2015), the large positive gains students made at immediate posttest were sustained at the three-week follow-up testing time point. Similarly, one reading comprehension intervention study (Berkeley et al., 2011), reported that substantial gains made at immediate posttest on a researcher-developed comprehension measure was sustained at the six-week follow-up time point. These findings accentuate the need for intensive reading interventions for students who continue to struggle in middle and high school as this may be the final window of time within which their reading skills can be improved before they exit the school system.

In a review of a century of progress in reading interventions, Scammacca et al. (2016) noted that a majority of reading intervention studies since the year 2000 were designed to deliver multicomponent reading strategies. Two studies included in this synthesis delivered multicomponent reading interventions that targeted more than one reading component. While differences between treatment and control group students at immediate posttest were not significant on multiple reading measures, it was observed that treatment group students maintained gains made from baseline to immediate posttest almost two years after the intervention concluded.

However, more studies are needed to substantiate the claim that effects of interventions are sustained at follow-up time points. For instance, only two studies in the current synthesis implemented multicomponent reading interventions and only one study implemented vocabulary interventions. Considering the paucity of studies, generally small sample sizes, and effectiveness of programs being measured on predominantly researcher-developed measures, there is less certainty about the long-term effectiveness of reading-intervention approaches especially in the area of vocabulary and word reading development for adolescent struggling readers.

### **Conclusion**

It is important to acknowledge the challenges researchers face in collecting follow-up data. One of the biggest challenges may be the need for access to continuous resources including personnel to collect data at follow-up time points. Other challenges relate to threats to internal validity that may arise when collecting follow-up data. For instance, follow-up study design is more susceptible to high rates of attrition due to participants leaving the school district, getting home schooled, or dropping out of the school system. Another threat to the internal validity of this study design is testing effects. Students may get familiar with the testing instrument over multiple exposures and their response to tests could be mistaken for treatment effects.

However, we contend that the benefits to the field of collecting and measuring follow-up data may outweigh the inherent challenges. Studies that collect follow-up data after the completion of interventions can provide unique insights into the long-term efficacy of academic interventions. Collecting follow-up data can provide powerful evidence concerning students' response to intervention, their reading development over time, and the extent to which their reading problems persist post-intervention. Furthermore, individual reading interventions differ in terms of their intensity, duration, and instructional techniques. The long-term impact of

reading interventions that vary on these key variables also needs to be tested to improve our understanding of the components of interventions that yield long-term effects. Similar to the conclusion made by Suggate (2016), the authors hope that findings from the current study will encourage researchers to collect follow-up data for this student population to improve delivery methods that translate to sustained intervention effects.

### References

References marked with an asterisk indicate studies included in this synthesis

- Antoniou, F., & Souvignier, E. (2007). Strategy instruction in reading comprehension: An intervention study for students with learning disabilities. *Learning Disabilities: A Contemporary Journal*, 5, 41–57.
- Berkeley, S., & Larsen, A. (2018). Fostering Self-Regulation of Students with Learning Disabilities: Insights from 30 Years of Reading Comprehension Intervention Research. *Learning Disabilities Research & Practice*, 33(2), 75-86.
- \*Berkeley, S., Mastropieri, M. A., & Scruggs, T. E. (2011). Reading comprehension strategy instruction and attribution retraining for secondary students with learning and other mild disabilities. *Journal of Learning Disabilities*, 44(1), 18-32.
- Blachman, B. A., Schatschneider, C., Fletcher, J. M., Francis, D. J., Clonan, S. M., Shaywitz, B. A., & Shaywitz, S. E. (2004). Effects of intensive reading remediation for second and third graders and a 1-year follow-up. *Journal of Educational Psychology*, 96(3), 444.
- Blachman, B. A., Schatschneider, C., Fletcher, J. M., Murray, M. S., Munger, K. A., & Vaughn, M. G. (2014). Intensive reading remediation in grade 2 or 3: Are there effects a decade later?. *Journal of Educational Psychology*, 106(1), 46.
- Boardman, A. G., Roberts, G., Vaughn, S., Wexler, J., Murray, C. S., & Kosanovich, M. (2008). *Effective instruction for adolescent struggling readers: A practice brief*. Portsmouth, NH: RMC Research Corporation, Center on Instruction.
- Borkowski, J. G., Weyhing, R. S., & Carr, M. (1988). Effects of attributional retraining on strategy-based reading comprehension in learning disabled students. *Journal of Educational Psychology*, 80, 46–53.

- Boulay, Beth, Barbara Goodson, Michael Frye, Michelle Blocklin, and Cristofer Price. (2015). *Summary of research generated by Striving Readers on the effectiveness of interventions for struggling adolescent readers (NCEE 2016-4001)*. Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.
- Burns, C. S., Ii, R. M. S., Nix, T., & Huber, J. T. (2019). Search results outliers among MEDLINE platforms. *Journal of the Medical Library Association*, *107*(3). doi: 10.5195/jmla.2019.622
- Burns, M. K., Senesac, B. J., & Silbergliitt, B. (2008). Longitudinal effect of a volunteer tutoring program on reading skills of students identified as at-risk for reading failure: A two-year follow-up study. *Literacy Research and Instruction*, *47*(1), 27-37.
- Byrne, B., & Fielding-Barnsley, R. (1993). Evaluation of a program to teach phonemic awareness to young children: A 1-year follow-up. *Journal of Educational Psychology*, *85*(1), 104.
- Byrne, B., & Fielding-Barnsley, R. (1995). Evaluation of a program to teach phonemic awareness to young children: A 2-and 3-year follow-up and a new preschool trial. *Journal of Educational Psychology*, *87*(3), 488.
- Cantrell, S. C., Carter, J. C., & Rintamaa, M. (2012). *Striving Readers Cohort II Evaluation Report: Kentucky*. KY: Collaborative Center for Literacy Development, University of Kentucky.
- Cantrell, S.C., Almasi, J.F., Carter, J.C., & Rintamaa M. (2011). *Striving Readers Final Evaluation Report: Danville, Kentucky*. Collaborative Center for Literacy Development,



University of Kentucky. Presented to the Striving Readers Program, U.S. Department of Education.

Cantrell, S. C., Almasi, J. F., Carter, J. C., Rintamaa, M., & Madden, A. (2010). The impact of a strategy-based intervention on the comprehension and strategy use of struggling adolescent readers. *Journal of Educational Psychology, 102*, 257.

Chall, J. S., & Jacobs, V. A. (2003). The classic study on poor children's fourth-grade slump. *American Educator, 27*(1), 14-15.

Cheung, A. C., & Slavin, R. E. (2016). How methodological features affect effect sizes in education. *Educational Researcher, 45*(5), 283-292.

\*Clarke, P. J., Paul, S. A. S., Smith, G., Snowling, M. J., & Hulme, C. (2017). Reading intervention for poor readers at the transition to secondary school. *Scientific Studies of Reading, 21*(5), 408-427.

Clearinghouse, W. W. (2017). Standards handbook (Version 4.0).

Cooper, H. M. (2017). *Research synthesis and meta-analysis a step-by-step approach*. Los Angeles: Sage.

Cooper, J. O., Heron, T. E., & Heward, W. L. (2008). *Applied behavior analysis*. Upper Saddle River, NJ: Pearson/Merrill-Prentice Hall.

Daniel, J., & Williams, K. J. (2019). Self-questioning strategy for struggling readers: A synthesis. *Remedial and Special Education, 074193251988033*.

<https://doi.org/10.1177/0741932519880338>

Deussen, T., Scott, C., Nelsestuen, K., Roccograndi, A., & Davis, A. (2012). *Washington Striving Readers: Year 1 Evaluation Report*. Portland, OR: Education Northwest.

- Dimitrov, D., Jurich, S., Frye, M., Lammert, J., Sayko, S., & Taylor, L. (2012). *Year One Evaluation Report/Impact Study: Illinois Striving Readers*. Arlington, VA: RMC Research Corporation.
- Dunn, L. M., Dunn, L. M., Bulheller, S., & Häcker, H. (1965). *Peabody picture vocabulary test*. Circle Pines, MN: American Guidance Service.
- Edmonds, M. S., Vaughn, S., Wexler, J., Reutebuch, C., Cable, A., Tackett, K. K., & Schnakenberg, J. W. (2009). A synthesis of reading interventions and effects on reading comprehension outcomes for older struggling readers. *Review of Educational Research*, 79(1), 262-300.
- Ellis, E. S., & Graves, A. W. (1990). Teaching rural students with learning disabilities: A paraphrasing strategy to increase comprehension of main ideas. *Rural Special Education Quarterly*, 10, 2–10.
- \*Esser, M. M. S. (2001). The effects of metacognitive strategy training and attribution retraining on reading comprehension in African-American students with learning disabilities. Unpublished doctoral dissertation, University of Wisconsin-Milwaukee.
- Faddis, B. J., Beam, M., Maxim, L., Gandhi, E. V., Hahn, K., & Hale, R., (2011). *Portland Public School's Striving Readers Program: Year 5 Evaluation Report*. Prepared for the U.S. Department of Education Office of Elementary and Secondary Education.
- Feldman, J., Schenck, A., Feighan, K., Coffey, D. & Rui, N., (2011). *Memphis Striving Readers Project: Evaluation Report, Year 4*. Philadelphia: Research for Better Schools. Presented to the Striving Readers Program, U.S. Department of Education.
- Fisher, Z., & Tipton, E. (2015). Robumeta: An R-package for robust variance estimation in meta-analysis. arXiv preprint arXiv:1503.02220.

- Flynn, L. J., Zheng, X., & Swanson, H. L. (2012). Instructing struggling older readers: a selective meta-analysis of intervention research. *Learning Disabilities Research & Practice, 27*(1), 21-32.
- Foster, H. & the National Foundation for Educational Research. (2008). *Single Word Reading Test 6–16*. London, UK:GL Assessment.
- Gates, A. I., & MacGinitie, W. H. (1964). *Gates-MacGinitie reading tests*. Teachers College Press, Columbia University.
- Graves, A. W. (1986). Effects of direct instruction and metacomprehension training on finding main ideas. *Learning Disabilities Research & Practice, 1*, 90–100.
- \* Haines, M. L., Husk, K. L., Baca, L., Wilcox, B., & Morrison, T. G. (2018). Longitudinal Effects of Reading Intervention on Below Grade Level, Preadolescent, Title I Students. *Reading Psychology, 39*(7), 690–710. doi: 10.1080/02702711.2018.1515135
- Hall, C., Roberts, G. J., Cho, E., McCulley, L. V., Carroll, M., & Vaughn, S. (2017). Reading instruction for English learners in the middle grades: A meta-analysis. *Educational Psychology Review, 29*(4), 763-794.
- Hedges, L. V., Tipton, E., & Johnson, M. C. (2010). Robust variance estimation in meta-regression with dependent effect size estimates. *Research Synthesis Methods, 1*(1), 39-65.
- Hock, M. F., Brasseur-Hock, I. F., Hock, A. J., & Duvel, B. (2017). The effects of a comprehensive reading program on reading outcomes for middle school students with disabilities. *Journal of Learning Disabilities, 50*(2), 195-212.
- Hofstetter, C.H., Strick, B., Der-Martirosian, C., Ong-Dean, C., Long, P., & Lou, Y. (2011). *Implementation and Impact of the Targeted and Whole School Interventions, Summary of Year 4 (2009-2010)*. San Diego Unified School District. San Diego, CA. Report

submitted to the U.S. Department of Education, Office of Elementary and Secondary Education.

\*Jitendra, A. K., Kay Hoppes, M., & Xin, Y. P. (2000). Enhancing main idea comprehension for students with learning problems: The role of a summarization strategy and self-monitoring instruction. *The Journal of Special Education, 34*(3), 127-139.

Johnson, L., Graham, S., & Harris, K. R. (1997). The Effects of Goal Setting and Self-Instruction on Learning a Reading Comprehension Strategy. *Journal of Learning Disabilities, 30*(1), 80–91. doi: 10.1177/002221949703000107

\*Kennedy, M. J., Deshler, D. D., & Lloyd, J. W. (2015). Effects of multimedia vocabulary instruction on adolescents with learning disabilities. *Journal of Learning Disabilities, 48*(1), 22-38.

Keogh, B. (2004). The importance of longitudinal research for early intervention practices. In P. McCardle & V. Chhabra (Eds.), *The voice of evidence in reading research*, (pp. 81-102). Baltimore, MD: Brookes Publishing Co.

\*Lane, S. S. (1997). Motivating inactive readers: the effects of video-modeled coping training on the metacognitive acquisition of a reading strategy by low-achieving readers. Unpublished doctoral dissertation, University of California, Los Angeles.

Lemons, C. J., King, S. A., Davidson, K. A., Berryessa, T. L., Gajjar, S. A., & Sacks, L. H. (2016). An inadvertent concurrent replication: Same roadmap, different journey. *Remedial and Special Education, 37*(4), 213-222.

Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis*. Sage Publications, Inc.

Loadman, W.E., Moore, R.J., Ren, W., Zhu, J., Zhao, J., & Lomax, R. (2011). *Striving Readers Year 5 Project Evaluation Report: Ohio. An Addendum to the Year 4 Report*.

Meisch, A., Hamilton, J., Chen, E., Quintanilla, P., Fong, P., Gray-Adams, K., et al. (2011).

*Striving Readers Study: Targeted and Whole School Interventions - Year 5.*

Miranda, A., Villaescusa, M. I., & Vidal-Abarca, E. (1997). Is Attribution Retraining Necessary?

Use of Self-Regulation Procedures for Enhancing the Reading Comprehension Strategies of Children with Learning Disabilities. *Journal of Learning Disabilities, 30*(5), 503–512.

doi: 10.1177/002221949703000506

\*Newbern, S. L. (1998). The effects of instructional settings on the efficacy of strategy

instruction for students with learning disabilities. Unpublished doctoral dissertation,

Johns Hopkins University.

Newman, D. L., Kundert, D. K., Spaulding, D. T., White, S. P., & Gifford, T. A., (2012).

*Striving Readers Project: New York State Department of Education/New York City*

*Department of Education: Striving Readers Local Evaluation.*

\*O'Connor, R. E., Beach, K. D., Sanchez, V. M., Kim, J. J., Knight-Teague, K., Orozco, G., &

Jones, B. T. (2019). Teaching Academic Vocabulary to Sixth-Grade Students With

Disabilities. *Learning Disability Quarterly, 073194871882109*. doi:

10.1177/0731948718821091

Roberts, G., Torgesen, J. K., Boardman, A., & Scammacca, N. (2008). Evidence-based strategies

for reading instruction of older students with learning disabilities. *Learning Disabilities*

*Research & Practice, 23*(2), 63-69.

Ryder, J. F., Tunmer, W. E., & Greaney, K. T. (2008). Explicit instruction in phonemic

awareness and phonemically based decoding skills as an intervention strategy for

struggling readers in whole language classrooms. *Reading and Writing, 21*(4), 349-369.

- Scammacca, N. K., Roberts, G. J., Cho, E., Williams, K. J., Roberts, G., Vaughn, S. R., & Carroll, M. (2016). A century of progress: Reading interventions for students in grades 4–12, 1914–2014. *Review of Educational Research, 86*(3), 756-800.
- Scammacca, N., Roberts, G., Vaughn, S., Edmonds, M., Wexler, J., Reutebuch, C. K., & Torgesen, J. K. (2007). *Interventions for adolescent struggling readers: A meta-analysis with implications for practice*. Portsmouth, NH: RMC Research Corporation, Center on Instruction.
- 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*(4), 369-390.
- Schenck, A., Jurich, S., Frye, M., Lammert, J., & Sayko, S. (2012). *Evaluation Report/Impact Study: Virginia Striving Readers Initiative (VSRII)*. Arlington, VA: RMC Research Corporation.
- Schiller, E., Wei, X., Thayer, S., Blackorby, J., Javitz, H., & Williamson, C. (2012). *A Randomized Control Trial of the Impact of the Fusion Reading Intervention on Reading Achievement and Motivation for Adolescent Struggling Readers*. Submitted for publication review by SRI International: Arlington, VA. June 2012.
- Scholastic. (2014). *SRI college & career technical guide*. New York, NY: Author.
- Scholastic. (2015). *READ 180 Reading intervention program: A comprehensive reading intervention solution*.
- Slavin, R., & Madden, N. A. (2011). Measures inherent to treatments in program effectiveness reviews. *Journal of Research on Educational Effectiveness, 4*(4), 370-380.

- 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*(1), 77-96.
- Swanlund, A., Dahlke, K., Tucker, N., Kleidon, B., Kregor, J., Davidson-Gibbs, D., & Hallberg, K. (2012). *Striving Readers: Impact Study and Project Evaluation Report: Wisconsin Department of Public Instruction (with Milwaukee Public Schools)*. Naperville, IL: American Institutes for Research.
- The Education Alliance at Brown University (2012). *Springfield-Chicopee School Districts Striving Readers (SR) Program: Final Report Years 1-5: Evaluation of Implementation and Impact*. Prepared for: Office of Elementary and Secondary Education, U.S. Department of Education Institute of Education Sciences, U.S. Department of Education.
- Tunik, J., Simon, A., Alemany, J., Zhu, J., Zacharia, J., Ramsey, L., et al. (2011). *Chicago Public Schools Striving Readers Initiative: Year Four Evaluation Report*. Submitted to Elizabeth Cadenas-Lopez, Project Director, Striving Readers, U.S. Department of Education.
- \*Vachon, V. L. (1999). Effects of mastery of multisyllabic word reading component skills and of varying practice contexts on word and text reading skills of middle school students with reading deficiencies. Unpublished doctoral dissertation, University of Oregon.
- Vadasy, P. F., & Sanders, E. A. (2013). Two-year follow-up of a code-oriented intervention for lower-skilled first-graders: The influence of language status and word reading skills on third-grade literacy outcomes. *Reading and Writing, 26*(6), 821-843.
- Vadasy, P. F., Sanders, E. A., & Peyton, J. A. (2006). Code-oriented instruction for kindergarten students at risk for reading difficulties: A randomized field trial with paraeducator implementers. *Journal of Educational Psychology, 98*(3), 508–528.

- Vaden-Kiernan, M., Caverly, S., Bell, N., Sullivan, K., Fong, C. & Atwood, E. (2012). *Louisiana Striving Readers: Final Report*. Austin TX: Southwest Educational Development Laboratory.
- Vaughn, S., Elbaum, B. E., Wanzek, J., Scammacca, N., & Walker, M. A. (2014). Code sheet and guide for education-related intervention study syntheses. Austin, TX: The Meadows Center for Preventing Educational Risk.
- Vaughn, S., Martinez, L. R., Williams, K. J., Miciak, J., Fall, A. M., & Roberts, G. (2019). Efficacy of a high school extensive reading intervention for English learners with reading difficulties. *Journal of Educational Psychology, 111*(3), 373.
- Vaughn, S., Roberts, G., Swanson, E. A., Wanzek, J., Fall, A. M., & Stillman-Spisak, S. J. (2015). Improving middle-school students' knowledge and comprehension in social studies: A replication. *Educational Psychology Review, 27*(1), 31-50.
- Vaughn, S., Roberts, G., Wexler, J., Vaughn, M. G., Fall, A. M., & Schnakenberg, J. B. (2015). High school students with reading comprehension difficulties: Results of a randomized control trial of a two-year reading intervention. *Journal of Learning Disabilities, 48*(5), 546-558.
- Vaughn, S., Wanzek, J., Murray, C. S., & Roberts, G. (2012). *Intensive Interventions for Students Struggling in Reading and Mathematics. A Practice Guide*. Center on Instruction.
- Wanzek, J., Vaughn, S., Scammacca, N. K., Metz, K., Murray, C. S., Roberts, G., & Danielson, L. (2013). Extensive reading interventions for students with reading difficulties after grade 3. *Review of Educational Research, 83*(2), 163-195.



Woodcock, R. W. (1987). *Woodcock reading mastery tests-revised*. Circle Pines, MN: American Guidance Service.

Table 1

*Examples of Studies Not Included in this Synthesis.*

Study	Reason for Exclusion
Antoniou and Souvignier (2007)	Intervention targeted German language reading related skills. Current synthesis only included studies that targeted English language related skills.
Borkowski et al., (1988)	Study publication year did not meet our search time frame that included studies published on or after 1996.
Burns, Senesac, and Silbergliitt (2008)	Study reports two-year follow-up data for sixth grade students. However, students were in fourth grade when intervention concluded.
Ellis and Grave (1990)	Study publication year did not meet our search time frame that included studies published on or after 1996.
Graves (1986)	Study publication year did not meet our search time frame that included studies published on or after 1996.
Hock, Brasseur-Hock, Hock, and Duvel (2017)	Treatment group received two years of intervention. Control group was waitlisted for year 1 and received intervention in year 2. No follow-up data were collected after the intervention ended in year 2.
Johnson et al., (1997)	The study included students in Grades 4 – 6. Johnson et al. were unable to confirm the majority of students were in Grade 6 or provide disaggregated data for 6 <sup>th</sup> grade students.
Miranda et al. (1997)	Intervention targeted Spanish language reading related skills. Current synthesis only included studies that targeted English language related skills.
Vaughn et al. (2015)	Study provides whole class instruction to a heterogenous population of readers. No small-group instruction provided.

Table 2  
*Study Information*

Study	Design Information	Treatment ( <i>n</i> = sample size)	Comparison ( <i>n</i> =sample size)
Berkeley et al., 2011 <ul style="list-style-type: none"> <li>▪ Peer-reviewed article</li> <li>▪ Learning disability</li> <li>▪ 7<sup>th</sup>, 8<sup>th</sup> and 9<sup>th</sup> graders</li> <li>▪ 3.7 to 4.2 years below grade-level</li> </ul>	<ul style="list-style-type: none"> <li>▪ Experimental (Treatment/Comparison)</li> <li>▪ 6 hours</li> <li>▪ NR / 12</li> </ul>	T1: Direct instruction in using comprehension strategies ( <i>n</i> = 20).  T2: Direct instruction in using comprehension strategies plus attribution retraining to improve student self-belief ( <i>n</i> = 19).	CO: Students read text and made predictions, practiced repeated reading, answered comprehension questions and graphed their fluency scores ( <i>n</i> = 20).
Clarke et al., 2017 <ul style="list-style-type: none"> <li>▪ Peer-reviewed article</li> <li>▪ Struggling readers</li> <li>▪ 7<sup>th</sup> and 8<sup>th</sup> graders</li> <li>▪ &lt;90 standard score on standardized reading measure</li> </ul>	<ul style="list-style-type: none"> <li>▪ Experimental (Treatment/Comparison)</li> <li>▪ 35 hours</li> <li>▪ 3 times a week / 60</li> </ul>	T1: Instruction in reading fluency, phonics, and writing ( <i>n</i> = 95).  T2: Instruction in reading fluency, phonics, writing, vocabulary, listening comprehension, and strategy instruction (graphic organizer) ( <i>n</i> = 94).	CO: Control group students were wait-listed to receive intervention. While wait-listed, students were received business-as-usual instruction. ( <i>n</i> = 89).
Esser, 2001 <ul style="list-style-type: none"> <li>▪ Unpublished doctoral dissertation</li> <li>▪ Learning disability</li> <li>▪ 6<sup>th</sup> and 7<sup>th</sup> graders</li> <li>▪ NR</li> </ul>	<ul style="list-style-type: none"> <li>▪ Experimental (Treatment/Comparison)</li> <li>▪ 5 hours</li> <li>▪ Twice a week / 6</li> </ul>	T1: Direct instruction in using comprehension strategies ( <i>n</i> = 20) T2: A combination of direct instruction in comprehension strategies and attribution retraining ( <i>n</i> = 20).	CO: Students read text and answered comprehension questions ( <i>n</i> = 20).

<p>Haines et al., 2018</p> <ul style="list-style-type: none"> <li>▪ Peer-reviewed article</li> <li>▪ At-risk students</li> <li>▪ 6<sup>th</sup> graders</li> <li>▪ Failed the AIMS test</li> </ul>	<ul style="list-style-type: none"> <li>▪ Experimental (Treatment/Comparison)</li> <li>▪ NR</li> <li>▪ Five times a week / NR</li> </ul>	<p>T1: Implemented the Read 180 program (<math>n = 7</math>)</p>	<p>CO: Business-as-usual instruction (<math>n = 7</math>).</p>
<p>Jitendra et al., 2000</p> <ul style="list-style-type: none"> <li>▪ Peer-reviewed article</li> <li>▪ Learning disability</li> <li>▪ 6<sup>th</sup>, 7<sup>th</sup> and 8<sup>th</sup> graders</li> <li>▪ Two years below grade-level</li> </ul>	<ul style="list-style-type: none"> <li>▪ Experimental (Treatment/Comparison)</li> <li>▪ 7.5 to 10 hours</li> <li>▪ NR / 15</li> </ul>	<p>T1: Direct instruction in generating main idea after reading the text and self-monitor during reading using cue cards (<math>n = 18</math>)</p>	<p>CO: Systematic reading instruction that emphasized decoding and comprehension activities (<math>n = 15</math>).</p>
<p>Kennedy et al., 2015</p> <ul style="list-style-type: none"> <li>▪ Peer-reviewed article</li> <li>▪ Learning disability</li> <li>▪ 10<sup>th</sup> graders</li> <li>▪ NR</li> </ul>	<ul style="list-style-type: none"> <li>▪ Experimental (Treatment/Comparison)</li> <li>▪ 1 hour</li> <li>▪ Three times a week / 3</li> </ul>	<p>Multimedia-based instruction on vocabulary words,  T1: that contained explicit instruction with text and images (<math>n = 7</math>)  T2: that contained keyword mnemonic strategy (<math>n = 8</math>)  T3: that contained explicit instruction and keyword mnemonic strategy (<math>n = 7</math>)</p>	<p>CO: Multimedia-based instruction on vocabulary words that contained text and no images. (<math>n = 8</math>)</p>
<p>Lane, 1997</p> <ul style="list-style-type: none"> <li>▪ Unpublished doctoral dissertation</li> <li>▪ Struggling Readers</li> <li>▪ 6<sup>th</sup> graders</li> <li>▪ 2 to 3 years below-grade level</li> </ul>	<ul style="list-style-type: none"> <li>▪ Experimental (Treatment/Comparison)</li> <li>▪ 8.3 hours</li> <li>▪ Five times a week / 10</li> </ul>	<p>T1 and T2: Students were taught reading comprehension strategy and how to cope with failure or respond positively to teacher instruction (<math>n = 138</math>)</p>	<p>CO: Business-as-usual instruction (<math>n = 98</math>).</p>

<p>Newbern, 1998</p> <ul style="list-style-type: none"> <li>▪ Unpublished doctoral dissertation</li> <li>▪ Learning disability</li> <li>▪ 7<sup>th</sup> and 8<sup>th</sup> graders</li> <li>▪ 1 to 2 years below grade-level</li> </ul>	<ul style="list-style-type: none"> <li>▪ Experimental (Treatment/Comparison)</li> <li>▪ 3 hours</li> <li>▪ Once a week / 6</li> </ul>	<p>T1: Instruction in using the RAP (Read-Ask-Paraphrase) strategy in a small-group setting (<math>n = 16</math>)</p>	<p>CO: No strategy related instruction was provided (<math>n = 13</math>).</p>
<p>O'Connor et al., 2019</p> <ul style="list-style-type: none"> <li>▪ Peer-reviewed article</li> <li>▪ Learning disability</li> <li>▪ 6<sup>th</sup> graders</li> <li>▪ 2.5 years below grade-level</li> </ul>	<ul style="list-style-type: none"> <li>▪ Experimental (Treatment/Comparison)</li> <li>▪ 55 hours</li> <li>▪ Five times a week / 60</li> </ul>	<p>T1: Direct instruction in phonics and vocabulary (<math>n = 32</math>).</p>	<p>CO: Direct instruction in phonics and reading text fluently (<math>n = 20</math>).</p>
<p>Vachon, 1998</p> <ul style="list-style-type: none"> <li>▪ Unpublished doctoral dissertation</li> <li>▪ Struggling Readers</li> <li>▪ 6<sup>th</sup>, 7<sup>th</sup> and 8<sup>th</sup> graders</li> <li>▪ 1 to 3 years below grade-level</li> </ul>	<ul style="list-style-type: none"> <li>▪ Experimental (Treatment/Comparison)</li> <li>▪ 17.3 hours</li> <li>▪ Five times a week / 25</li> </ul>	<p>All groups received instruction in multisyllabic word reading.</p> <p>T1 and T2: When students achieved 90% mastery, they moved to next set of words. They also read grade-level passages or sentences (<math>n = 32</math>).</p> <p>T3 and T4: Students did not have to achieve mastery to move to next lesson. They also read grade-level passages or sentences (<math>n = 33</math>).</p>	<p>CO: There was no control group</p>

Note: T = Treatment; CO = Control. AIMS = Arizona's Instrument to Measure Standards test

Table 3  
*Study measures and outcomes.*

Study	Intervention Type	Dependent Measure(s)	Std	Group	PT Sample Size	PT		FU Sample Size	FU		Week
						<i>g</i>	<i>95% CI</i>		<i>g</i>	<i>95% CI</i>	
Berkeley et al., 2011	Comprehension	Summary Test	N	T1 – CO	59	1.39	0.70 to 2.08	59	1.12	0.45 to 1.79	6
				T2 – CO		0.92	0.26 to 1.58		0.67	0.03 to 1.32	
	Passage Test	N	T1 – CO	0.13	-0.49 to 0.75	0.25	-0.37 to .88				
			T2 – CO	-0.16	-0.78 to 0.47	0.05	-0.58 to 0.68				
Clarke et al., 2017	Multi-component	NGRT	Y	T1 – CO	278	0.21	-0.21 to 0.54	145	*Due to the wait list control group study design, follow-up data were only available for treatment group students.	20	
				T2 – CO		0.45	0.12 to 0.74				
		TOWRE-Sight Word	Y	T1 – CO	0.01	-0.34 to 0.32					
				T2 – CO	0.06	-0.26 to 0.39					
		TOWRE-Phonemic Decoding	Y	T1 – CO	0.11	-0.22 to 0.44					
				T2 – CO	0.22	-0.11 to 0.53					
		SWRT	Y	T1 – CO	0.17	-0.15 to 0.50					
				T2 – CO	0.03	-0.30 to 0.36					
		WIAT III RC	Y	T1 – CO	-0.23	-0.56 to 0.10					
				T2 – CO	0.05	-0.29 to 0.37					
		WASI II Vocab	Y	T1 – CO	-0.03	-0.37 to 0.29					
				T2 – CO	0.08	-0.24 to 0.48					
		Taught Words	N	T1 – CO	-0.12	-0.46 to 0.20					
				T2 – CO	0.24	-0.09 to 0.57					
Nontaught Words	N	T1 – CO	-0.04	-0.37 to 0.28							
		T2 – CO	0.18	-0.15 to 0.51							

Esser, 2001	Comprehension	Comprehension quiz	N	T1 – CO T2 – CO	60	0.58 1.23	-0.05 to 1.21 0.56 to 1.91	60	-0.19 -0.22	-0.82 to 0.43 -0.84 to 0.41	6
Haines et al., 2018	Multi-component	Scholastic reading inventory	Y	T1 – CO	14	-0.15	-1.20 to 0.90	14	-0.56 -0.99	-1.63 to 0.51 -2.03 to 0.18	52 104
Jitendra et al., 2000	Comprehension	Main idea generation - Training <sup>1</sup> - Near Transfer <sup>1</sup> - Far Transfer <sup>1</sup>	N	T1 – CO	33	2.65 1.22 2.07	1.71 to 3.58 0.47 to 1.96 1.22 to 2.91	33	1.26 1.02 0.58	0.51 to 2.01 0.30 to 1.75 -0.12 to 1.28	6
		Multiple Choice - Training - Near Transfer - Far Transfer	N	T1 – CO		0.93 1.44 0.63	0.21 to 1.65 0.67 to 2.21 -0.07 to 1.33		0.02 0.75 0.12	-0.67 to 0.71 0.04 to 1.45 -0.57 to 0.80	
Kennedy et al., 2015	Vocabulary	Multiple Choice	N	T1 – CO T2 – CO T3 – CO	30	0.96 0.86 1.57	-0.11 to 2.03 -0.16 to 1.89 0.41 to 2.73	30	0.73 1.41 2.33	-0.32 to 1.78 0.31 to 2.50 1.01 to 3.64	3
		Open Ended	N	T1 – CO T2 – CO T3 – CO		1.74 1.57 2.81	0.55 to 2.93 0.45 to 2.70 1.38 to 4.23		1.80 1.67 2.88	0.60 to 3.01 0.53 to 2.81 1.43 to 4.32	
Lane, 1997	Comprehension	Multiple Choice Main idea generation Gates MacGinitie	N N Y	T1 & T2 – CO T1 & T2 – CO T1 & T2 – CO	236	0.14 0.55 0.21	-0.11 to 0.40 0.28 to 0.80 -0.04 to 0.47	226	0.34 0.49 0.33	0.07 to 0.60 0.22 to 0.76 0.06 to 0.59	2
Newbern, 1998	Comprehension	Reading comprehension test	N	T1 – CO	29	1.45	0.63 to 2.27	NR	NR <sup>2</sup>	NR <sup>2</sup>	2

O’Conno r et al., 2019	Vocabulary	Vocabulary Use	N	T1 – CO	52	1.88	1.21 to 2.54	32	*Since control group students did not learn vocabulary words, follow-up tests were only administered to treatment group.	4	
		Multiple Choice – Vocabulary	N	T1 – CO		2.31	1.60 to 3.02				
Vachon, 1998	Phonics	WRMT – Word identification	Y	T1 & T2 – T3 and T4	65	-0.02	-0.51 to 0.47	65	0.00	-0.49 to 0.49	7
		WRMT – Word attack Passage reading error	Y	T1 & T2 – T3 and T4		0.00	-0.49 to 0.49		0.16	-0.33 to 0.65	
			N	T1 & T2 – T3 and T4		-0.05	-0.54 to 0.43		-0.22	-0.71 to 0.27	

Note:

Std = Standardized measure. T = Treatment; CO = Control.

PT = Posttest. FU = Follow-up. NR = Not reported.

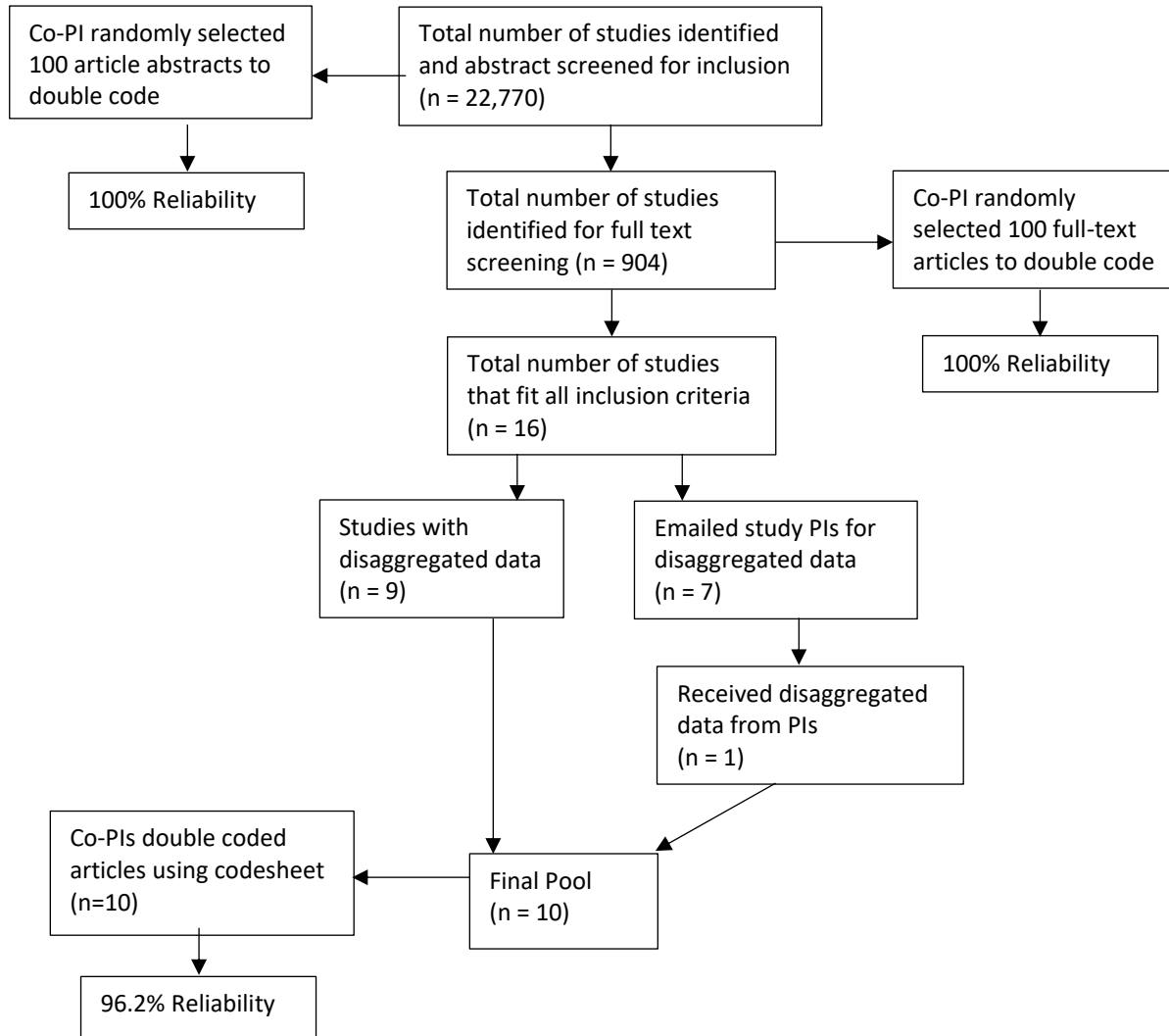
SWRT = Single Word Reading Test; NGRT = New Group Reading Test Digital; TOWRE II = Test of Word Reading Efficiency–2; WIAT II = Wechsler Individual Achievement Test–2nd Edition; WASI II = Wechsler Abbreviated Scale of Intelligence; WRMT = Woodcock Johnson Reading Mastery Test.

1. Training test items were similar to items students were trained on; Near Transfer measure items were based on similar narrative text that was not used in training; Far Transfer items were based on expository passages from social studies text.

2. Author reported that F-test was not significant for follow-up test; however, F-score was not reported.

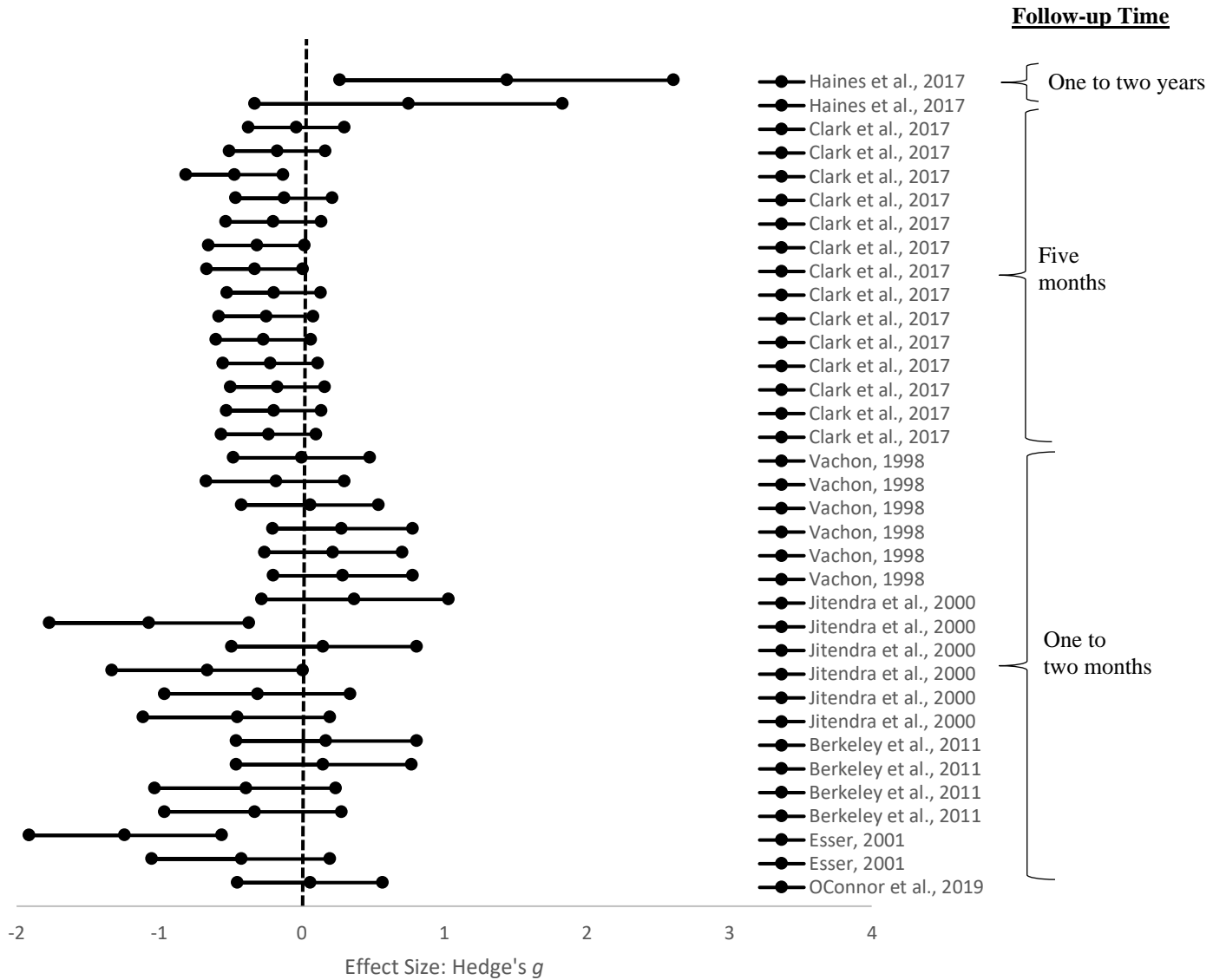


Figure 1. Flowchart for inclusion of studies.



\*PI = Primary investigator

Figure 2. Reading outcome comparison for treatment groups at follow-up and immediate posttest time points.



Note:

1. Effect size is significant when 95% confidence interval does not contain zero.
2. Positive effects (does not contain zero) indicate significantly greater performance at follow-up compared to immediate posttest.
3. Negative effects (does not contain zero) indicate significantly lower performance at follow-up compared to immediate posttest.
4. Kennedy et al, (2015) was not included because at follow-up researchers administered a truncated version of the posttest measure
5. Lane (1997) and Newbern (1998) were not included because raw data or summary statistics for follow-up time point are not reported; authors only report F-statistic for difference between treatment and control groups.