Read Writ (2017) 30:1639–1665
DOI 10.1007/s11145-017-9743-3

The impact of tier 1 reading instruction on reading outcomes for students in Grades 4–12: A meta-analysis

Elizabeth Swanson[1] · Elizabeth A. Stevens[1] · Nancy K. Scammacca[1] · Philip Capin[1] · Alicia A. Stewart[1] · Christy R. Austin[1]

Published online: 29 April 2017

Abstract Understanding the efficacy of evidence-based reading practices delivered in the Tier 1 (i.e. general classroom) setting is critical to successful implementation of multi-tiered systems, meeting a diverse range of student learning needs, and providing high quality reading instruction across content areas. This meta-analysis presents evidence on the effects of Tier 1 reading instruction on the reading outcomes of students in Grades 4–12, and a synthesis of effects for students identified as struggling readers. Results from this meta-analysis of 37 publications conducted between 2000 and 2015 reveal significant, positive effects for Tier 1 reading instruction on comprehension and vocabulary outcomes. A synthesis of the results for struggling readers indicates that they maintained or improved reading comprehension over struggling readers receiving typical instruction.

Keywords Tier 1 · Response to intervention · Reading instruction · Meta-analysis

Introduction

In the wake of national legislation (i.e., No Child Left Behind Act, Individuals with Disabilities Education Act), many school systems are implementing multi-tiered instructional models (e.g., Response to Intervention) to implement research-based practices and meet the needs of diverse learners. Frameworks such as these aim to improve student academic and behavioral outcomes by providing students with the appropriate level of classroom support (Fletcher & Vaughn, 2009; Vaughn & Fuchs, 2003). The success of a multi-tiered framework begins with establishing school-wide, high-quality general classroom instruction via professional development in

✉ Elizabeth Swanson
easwanson@austin.utexas.edu

[1] The University of Texas at Austin, Austin, TX, USA

evidence-based instructional procedures and classroom support from instructional leaders (i.e., Tier 1; Fletcher & Vaughn, 2009).

Although classroom teachers may implement scientifically validated techniques in Tier 1, it is unlikely that one instructional approach or program will meet the needs of all students (Fuchs & Deshler, 2007). Therefore, Tier 1 should consist of instruction that is a "'good bet'" for most students (Fuchs & Deshler, 2007, p. 132). Even with effective Tier 1 instruction, it is likely that only 80% of students will respond, leaving 20% that require Tier 2 or Tier 3 intervention (Vaughn & Fletcher, 2012). When Tier 1 instruction is successful and meets the needs of a higher percentage of students, fewer require services at the Tier 2 or Tier 3 level. In this way, it is critically important that Tier 1 instruction is as efficacious as possible.

Identifying Tier 1 reading instruction that benefits most students is critical to the successful implementation of multi-tiered systems and meeting a diverse range of student learning needs. In order to provide teachers with targeted, ongoing professional development, effective Tier 1 instructional practices must be identified. Prior reviews provide information on the effects of specific components of Tier 1 practices (e.g., cooperative leaning, vocabulary instruction, and phonemic awareness; Cisco & Padron, 2012; Ehri et al., 2001; Faggella-Luby, Drew, & Schumaker, 2015; Puzio & Colby, 2013; Reznitskaya et al., 2009;); however, the broader corpus of studies examining Tier 1 reading instruction for students in Grades 4–12 has not yet been synthesized.

In both the Common Core State Standards and progressive state standards, teachers are expected to infuse content area instruction with literacy practices. Therefore, we include Tier 1 reading instruction (e.g., word reading, reading fluency, vocabulary, reading comprehension, multicomponent) delivered in both English language arts/reading classes as well as the content areas. Our review also uniquely extends prior research by defining the population more broadly, including all students taught in the general education setting (i.e., typically achieving students and students with reading difficulties).

We also used evidence from prior research to choose several moderators that may impact student outcomes. For example, in one meta-analysis of reading instruction delivered using social studies materials, effect sizes did not differ based on duration of intervention (Swanson et al., 2012). This finding was reported in a meta-analysis of Tier 3 interventions as well (Wanzek et al., 2013). In our meta-analysis as well, we do not expect duration to impact effect sizes. Another moderator of interest was grade level. According to Bloom, Hill, Black, and Lipsey, (2008), annual effects on key measures of reading comprehension are smaller among older students (e.g., $ES = 0.06$ in 12th grade) than among younger students (e.g., $ES = 0.36$ in 4th grade). Therefore, we hypothesize effect sizes to be larger in the lower grades (i.e., 4th–5th) than in higher grades (i.e., 9th–12th).

The purpose of this meta-analysis is to analyze the effects of Tier 1 instruction for students in Grades 4–12 from 2000 to 2015. We address the following research questions: (a) What are the effects of Tier 1 reading instruction on the reading outcomes (i.e., reading, vocabulary, oral reading fluency, reading comprehension, phonics or word reading) of students in Grades 4–12?; (b) What variables (e.g., intervention type, hours of treatment, grade level, research design) moderate the

effect of Tier 1 instruction on reading outcomes for this population? We also conducted a narrative synthesis of the studies that disaggregated data for struggling readers and addressed the question: What are the effects of Tier 1 reading instruction on the reading outcomes of struggling readers in Grades 4–12?

Method

We conducted a comprehensive search of the literature using a three-step process. First, we conducted an electronic search of the ERIC, PsycINFO, and Academic Search Complete databases to identify peer-reviewed studies published from 2000 to 2015. We selected this range of years to reflect the most current research on this topic. Key search terms and roots related to Tier I (Tier 1 or classroom instruction or full class or whole class or general education or regular education or interven*) coupled with reading search terms and roots (read* or vocabulary or oral reading fluency or comprehen*) were used to capture the highest number of potentially relevant articles. Second, a backwards search (Cooper, Hedges & Valentine, 2009) was used to identify relevant studies referenced in prior related syntheses (Cheung & Slavin, 2012; Cisco & Padrón, 2012; Ehri et al., 2001; Faggella-Luby et al., 2015; Puzio & Colby, 2013; Reichrath, de Witte, & Winkens, 2010; Reznitskaya et al., 2009; USDOE, IES, & WWC, 2009, 2013). Last, a hand search was conducted of the three journals that commonly report reading intervention studies (*Reading Research Quarterly, Journal of Research in Educational Effectiveness, Reading and Writing*).

Figure 1 provides an overview of the search and screening process. The database, backwards, and hand searches yielded 4325 studies. We screened all abstracts and then evaluated the full texts of those records that met the initial screening for eligibility. Included studies met the following a priori inclusion criteria:

- A majority of the sample participants were students in Grades 4 through 12 or aged 9 to 18 or data was disaggregated by grade level.
- The reading instruction was provided in an alphabetic language and delivered in a general education classroom.
- The dependent variable addressed reading performance outcome(s) (i.e., vocabulary, oral reading fluency, comprehension, phonics/word study).
- The research design was experimental, quasi-experimental, or multiple treatment.
- The study was published in English in a peer-reviewed journal from 2000 through 2015.
- The study provided sufficient data for computing a standardized mean difference effect size.

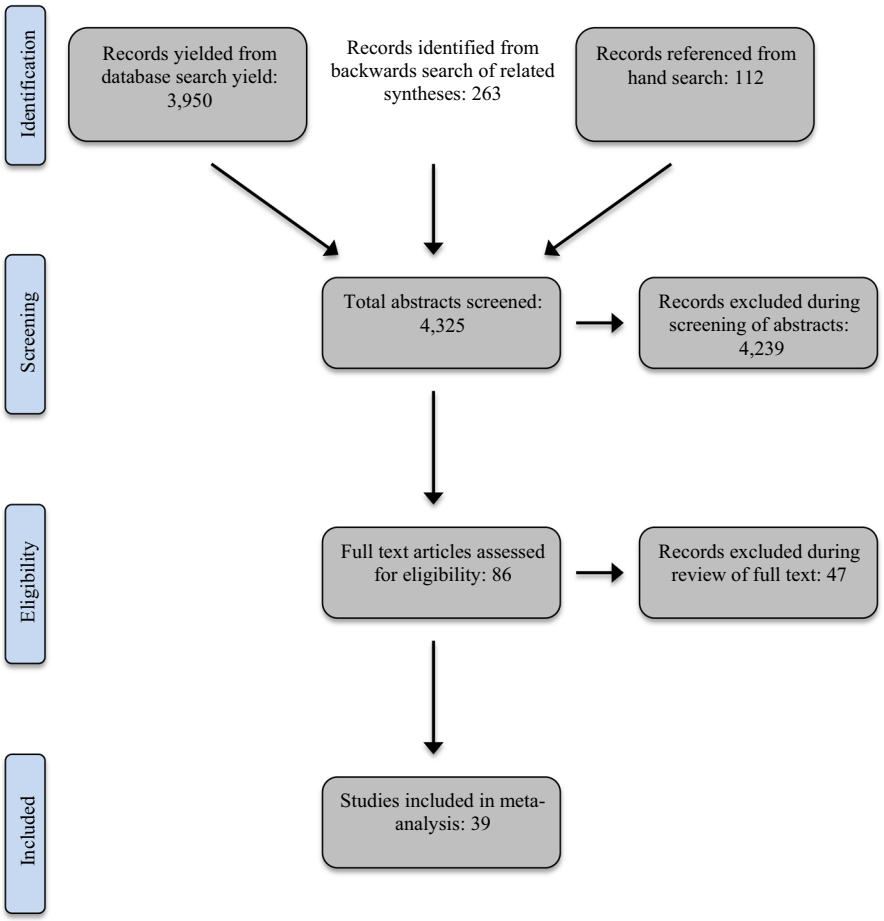

Fig. 1 Manuscript search and screening flow chart. Articles were excluded during the screening and eligibility phases for not meeting any of the following criterion: (1) A majority of the sample participants were students in Grades 4 through 12 or aged 9–18 or data was disaggregated by grade level; (2) The reading instruction was provided in an alphabetic language and delivered in a general education classroom; (3) The dependent variable addressed reading performance outcome(s) (i.e. vocabulary, oral reading fluency, comprehension, phonics); (4) The research design was experimental, quasi-experimental, or multiple treatment; (5) The study was published in English in a peer-reviewed journal from 2000 through 2015; (6) The study provided sufficient data for including in meta-analysis

Coding procedures

We employed meticulous coding procedures to collect and organize information from each study. We used the Vaughn, Elbaum, Wanzek, Scammacca, and Walker (2014) codesheet that was designed to align with the study features detailed in the What Works Clearinghouse (WWC) Design and Implementation Assessment Device (Valentine & Cooper, 2008). This codesheet was utilized in numerous previous reading syntheses (e.g., Swanson et al., 2011, 2012; Wanzek et al., 2015).

A combination of forced-choice items and open-ended items were used to record information related to: (a) participants, (b) methodology, (c) intervention and comparison descriptions, (d) measures, (e) results, and (f) potential moderators.

Three graduate research assistants studying reading intervention research participated in the coder training and reliability process. Initially, 4 h of training were provided to the graduate research assistants on the meaning of each codesheet item and examples of appropriate codes. Next, a researcher with experience using the codesheet modeled step-by-step how to complete a codesheet for one study. The graduate research assistants were assigned an article to code for the group to discuss collectively. Lastly, we used gold standard method (Gwet, 2001) to establish reliability between each of the graduate research assistants and the researcher. Interrater reliability was assessed as the number of items in agreements divided by the total number of items. To establish reliability, an overall interrater reliability score of .9 was required for the entire codesheet. The overall reliability scores ranged from .91 to .98.

Once initial reliability was established, two graduate research assistants studying reading intervention research independently coded each study and then met to identify and resolve coding discrepancies. When the coders were uncertain about a specific item, the trainer reviewed the study and the team made final decisions by consensus. Reliability was maintained through independent double-coding of each article. Additionally, a second reliability check was conducted using the gold standard method occurred four weeks after the initial reliability check (the coding process lasted a total of nine weeks). The overall interrater reliability scores for the second reliability check ranged from .92 to .96.

Effect size calculation

For all studies, Hedges's g was calculated using the means and standard deviations for treatment and comparison groups when such data were provided. In some cases, Cohen's d effect sizes and the treatment and comparison group sample sizes were used to calculate Hedges's g because means and standard deviations were not reported. All effect sizes and their standard errors were computed using the Comprehensive Meta Analysis (Version 3.3.070) software (CMA; Borenstein, Hedges, Higgins, & Rothstein, 2013).

Meta-analysis procedures

Separate meta-analyses were conducted for standardized and unstandardized measures because previous research has shown that effect sizes in reading intervention studies from standardized and unstandardized measures differ in magnitude (Scammacca, Roberts, Vaughn, & Stuebing, 2015; Willingham, 2007). In the meta-analysis of unstandardized outcome measures, 16 of the 20 studies contributed multiple effect sizes; 16 of 25 studies in the meta-analysis of standardized outcome measures contributed multiple effect sizes. Multiple effect sizes resulted from multiple measures being used to determine the treatment effect, more than one pair of treatment-comparison group contrasts, and multiple subgroup

comparisons (e.g., same group comparisons broken out by multiple grades). As a result, the meta-analytic data contained dependency from three sources.

To accommodate the dependency in the data, we conducted the meta-analyses using robust variance estimation (RVE; Hedges, Tipton, & Johnson, 2010) to adjust the standard errors via the robumeta package for R (Fisher & Tipton, 2013) instead of CMA. In RVE, the mean correlation between all pairs of effect sizes within a study (ρ) must be specified in order to estimate the study weights and calculate the between-study variance. Hedges et al. (2010) demonstrated that the value selected for ρ generally does not affect results very much and recommended implementing a sensitivity analysis by analyzing models with varying ρ values. Using .2, .5, and .8, we found no meaningful difference in the results across models for either unstandardized or standardized measures. The results reported below used a ρ of .8.

Using RVE, we estimated a series of meta-regression models for the meta-analyses of the standardized and unstandardized measures. RVE results have been shown to inflate Type I error when the number of studies included in the meta-analysis is less than 40 (Tipton, 2015). Therefore, the small-sample correction developed by Tipton (2015) was implemented in robumeta in all models. In each meta-analysis, an intercept-only model was run first to estimate the overall mean effect size. Additional meta-regression models were run to conduct four moderator analyses: intervention type (reading comprehension only vs. reading comprehension and vocabulary), hours of treatment (less than 30 h vs. 30 h or more), grade level (Grades 4–5 vs. Grades 6–8), and research design (quasi-experimental vs. experimental). These moderator variables were coded as categorical in order to maximize the number of studies that could be included in each moderator analysis given the small total number of studies in the meta-analysis and the information reported on these moderators in each study coded. As noted in Borenstein, Hedges, Higgins, and Rothstein (2009), statistical power is very low when fewer than five studies per category are included in a moderator analysis. Hours of treatment could not be operationalized as a continuous variable because this information tended to be reported as a range or a mean in the included studies.

Ideally, one meta-regression model with covariates for all moderators of interest would have been run for each meta-analysis. However, given that the overall number of studies that met the inclusion criteria was small and that not every study included information that allowed all moderator variables to be coded, this approach would not have yielded interpretable results due to insufficient degrees of freedom. Instead, we conducted each of the four moderator analyses in separate meta-regression models with the moderator as a covariate. In each model, the moderator variables were dummy coded 0 (first level of the variable in the comparison) and 1 (second level of the variable in the comparison) and included as covariates in the model. Because we ran four RVE regression models and wanted to maintain a $p < .05$ criteria for the moderator analysis, we adjusted the p value for determining statistical significance in each of the four moderator analyses to .0125 (.05 divided by 4; Abdi, 2007). To estimate a mean effect size for each category of the moderator variables, intercept-only models also were run for each level of the moderator. We recognize that power for the moderator analyses was low and consider these analyses to be exploratory in nature.

Results

The results section is divided into three sections. First, we provide information about the methodological characteristics of studies. Second, we present results from the meta-analyses that provide information about the effects of Tier 1 instruction for all students. Third, we present a narrative synthesis of 10 studies that disaggregated data for struggling readers.

Study characteristics

The literature search yielded 37 publications containing 40 studies (publications with two studies: Gayo et al., 2014; Johnston, McGeown & Watson, 2015; Vaughn et al., 2009). Research designs were equally represented with 20 experimental and 20 quasi-experimental studies. There were a total of 15,856 participants (range = 24–2082; median = 230). In every study, a general education teacher delivered instruction. See Tables 1 and 2 for additional information about the studies included in the meta-analyses and the struggling reader synthesis. Tables 3 and 4 contain Hedges's g effect sizes for all standardized (Table 3) and unstandardized (Table 4) outcome measures by study.

Hours of treatment and grade level

A total of 31 studies reported hours of treatment. The total hours of treatment across all studies was 1183, with a range of 4.5–125 and a mean of 38 (SD = 32.4). Sixteen studies were conducted in 4th through 5th grades, 17 studies were conducted in 6th through 8th grades and six studies were conducted in 9th through 12th grades. One study spanned 7th through 10th grades (Simmons et al., 2014).

Meta-analytic results

Standardized outcome measures

The meta-analysis of the standardized outcome measures included 70 effect sizes from 25 studies. The estimate of the mean effect size for these studies was 0.09 ($SE = .03$, $p = .008$, 95% CI [0.03, 0.16]), indicating a small but non-zero positive effect. The I^2 estimate of the percentage of between-study heterogeneity not due to chance variation in effects was 56.03%, with a τ^2 estimate of the true variance in the population of effects of .02. Differences in effect size due to moderator variables were investigated. Due to the small number of studies phonics/word recognition and fluency categories, only reading comprehension and reading comprehension plus vocabulary studies were included as the two intervention categories (refer to Tables 3, 4 for phonics/word recognition and fluency study effect sizes). We faced a similar situation with Grade 9–12 studies (refer to Tables 3, 4 for effect sizes). Because not enough of these studies were available, only Grades 4–5 and 6–8 studies were included. None of the moderator variables were statistically significant

Table 1 Intervention components

Study name	Intervention component			
	Reading comprehension	Vocabulary	Phonics/word recognition	Fluency
Alfassi (2009)	+			
Andreassen and Braten (2011)	+			
Baumann et al. (2002)		+		
Bowers and Kirby (2010)		+		
Bui and Fagan (2013)	+	+		
Chamberlain et al. (2009)	+	+	+	+
Fogarty et al. (2014)	+			
Gayo et al. (2014) Study 1	+			
Gayo et al. (2014) Study 2	+			
Guthrie and Lutz Klauda (2014)	+			
Harris et al. (2011)		+		
Huff and Nietfeld (2009)	+			
Johnston et al. (2015) Study 1			+	
Johnston et al. (2015) Study 2			+	
Kent et al. (2015)	+	+		
Klingner et al. (2004)	+			
Lesaux et al. (2010)	+	+		
Lesaux et al. (2014)	+	+		
Levine (2014)	+			
Lubliner and Smetana (2005)	+	+		
McCown and Thomason (2014)	+	+		
Reis et al. (2011)	+			+
Reisman (2012)	+			
Schunemann, Sporer, and Brunstein (2013)	+			
Shaaban (2006)	+	+		
Shippen et al. (2006)	+	+		+
Simmons et al. (2010)	+	+		
Simmons et al. (2014)	+	+		
Slavin et al. (2009)	+	+		
Stoeger et al. (2014)	+			
Swanson et al. (2015)	+	+		
Van Keer and Verhaeghe (2005)	+			
Vaughn et al. (2009) Study 1	+	+		
Vaughn et al. (2009) Study 2	+	+		
Vaughn et al. (2013)	+	+		
Vaughn et al. (2011)	+	+		
Vaughn et al. (2015)	+	+		

Table 2 Study characteristics

Study name	Study design	*N*	Struggling readers (*n*)	Grade level	Total treatment hours
Alfassi (2009)	E	115		6–8	24
Andreasse and Braten (2011)	Q	216		4–5	67.5
Baumann et al. (2002)	Q	89	LA = 4	4–5	10
Bowers & Kirby (2010)	E	81		4–5	16.66
Bui & Fagan (2013)	MT	49		4–5	6.66
Chamberlain et al. (2009)	E	405		6–8	125
Fogarty et al. (2014)	E	859		6–8	30
Gayo et al. (2014) Study 1	E	49		4–5	24–36
Gayo et al. (2014) Study 2	E	45		6–8	24–36
Guthrie and Lutz Klauda (2014)	E	557		6–8	NR
Harris et al. (2011)	Q	230	SWD = 24	9–12	7.5
Huff and Nietfeld (2009)	E	92		4–5	6–8
Ismail and Alexander (2005)	MT	48		9–12	4.5–5
Johnston et al. (2015) Study 1	MT	393		4–5	NR
Johnston et al. (2015) Study 2	MT	64		9–12	NR
Kent et al. (2015)	E	24		4–5	37.5–41.25
Klingner et al. (2004)	Q	212	LD = 29; LA = 48	6–8	NR
Lesaux et al. (2010)	Q	476		6–8	54
Lesaux et al. (2014)	E	2082		9–12	67.5
Levine (2014)	Q	37		4–5	NR
Lubliner and Smetana (2005)	Q	71		4–5	18
McCown and Thomason (2014)	Q	97		4–5	NR
Reis et al. (2011)	E	1192		9–12	100
Reisman (2012)	Q	200	SR = 42	4–5	NR
Schünemann et al. (2013)	Q	306		4–5	10.5
Shaaban (2006)	E	44		6–8	80
Shippen et al. (2006)	MT	44	SW = 44	4–5	NR
Simmons et al. (2010)	E	911		7–10	27
Simmons et al. (2014)	E	911	SR = 276	6–8	26.5
Slavin et al. (2009)	E	788	LA = 260	4–5	125
Stoeger et al. (2014)	Q	763		6–8	24–28
Swanson et al. (2015)	E	130	SWD = 130	6–8	25
Van Keer and Verhaeghe (2005)	Q			5	NR
Vaughn et al. (2009) Study 1	E	334		6–8	37.5
Vaughn et al. (2009) Study 2	E	453		6–8	37.5
Vaughn et al. (2013)	E	511	SR = 51	6–8	30
Vaughn et al. (2011)	E	723	SR = 92	6–8	29
Vaughn et al. (2015)	E	1442		6–8	26.25
Vaughn et al. (2013)	E	419		6–8	25–27
Wanzek et al. (2014)	E	394	LP = 76	9–12	37.5–41.25

E experimental, *Q* quasi-experimental, *LA* low-achieving, *MT* multiple treatment, *NR* not reported, *SWD* students with disabilities, *LD* learning disabilities, *SR* struggling readers

predictors of effect size. See Table 5 for results of the moderator analyses and Table 6 for the breakdown by each level of the moderators for the standardized outcome measures.

Unstandardized outcome measures

The meta-analysis of unstandardized outcome measures included 94 effect sizes from 20 studies. The estimate of the mean effect size for these studies was 0.47 ($SE = .11$, $p = .005$, 95% CI [0.20, 0.74]), indicating a moderate, non-zero positive effect. The I^2 estimate of the percentage of between-study heterogeneity not due to chance variation in effects was 92.99%, with a τ^2 estimate of the true variance in the population of effects of .21. None of the moderator variables were statistically significant predictors of effect size. See Table 5 for results of the moderator analyses and Table 7 for the breakdown by each level of the moderators for the standardized outcome measures.

Publication bias

Given that unpublished studies were not sought out for this meta-analysis, publication bias is a threat to the validity of our results. To evaluate the potential impact of publication bias, we implemented the trim-and-fill method (Duval & Tweedie, 2000) using a random effects model in CMA. Based on a funnel plot, this method removes effect sizes that cause asymmetry in the plot, calculates a mean effect, and then imputes additional effect sizes to make the plot symmetrical. In the process, it identifies the number of studies that may be missing from the meta-analysis due to publication bias and calculates an effect size that reflects adding in these missing studies.

Results of the trim-and-fill analysis indicated that publication bias affected the mean effect size estimate for the meta-analysis of standardized outcomes, with eight studies that had effect sizes that were smaller than the mean effect likely missing from the analysis. Estimating these effects resulted in an adjusted mean effect size of 0.02, with a 95% CI that includes zero [−0.05, 0.08]. In the meta-analysis of unstandardized outcomes, the trim-and-fill results indicated that no studies were likely missing that had effect sizes smaller than the mean effect.

Synthesis of effects of tier 1 reading instruction on struggling readers' outcomes

Struggling readers defined

Ten studies included in the meta-analysis provided disaggregated data for a subsample of struggling readers (Harris, Schumaker, & Deshler, 2011; Klingner, Vaughn, Arguelles, Hughes, & Leftwich, 2004; Reisman, 2012; Shippen, Houchins, Calhoon, Furlow, & Sartor, 2006; Simmons et al., 2014; Slavin, Chamberlain, Daniels, & Madden, 2009; Swanson, Wanzek, Vaughn, Roberts, & Fall, 2015;

Table 3 Effect sizes from standardized measures

Study	Comprehension measure	Comparison	*g*	Vocabulary Measure	Comparison	*g*	Fluency measure	Comparison	*g*
Alfassi (2009)	ORT Mult	T1 v. C	−0.08						
Andreasse and Braten (2011)	NLCT	T1 v. C	−0.12						
Chamberlain et al. (2009)	Gates-RC	T1 v. C	0.11	Gates-V	T1 v. C				
Fogarty et al. (2014)	Gates-RC	T1 v. C	0.02						
Gayo et al. (2014) Study 1	ACL	T1 v. C	0.82						
Gayo et al. (2014) Study 2	ACL	T1 v. C	0.24						
Huff and Nietfeld (2009)	Gates-RC	T1 v. C1	0.20						
		T1 v. C2	0.43						
		T2 v. C2	−0.42						
		T2 v C2	−0.34						
Johnston et al. (2015) Study 1	GRT	T1 v. C	0.35	EPVT	T1 v. C	0.15	WRAT-WR	T1 v. C	0.73
Johnston et al. (2015) Study 2	GRT	T1 v. C	0.21	EPVT	T1 v. C	0.21	WRAT-WR	T1 v. C	0.14
Klingner et al. (2004)	Gates-RC	T1 v. C	0.49						
Lesaux et al. (2010)	Gates-RC	T1 v. C	0.15	SAT-10	T1 v. C	0.00			
Lesaux et al. (2014)	Gates-RC	T1 v. C	−0.05						

Table 3 continued

Study	Comprehension measure	Comparison	*g*	Vocabulary Measure	Comparison	*g*	Fluency measure	Comparison	*g*
McCown & Thomason (2014)	GCRCT Comp	T1 v. C	0.34	GCRCT Voc	T1 v. C	0.10			
Reis et al. (2011)	ITBS	T1 v. C	0.11 (MS4)				ORF	T1 v. C	−0.18 (MS4)
	ITBS	T1 v. C	0.44 (MS5)				ORF	T1 v. C	−0.14 (MS5)
	ITBS	T1 v. C	−0.29 (SR4)				ORF	T1 v. C	−0.20 (SR4)
	ITBS	T1 v. C	−0.11 (SR5)				ORF	T1 v. C	0.00 (SR5)
	ITBS	T1 v. C	−0.74 (SS4)				ORF	T1 v. C	−0.36 (SS4)
	ITBS	T1 v. C	0.13 (SS5)				ORF	T1 v. C	0.63 (SS5)
	ITBS	T1 v. C	−0.12 (UM4)				ORF	T1 v. C	−0.24 (UM4)
	ITBS	T1 v. C	0.26 (UM5)				ORF	T1 v. C	0.31 (UM5)
	ITBS	T1 v. C	0.81 (US4)				ORF	T1 v. C	0.80 (US4)
	ITBS	T1 v. C	0.51 (US5)				ORF	T1 v. C	0.59 (US5)
Reisman (2012)	Gates-RC	T1 v. C	0.30						
Schünemann et al. (2013)	FRCT	T1 v. C	0.08						
	FRCT	T2 v. C	0.34						
Shaaban (2006)	Gates-RC	T1 v. C	0.06	Gates-V	T1 v. C	0.40			
Shippen et al. (2006)	WJIII-T[b]	T1 v. C	−0.63[a]				GORT-4	T1 v. C	−0.37
							WJIII-B[b]	T1 v. C	−0.64
Simmons et al. (2010)	Gates-RC	T1 v. C	0.38						
	Gates-RC	T2 v. C	0.36						
	TORC 3 SS	T1 v. C	0.27						
	TORC 3 SS	T2 v. C	0.26						
Simmons et al. (2014)	Gates-RC	T1 v. C	0.01						
Slavin et al. (2009)	Gates-RC	T1 v. C	0.09						
Stoeger et al. (2014)	HAMLET 3–4	T1 v. C	−0.12						
	HAMLET 3–4	T2 v. C	0.00						

Table 3 continued

Study	Comprehension measure	Comparison	*g*	Vocabulary Measure	Comparison	*g*	Fluency measure	Comparison	*g*
Swanson et al. (2015)	Gates-RC	T1 v. C	0.07						
Van Keer and Verhaeghe (2005)	RCAD	T1 v. C	0.39						
	RCAD	T2 v. C	0.32						
	RCAD	T3 v. C	0.33						
Vaughn et al. (2013)	Gates-RC	T1 v. C	0.11						
	TOSREC	T1 v. C	0.05						
Vaughn et al. (2011)	AIMSweb	T1 v. C	−0.05						
	Gates-RC	T1 v. C	0.12						
	TOSREC	T1 v. C	0.02						
Vaughn et al. (2015)	Gates-C-SS	T1 v. C	−0.1						
Vaughn et al. (2013)	Gates-RC	T1 v. C	0.31						
Wanzek et al. (2014)	Gates-RC	T1 v. C	0.03						

ORT Mult: Ortar Reading Test (standardized multiple-choice section), T: Treatment Condition, C: Comparison Condition, NLCT: Norwegian Learning Center Test, Gates-RC: Gates-MacGinitie Reading Comprehension Test, Gates-V: Gates-MacGinitie Reading Vocabulary Test, ACL: Reading Comprehension Assessment (5th/6th grade posttest), GRT: Group Reading Test, EPVT: English Picture Vocabulary Test, WRAT-WR: Wide Range Achievement Test (word reading subtest), SAT-10: Stanford Achievement Test, 10th Edition (reading vocabulary subtest), GCRCT Comp: Georgia's Criterion-Referenced Competency Test (comprehension subtest), GCRCT Voc: Georgia's Criterion-Referenced Competency Test (vocabulary subtest), ITBS: Iowa Tests of Basic Skills (reading comprehension subtest), MS4: Midwest Suburban 4th grade, MS5: Midwest Suburban 5th grade, SR4: Southern Rural 4th grade, SR5: Southern Rural 5th grade, SS4: Suburban South 4th grade, SS5: Suburban South 5th grade, UM4: Urban Magnet 4th grade, UM5: Urban Magnet 5th grade, US4: Urban Southeast 4th grade, US5: Urban Southeast 5th grade, ORF: Oral Reading Fluency, FRCT: Frankfurt Reading Comprehension Test, WJII-T: Woodcock Reading Mastery Total Reading Short Scale, GORT-4: Gray Oral Reading Test Fourth Edition Oral Reading Quotient, WJIII-B Woodcock Reading Mastery Basic Skills Cluster, TORC 3 SS: Test of Reading Comprehension, Social Studies Vocabulary Test, HAMLET 3–4: Hamburger Test for 3rd and 4th grade, RCAD: Reading Comprehension Achievement using Dutch Standards Test, TOSREC: Test of Reading Efficiency and Comprehension, AIMSweb: AIMSweb Curriculum Based Measure, Gates-C-SS: Gates MacGinitie Reading Comprehension in Social Studies

[a] Effect size calculated using the Woodcock Reading Mastery Tests—Revised which included both a fluency and a comprehension subtest

[b] Measure includes both a fluency and a comprehension subtest

Table 4 Effect sizes from unstandardized measures

Study	Reading component	Comparison	*g*	Reading component	Comparison	*g*
Alfassi (2009)	Comprehension	T1 v. C	0.50			
Andreassen and Braten (2011)	Comprehension	T1 v. C	−0.16, 0.12			
Baumann et al. (2002)	Vocabulary	T1 v. C	Range −0.11 to 1.76			
		T2 v. C	Range −0.05 to 1.00			
		T3 v. C	Range −0.07 to 0.89			
Bowers and Kirby (2010)	Vocabulary	T1 v. C	Range −0.23 to 0.74			
Bui & Fagan (2013)	Comprehension	T1 v. C	−0.06, 0.07	Vocabulary	T1 v. C	0.26
Fogarty et al. (2014)	Comprehension	T1 v. C	−0.20 (E)			
	Comprehension	T1 v. C	−0.01 (N)			
Guthrie and Lutz Klauda (2014)	Comprehension	T1 v. C	0.19			
Harris et al. (2011)	Comprehension	T1 v. C	4.48	Vocabulary	T1 v. C	2.52
		T2 v. C	0.51		T2 v. C	0.51
Ismail and Alexander (2005)		T1 v. T2	2.02			
		T1 v. T3	3.52			
		T2 v. T3	0.51			
Johnston et al. (2015) Study 2	Fluency ES	T1 v. C	Range 0.01–.64			
Kent et al. (2015)	Comprehension	T1 v. C	0.52	Vocabulary	T1 v. C	1.11
Lesaux et al. (2010)	Vocabulary	T1 v. C	Range 0.15–1.11			
Lesaux et al. (2014)	Comprehension	T1 v. C	0.06	Vocabulary	T1 v. C	0.14, 0.33
Levine (2014)	Comprehension	T1 v. C	1.69	Vocabulary	T1 v. C	Range 0.18–0.41
Lubliner and Smetana (2005)	Comprehension	T1 v. C	−0.31	Vocabulary	T1 v. C	−0.35
McCown and Thomason (2014)	Comprehension	T1 v. C	0.92			
Simmons et al. (2010)	Comprehension	T1 v. C	0.10	Vocabulary	T1 v. C	0.10
	Comprehension	T2 v. C	0.85		T2 v. C	0.85

Table 4 continued

Study	Reading component	Comparison	*g*	Reading component	Comparison	*g*
Simmons et al. (2014)	Comprehension	T1 v. C	0.05 (E)			
		T1 v. C	0.01 (N)			
Swanson et al. (2015)	Comprehension	T1 v. C	0.01			
Vaughn et al. (2009) Study 1	Comprehension	T1 v. C	0.71 (ELL)	Vocabulary	T1 v. C	0.63 (ELL)
		T1 v. C	0.72		T1 v. C	0.48
Vaughn et al. (2009) Study 2	Comprehension	T1 v. C	0.83 (ELL)	Vocabulary	T1 v. C	0.50 (ELL)
		T1 v. C	0.38		T1 v. C	0.31
Vaughn et al. (2015)	Comprehension	T1 v. C	−0.01			
Vaughn et al. (2013)	Comprehension	T1 v. C	0.44			

Ranges are reported for studies with more than two effect sizes

T Treatment Group, *C* Control Group, *E* Expository Text, *N* Narrative Text, *ELL* English Language Lear

Table 5 Moderator analysis results

	Coeff	SE	95% CI	*p*	*df*	I^2	τ^2	*n*	*k*	ρ	
Standardized outcomes											
Intervention type	−0.09	0.06	−0.21	0.03	.14	15	40.36	0.01	36	22	.8
Constant	0.11	0.06	−0.01	0.22	.07	8					
Hours of treatment	−0.02	0.06	−0.14	0.11	.80	14	55.87	0.01	52	20	.8
Constant	0.06	0.05	−0.05	0.18	.26	8					
Grade level	−0.12	0.07	0.27	0.03	.11	15	54.91	0.01	67	24	.8
Constant	0.16	0.06	0.01	0.30	.04	8					
Research design	−0.04	0.07	−0.19	0.11	.57	12	56.63	0.01	61	24	.8
Constant	0.11	0.06	−0.03	0.26	.11	6					
Unstandardized outcomes											
Intervention type	−0.10	0.16	−0.45	0.26	.56	10	82.32	0.05	41	17	.8
Constant	0.27	0.14	−0.10	0.64	.11	4					
Hours of treatment	−0.28	0.28	−0.94	0.37	.34	7	94.51	0.25	31	9	.8
Constant	0.53	0.23	0.02	1.05	.04	11					
Grade level	0.02	0.14	−0.28	0.32	.89	13	85.25	0.06	82	17	.8
Constant	0.22	0.12	−0.06	0.51	.11	6					
Research design	−0.27	0.43	−1.21	0.67	.54	11	93.70	0.16	77	19	.8
Constant	0.58	0.42	−0.48	1.63	.23	5					

Coeff coefficient, *n* number of effect sizes, *k* number of studies

Table 6 Effect size by moderator, standardized measures

	Coeff	SE	95% CI		*p*	*df*	I^2	τ^2	*n*	*k*
Intervention type										
Reading comprehension	0.12	0.05	0.00	0.23	.05	10	52.17	0.02	20	12
Comprehension and vocabulary	0.00	0.03	−0.07	0.07	.94	6	14.60	0.00	16	10
Hours of treatment										
Less than 30	0.06	0.05	−0.05	0.18	.23	9	66.03	0.02	21	11
30 or more	0.04	0.03	−0.05	0.12	.30	5	29.39	0.00	31	9
Grade level										
4 and 5	0.18	0.06	0.04	0.32	.02	10	61.48	0.04	48	12
6, 7, and 8	0.04	0.03	−0.04	0.11	.31	9	48.00	0.01	19	13
Research design										
Experimental	0.07	0.04	−0.01	0.15	.06	12	60.49	0.01	47	16
Quasi-experimental	0.11	0.06	−0.03	0.26	.11	6	45.30	0.01	14	8

Coeff coefficient, *n* number of effect sizes, *k* number of studies

Vaughn et al., 2011, 2015; Wanzek et al., 2014). Due to the limited number of studies, these data were not meta-analyzed separately, but are synthesized here. We defined students experiencing reading difficulty as:

Table 7 Effect size by moderator, unstandardized measures

	Coeff	SE	95% CI		*p*	*df*	I^2	τ^2	*n*	*k*
Intervention type										
Reading comprehension	0.33	0.16	−0.08	0.75	.10	5	86.74	0.11	11	7
Comprehension and vocabulary	0.17	0.07	0.00	0.33	.05	7	77.27	0.03	30	10
Hours of treatment										
Less than 30	0.54	0.24	0.02	1.07	.04	11	95.53	0.33	56	12
30 or more	0.21	0.14	−0.22	0.63	.24	4	86.22	0.08	24	5
Grade level										
4 and 5	0.22	0.11	−0.05	0.49	.09	7	85.19	0.21	57	8
6, 7, and 8	0.23	0.07	0.06	0.41	.02	8	85.30	0.04	25	9
Research design										
Experimental	0.28	0.07	0.11	0.45	.004	10	89.04	0.06	33	12
Quasi-experimental	0.67	0.43	−0.38	1.72	.17	6	96.44	1.01	44	7

Coeff coefficient, *n* number of effect sizes, *k* number of studies

- students with disabilities (Harris et al., 2011; Shippen et al., 2006; Swanson et al., 2015),
- low achievers based on pretest scores (Klingner et al., 2004; Simmons et al., 2014; Slavin et al. 2009; Wanzek et al., 2014), or
- struggling readers as determined by failing scores on state reading assessments or scoring below the 25th percentile on the Gates-MacGinitie reading comprehension subtest at pretest (Vaughn et al., 2011, 2015).

Study characteristics

One study was conducted with fourth-grade students (Klingner et al., 2004), five were conducted at the middle school level (Shippen et al., 2006; Slavin et al., 2009; Swanson et al., 2015; Vaughn et al., 2011, 2015), three at the high school level (Harris et al., 2011; Reisman, 2012; Wanzek et al., 2014), and one across Grades 7 through 10 (Simmons et al., 2014). Samples of struggling readers ranged from 24 to 276 students, with a mean of 107.5 and a median of 76.5. Six studies were experimental (Simmons et al., 2014; Slavin et al., 2009; Swanson et al., 2015; Vaughn et al., 2011, 2015; Wanzek et al., 2014), three were quasi-experimental (Harris et al., 2011; Klingner et al., 2004; Reisman, 2012), and one was a multiple-treatment study (Shippen et al., 2006). All studies reported fidelity of implementation data. In eight studies authors used standardized measures of reading comprehension; three of those studies also used standardized measures of reading fluency and one used a standardized measure of vocabulary. Two studies included unstandardized measures of reading comprehension and one included unstandardized measures of vocabulary. Finally, two studies reported unstandardized measures

of content knowledge. Effect sizes in the struggling reader synthesis were calculated using the disaggregated struggling reader sample alone. Therefore, these effects represent the difference in outcomes between struggling readers who received a treatment and struggling readers who did not.

Tier 1 reading instruction effects for struggling readers

Three studies reported the effects of reading comprehension interventions (Klingner et al., 2004; Reisman, 2012; Wanzek et al., 2014), five studies investigated the effects of multicomponent reading comprehension plus vocabulary interventions (Simmons et al., 2014; Slavin et al., 2009; Swanson et al., 2015; Vaughn et al., 2011, 2013), and one study investigated vocabulary interventions (Harris et al., 2011). Refer to Table 1 for an additional summary of these studies' intervention components. Overall, effect sizes among struggling readers for standardized measures of reading ranged from −0.05 to 0.49. Effect sizes among struggling readers on unstandardized measures of reading ranged from 0.01 to 2.52. Following is a description of effects by intervention type.

Reading comprehension intervention effects for struggling readers Klingner et al. (2004) reported that reading comprehension instruction improved struggling readers performance on the Gates-MacGinitie Reading Comprehension test with an effect size of 0.52, although this difference was not statistically significant. Two additional studies investigated the effects of comprehension instruction on unstandardized measures of content knowledge, with effects ranging from −0.30 to 0.22 (Reisman, 2012; Wanzek et al., 2014).

Multicomponent intervention effects for struggling readers Four of the five studies that investigated the effects of multicomponent reading comprehension plus vocabulary instruction on standardized measures of reading comprehension reported effects ranging from 0.00 to 0.36 (Slavin et al., 2009; Swanson et al., 2015; Vaughn et al., 2011, 2015). Simmons et al. (2014) did not provide information to calculate effect sizes; however, the authors indicated higher-performing readers made greater gains than lower-performing readers. Two of these studies also reported an effect size of 0.17 on the Gates Vocabulary test (Slavin et al., 2009) and 0.25 on a standardized measure of reading fluency and comprehension (Vaughn et al., 2015). Finally, Swanson et al. (2015) reported effect sizes of 0.35 and 0.30 on researcher-developed content reading comprehension and content knowledge measures, respectively.

Vocabulary intervention effects for struggling readers One study investigated vocabulary instruction targeting morphemic analysis and memory strategies (e.g., keywords and visual imagery). Harris et al. (2011) reported differences in favor of word mapping instruction ($ES = 2.12$; $ES = 1.32$) and a visual memory strategy ($ES = 2.12$; $ES = 0.26$) on researcher-developed tests of word knowledge and morphological analysis, respectively.

Discussion

Results from this meta-analysis of 37 publications conducted between 2000 and 2015 reveal significant, positive effects for Tier 1 reading instruction on comprehension and vocabulary outcomes, indicating that fourth through 12th graders who receive Tier 1 instruction that includes at least one reading component outperform their peers who did not receive the reading components on reading outcome measures. Results from standardized measures indicate somewhat smaller gains of around one tenth of one standard deviation for students receiving the intervention. However, on unstandardized measures that are more closely aligned with the intervention itself, gains were larger and equaled about one half of a standard deviation for students who received Tier 1 reading instruction. The finding of differences in results using standardized versus unstandardized measures is aligned with prior meta-analytic results that indicate when researchers use standardized measures, smaller effect sizes are reported (Scammacca et al., 2015; Swanson, Hoskyn, & Lee, 1999).

Although we conducted moderator analyses on four variables (intervention type, hours of treatment, grade level, and research design) due to the amount of heterogeneity detected in the overall effect sizes for standardized and unstandardized measures, the small number of studies substantially limits our ability to draw conclusions about their effects. No variable was a statistically significant predictor of effect size. This does not mean these variables do not affect the potency or the efficacy of Tier 1 reading instruction. Instead, it means that with additional studies, the predictive power of these variables might become more apparent. In other words, we do not know whether these variables are influential or not. To collect additional studies for the meta-analyses, we considered extending the search years. However, there is some indication that year of publication impacts effect size and that studies published between 1980 and 2004 may very well come from a different population of studies (Scammacca et al., 2015). For that reason, we chose to reduce the variability that would be introduced by going back further in time and instead maintained the restricted range of dates to maintain precision of results.

Grade level

The effect size for interventions conducted in Grades 4 and 5 was 0.18 on standardized measures and 0.22 on unstandardized measures. Based on prior work (e.g., Bloom, Hill, Black, & Lipsey, 2008; Scammacca et al., 2015), we expected higher effect sizes in lower grades (i.e., Grades 4 and 5). One possible explanation is the changing nature of the business as usual (BAU) condition, particularly at the lower grades. When an intervention is compared to BAU instruction received by a comparison group, we are not comparing the intervention to no instruction. To the contrary, we are comparing two sets of instruction—that is (a) intervention and (b) instruction typically provided by teachers. Lemons, Fuchs, Gilbert, and Fuchs (2014) hypothesized that the BAU condition represents a population that may have collectively shift[ed] their behavior over time, resulting in an increase in the overall

quality of instruction in BAU conditions and thereby decreasing the measured impact of interventions compared to the BAU condition. To test this hypothesis, they conducted a retrospective analysis of 9 years worth of randomized control trials investigating the efficacy of the same kindergarten intervention. They investigated the outcome scores for students assigned to the BAU condition and noticed, "the relative value [of the intervention] lessened over time because the performance of control students increased markedly" (p. 245). Results from our moderator analysis may support further support for this hypothesis. We expected elementary effects to be larger and believe that improved BAU practice at the elementary level obscured the measured effect of treatments at the elementary level, resulting in an inability to detect differences in effect between elementary and secondary students.

Intervention type

Researchers conducted studies on a variety of intervention types, including single component interventions—reading comprehension (e.g., Fogarty et al., 2014), vocabulary (e.g., Bauman et al., 2002), and phonics/word recognition (e.g., Johnston et al., 2015)—and multi-component interventions that included comprehension plus vocabulary (e.g., Swanson et al., 2015) or phonics/word recognition plus fluency (e.g., Chamberlain et al., 2009). We only had enough studies in two categories (i.e., comprehension and comprehension plus vocabulary) to investigate the moderating effect of intervention type on outcomes. The effect of reading comprehension interventions as measured by standardized assessments was 0.12 ($SE = .05$; 95% CI [0.00–0.23]). Although this is considered small and the confidence interval includes 0, the mean effect is not inconsequential, particularly for students in the 4th through 12th grade range. The mean effect size is aligned with other reported effect sizes on standardized reading comprehension measures of 0.06–0.19 at the high school level (Bloom et al., 2008).

Our finding of no differences could reflect the effect of reading comprehension instruction in both categories of studies. If comprehension instruction were the primary driver of differences in outcomes, the addition of vocabulary instruction might not have a powerful enough impact to result in significantly larger effects in these multi-component interventions. In a recent meta-analysis of vocabulary interventions (Elleman, Lindo, Morphy, & Compton, 2009), vocabulary interventions produced an average effect of 0.10 on standardized comprehension outcomes, providing evidence that vocabulary interventions alone impact comprehension outcomes to some extent. However, they also concluded that they were unable to provide recommendations about which vocabulary interventions best impact comprehension outcomes. Additional studies of single and multi-component Tier 1 reading interventions should be conducted in order to determine which components are most potent in producing effects on reading outcomes for fourth through twelfth graders.

Duration

The finding of no differences based on hours of treatment when studies were categorized as less than 30 h vs. 30 h or more may reflect several factors. This finding is aligned with Wanzek et al. (2013) investigation of the moderating role of hours of treatment among Tier 3 interventions. While it may align with prior meta-analytic findings, we should also consider reasons that may mitigate the finding of no differences based on hours of treatment. First, the inability to examine study duration as a continuous variable may blur the true impact of the treatment's duration. Three other meta-analyses have examined duration as a moderator of the impact of Tier 2 and 3 reading interventions (e.g., Flynn, Zheng, & Swanson, 2012; Scammacca et al., 2015; Wanzek et al., 2013). These authors also categorized interventions by duration rather than examining duration as a continuous variable due to the same issue we encountered with the way in which intervention studies reported their duration. Without more precise information about the hours of intervention provided, unpacking the true effect of duration may not be possible.

Second, we chose 30 h as the division point between the groups of studies in large part because it created two groups of sufficient size to allow for moderator analysis in both meta-analyses. Alternatively, Scammacca et al. (2015) divided studies into three groups: less than 5, 6–12, and more than 12 h and reported significantly smaller effects for longer interventions compared to shorter interventions. In the current meta-analysis, another dividing point might reveal the true impact of duration as a moderator of the impact of Tier 1 interventions. As the corpus of Tier 1 studies grows, it may be possible to divide duration in ways that expose a more accurate estimate of effect due to duration.

Duration should also be considered when interpreting the educational implications of small effect sizes. For example, the reading comprehension effect size was only 0.12. However, as acknowledged prior, this effect "is not inconsequential" for students in grades 4th through 12th grades and could be an artifact of rather short intervention durations. Among the reading comprehension studies, duration ranged from 4.5 to 100 h with a mean of 19.9 h. Even considering the study reporting 100 h of instruction (Reis, McCoach, Little, Muller, & Kanisken, 2011), this equals approximately 14 school days. This suggests that longer-term applications of reading comprehension interventions are necessary to improve student, reading outcomes.

Struggling reader outcomes

The synthesis of studies in which authors disaggregated data for struggling readers indicates that multi-component interventions, combining reading comprehension and vocabulary instruction delivered in the Tier 1 setting may be effective. One intervention in particular—Collaborative Strategic Reading (Klingner et al., 2004)—produced a moderate effect on a standardized measure of reading comprehension. In Collaborative Strategic Reading, students are taught to identify the main idea of a section of text, recognize vocabulary they do not understand, and then use context clue fix-up strategies to learn the meaning of the word. Authors of

another study investigating the effects of a multicomponent intervention for struggling readers (PACT; Swanson et al., 2015; Wanzek et al., 2014) reported small-to-moderate effects on researcher-developed measures of reading comprehension. In this approach, teachers lead students in text reading and classroom discourse guided by careful questioning. Teachers also provide explicit vocabulary instruction and periodic review over the course of instructional units.

With this in mind, we can tentatively conclude that struggling readers who receive multi-component reading comprehension plus vocabulary instruction in Tier 1 settings will outperform their struggling reader peers who do not receive the intervention. However, there are two caveats. First, few of these effects were statistically significant, indicating that struggling readers may require, in addition to high quality Tier 1 instruction, Tier 2 support to produce group differences in outcomes. Second, it is unclear if these gains begin to close the gap between struggling readers and typically-achieving students. This is a question largely unanswered by the corpus of Tier 1 studies that provided disaggregated data on struggling readers.

It is important to determine the effect of Tier 1 reading intervention on outcomes for struggling readers since students who struggle with reading but are not identified with a disability almost always receive instruction within the Tier 1 setting—particularly at the middle and high school levels. In addition, more than half of students with disabilities are educated in the Tier 1 setting for a majority of their day (National Center for Education Statistics, 2015). It is possible to conduct sub-group analysis of struggling readers' outcomes within the context of large randomized control trials focused on the Tier 1 setting, but it takes careful planning. For example, consider a large-scale initial study and subsequent replication of a multi-component Tier 1 intervention delivered in general education social studies classrooms across two states (Vaughn et al., 2013, 2015). Researchers built a well-powered sample of students with disabilities (who were struggling readers) and conducted a sub-group analysis of effects for this particular sample (Swanson et al., 2015). The resulting body of work provides evidence of the effects of the intervention for all students served in the Tier 1 setting, and also supports its effect on students with disabilities who typically struggle with reading and are served within the same Tier 1 setting. Further large scale research designs should investigate the impact of Tier 1 instruction on closing the achievement gap for struggling readers.

Limitations

The data reported in this meta-analysis are limited by the number and type of studies conducted between 2000 and 2015 that investigated the effects of Tier 1 reading instruction on reading outcomes. One key finding is the limited number of studies examining Tier 1 reading instruction for students in Grades 4 through 12. Additional research in this area is needed. The relatively small sample of studies and the presence of multiple dependent effect sizes within these studies led us to implement RVE with a small-sample size correction in all of the moderator analyses with little power to detect statistically significant predictors of effect size. Second, authors

often provided incomplete descriptions of key variables needed for coding all moderator variables. For example, we included a duration moderator. Optimally, we would like to have treated this variable as continuous in nature. However, while 34 of the 37 articles included duration information, it was usually reported as a range, estimation, or average duration, forcing a categorical treatment of the duration variable and limiting our ability to detect differential effects due to duration. Finally, the analysis for duration did not take into account differences in instructional contexts. Both variables combined (duration + instructional context) might moderate effect sizes. However, we were not able to examine this combined effect due to the small number of studies available.

Conclusions

Results from the meta-analyses support the use of Tier 1 reading interventions in general education classrooms with limited evidence that Tier 1 reading instruction alone is effective for struggling readers. The greatest volume of evidence indicates infusing reading comprehension into English language arts/reading classes and content area classes would be beneficial to all students, including struggling readers. However, in order to examine the impact of other types of Tier 1 interventions, additional research must be conducted. With so few studies that include phonics/word recognition and fluency instruction, it is not possible from this meta-analysis to determine the impact of these intervention components when delivered in the Tier 1 setting. Finally, the duration of these studies was relatively short, with the longest study comprising 125 h, or approximately 17.9 days of school. When a greater quantity of Tier 1 interventions are investigated over longer periods of time, perhaps then will we have a better picture of the Tier 1 instructional influence for students in Grades 4–12.

Acknowledgements The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R305A150407 to The University of Texas at Austin. The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education.

References

Abdi, H. (2007). Bonferroni and Šidák corrections for multiple comparisons. In N. J. Salkind (Ed.), *Encyclopedia of measurement and statistics* (Vol. 3, pp. 103–107). Thousand Oaks, CA: Sage.

Alfassi, M. (2009). The efficacy of a dialogic learning environment in fostering literacy. *Reading Psychology, 30,* 539–563. doi:10.1080/02702710902733626.

Andreassen, R., & Braten, I. (2011). Implementation and effects of explicit reading comprehension instruction in fifth-grade classrooms. *Learning and Instruction, 21,* 520–537. doi:10.1016/j.learninstruc.2010.08.003.

Baumann, J. F., Edwards, E. C., Font, G., Tereshinski, C., Kame'enui, E. J., & Olejnik, S. (2002). Teaching morphemic and contextual analysis to fifth-grade students. *Reading Research Quarterly, 37,* 150–176. doi:10.1598/rrq.37.2.3.

Bloom, H. S., Hill, C. J., Black, A. R., & Lipsey, M. W. (2008). Performance trajectories and performance gaps as achievement effect-size benchmarks for educational interventions. *Journal of Research on Educational Effectiveness, 1,* 289–328. doi:10.1080/19345740802400072.

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. Chichester: Wiley.

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2013). *Comprehensive meta analysis* (Version 3.3.070), Englewood, NJ: Biostat.

Bowers, P., & Kirby, J. (2010). Effects of morphological instruction on vocabulary acquisition. *Reading and Writing, 23,* 515–537. doi:10.1007/s11145-009-9172-z.

Bui, Y., & Fagan, Y. (2013). The effects of an integrated reading comprehension strategy: A culturally responsive teaching approach for fifth-grade students' reading comprehension. *Preventing School Failure, 57,* 59–69. doi:10.1080/1045988x.2012.664581.

Chamberlain, A., Daniels, C., Madden, N., & Slavin, R. (2009). A randomized evaluation of the success for all middle school reading program. *Middle Grades Research Journal, 2*(1), 1–21.

Cheung, A. C., & Slavin, R. E. (2012). Effective reading programs for Spanish-dominant English language learners (ELLs) in the elementary grades: A synthesis of research. *Review of Educational Research, 82,* 351–395. doi:10.3102/0034654312465472.

Cisco, B. K., & Padrón, Y. (2012). Investigating vocabulary and reading strategies with middle grades English language learners: A research synthesis. *RMLE Online, 36*(4), 1–23. doi:10.1080/19404476.2012.11462097.

Cooper, H., Hedges, L. V., & Valentine, J. C. (Eds.). (2009). *The handbook of research synthesis and meta-analysis*. New York: Russell Sage Foundation.

Duval, S., & Tweedie, R. (2000). Trim and fill: A simple funnel-plot–based method of testing and adjusting for publication bias in meta-analysis. *Biometrics, 56,* 455–463. doi:10.1111/j.0006-341x.2000.00455.x.

Ehri, L. C., Nunes, S. R., Willows, D. M., Schuster, B. V., Yaghoub-Zadeh, Z., & Shanahan, T. (2001). Phonemic awareness instruction helps children learn to read: Evidence from the National Reading Panel's meta-analysis. *Reading Research Quarterly, 36,* 250–287. doi:10.1598/rrq.36.3.2.

Elleman, A. M., Lindo, E. J., Morphy, P., & Compton, D. L. (2009). The impact of vocabulary instruction on passage-level comprehension of school-age children: A meta-analysis. *Journal of Research on Educational Effectiveness, 2,* 1–44. doi:10.1080/19345740802539200.

Faggella-Luby, M. N., Drew, S. V., & Schumaker, J. B. (2015). Not such a simple story: Contradictory evidence from a review of story structure research for students at-risk. *Learning Disabilities Research & Practice, 30,* 61–75. doi:10.1111/1drp.12057.

Fisher, Z., & Tipton, E. (2013). *Robumeta: An R-package for robust variance estimation in meta-analysis*. Retrieved from https://arxiv.org/abs/1503.02220

Fletcher, J. M., & Vaughn, S. (2009). Response to intervention: Preventing and remediating academic difficulties. *Child Development Perspectives, 3,* 30–37. doi:10.1111/j.1750-8606.2008.00072.x.

Flynn, L. J., Zheng, X., & Swanson, H. (2012). Instructing struggling older readers: A selective meta-analysis of intervention research. *Learning Disabilities Research & Practice, 27,* 21–32. doi:10.1111/j.1540-5826.2011.00347.x.

Fogarty, M., Oslund, E., Simmons, D., Davis, J. J., Simmons, L., Anderson, L., et al. (2014). Examining the effectiveness of a multicomponent reading comprehension intervention in middle schools: A focus on treatment fidelity. *Educational Psychology Review, 26,* 425–449. doi:10.1007/s10648-014-9270-6.

Fuchs, D., & Deshler, D. D. (2007). What we need to know about responsiveness to intervention (and shouldn't be afraid to ask). *Learning Disabilities Research & Practice, 22,* 129–136. doi:10.1111/j.1540-5826.2007.00237.x.

Gayo, E., Deano, M., Conde, A., Ribeiro, I., Cadime, I., & Alfonso, S. (2014). Effect of an intervention program on the reading comprehension processes and strategies in 5th and 6th grade students. *Psicothema, 26,* 464–470. doi:10.7334/psicothema2014.42.

Guthrie, J., & Lutz Klauda, S. (2014). Effects of classroom practices on reading comprehension, engagement, and motivations for adolescents. *Reading Research Quarterly, 49,* 387–416. doi:10.1002/rrq.81.

Gwet, K. (2001). *Handbook of inter-rater reliability: How to estimate the level of agreement between two or multiple raters*. Gaithersburg, MD: STATAXIS Publishing Company.

Harris, M., Schumaker, J., & Deshler, D. D. (2011). The effects of strategic morphological analysis instruction on the vocabulary performance of secondary students with and without disabilities. *Learning Disability Quarterly, 34,* 17–33. doi:10.1177/073194871103400102.

Hedges, L. V., Tipton, E., & Johnson, M. C. (2010). Robust variance estimation in meta-regression with dependent effect size estimates. *Research Synthesis Methods, 1,* 39–65. doi:10.1002/jrsm.5.

Huff, J., & Nietfeld, J. (2009). Using strategy instruction and confidence judgments to improve metacognitive monitoring. *Metacognition and Learning, 4,* 161–176. doi:10.1007/s11409-009-9042-8.

Ismail, H., & Alexander, J. (2005). Learning within scripted and nonscripted peer-tutoring sessions: The Malaysian context. *The Journal of Educational Research, 99,* 67–77. doi:10.3200/joer.99.2.67-77.

Johnston, R., McGeown, S., & Watson, J. (2015). Long-term effects of synthetic versus analytic phonics teaching on the reading and spelling ability of 10 year old boys and girls. *Reading and Writing, 25,* 1365–1384. doi:10.1007/s11145-011-9323-x.

Kent, S., Wanzek, J., Swanson, E. A., & Vaughn, S. (2015). Team-based learning for students with high-incidence disabilities in high school social studies classrooms. *Learning Disabilities Research & Practice, 30,* 3–14. doi:10.1111/1drp.12048.

Klingner, J. K., Vaughn, S., Arguelles, M. E., Hughes, M. T., & Leftwich, S. A. (2004). Collaborative strategic reading: "Real world" lessons from classroom teachers. *Remedial and Special Education, 25,* 291–302. doi:10.1177/07419325040250050301.

Lemons, C. J., Fuchs, D., Gilbert, J. K., & Fuchs, L. S. (2014). Evidence-based practices in a changing world: Reconsidering the counterfactual in education research. *Educational Researcher, 43,* 242–252. doi:10.3102/0013189X14539189.

Lesaux, N., Kieffer, M., Faller, S. E., & Kelley, J. (2010). The effectiveness and ease of implementation of an academic vocabulary intervention for linguistically diverse students in urban middle schools. *Reading Research Quarterly, 45,* 196–228. doi:10.1598/rrq.45.2.3.

Lesaux, N., Kieffer, M., Kelley, J., & Harris, J. R. (2014). Effects of academic vocabulary instruction for linguistically diverse adolescents: Evidence from a randomized field trial. *American Educational Research Journal, 51,* 1–36. doi:10.3102/0002831214532165.

Levine, S. (2014). Making interpretation visible with an affect-based strategy. *Reading Research Quarterly, 49,* 283–303. doi:10.1002/rrq.71.

Lubliner, S., & Smetana, L. (2005). The effects of comprehensive vocabulary instruction on title I students' metacognitive word-learning skills and reading comprehension. *Journal of Literacy Research, 37,* 163–200. doi:10.1207/s15548430j1r3702_3.

McCown, M., & Thomason, G. (2014). Informational text comprehension: Its challenges and how collaborative strategic reading can help. *Reading Improvement, 2,* 237–253.

Puzio, K., & Colby, G. T. (2013). Cooperative learning and literacy: A meta-analytic review. *Journal of Research on Educational Effectiveness, 6,* 339–360. doi:10.1080/19345747.2013.775683.

Reichrath, E., de Witte, L. P., & Winkens, I. (2010). Interventions in general education for students with disabilities: A systematic review. *International Journal of Inclusive Education, 14,* 563–580. doi:10.1080/13603110802512484.

Reis, S. M., McCoach, D. B., Little, C., Muller, L., & Kaniskan, R. B. (2011). The effects of differentiated instruction and enrichment pedagogy on reading achievement in five elementary schools. *American Educational Research Journal, 48,* 462–501. doi:10.3102/0002831210382891.

Reisman, A. (2012). Reading like a historian: A document-based history curriculum intervention in urban high schools. *Cognition and Instruction, 30,* 86–112. doi:10.1080/07370008.2011.634081.

Reznitskaya, A., Kuo, L. J., Clark, A. M., Miller, B., Jadallah, M., Anderson, R. C., et al. (2009). Collaborative reasoning: A dialogic approach to group discussions. *Cambridge journal of education, 39,* 29–48. doi:10.1080/03057640802701952.

Scammacca, N., Roberts, G., Vaughn, S., & Stuebing, 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.

Schunemann, N., Sporer, N., & Brunstein, J. (2013). Integrating self-regulation in whole-class reciprocal teaching: A moderator–mediator analysis of incremental effects on fifth graders' reading comprehension. *Contemporary Educational Psychology, 38,* 289–305. doi:10.1016/j.cedpsych.2013.06.002.

Shaaban, K. (2006). An initial study of the effects of cooperative learning on reading comprehension, vocabulary acquisition, and motivation to read. *Reading Psychology, 27,* 377–403. doi:10.1080/02702710600846613.

Shippen, M., Houchins, D. E., Calhoon, M. B., Furlow, C., & Sartor, D. (2006). The effects of comprehensive school reform models in reading for urban middle school students with disabilities. *Remedial and Special Education, 27,* 322–328. doi:10.1177/07419325060270060101.

Simmons, D., Fogarty, M., Oslund, E. L., Simmons, L., Hairrell, A., Davis, J., et al. (2014). Integrating content knowledge-building and student-regulated comprehension practices in secondary english

language arts classes. *Journal of Research on Educational Effectiveness, 7,* 309–330. doi:10.1080/19345747.2013.836766.

Simmons, D., Hairrell, A., Edmonds, M., Vaughn, S., Larsen, R., Willson, V., et al. (2010). A comparison of multiple-strategy methods: Effects on fourth-grade students' general and content-specific reading comprehension and vocabulary development. *Journal of Research on Educational Effectiveness, 3,* 121–156. doi:10.1080/19345741003596890.

Slavin, R., Chamberlain, A., Daniels, C., & Madden, N. (2009). The reading edge: A randomized evaluation of a middle school cooperative reading program. *Effective Education, 1,* 13–26. doi:10.1080/19415530903043631.

Stoeger, H., Sontag, C., & Ziegler, A. (2014). Impact of a teacher-led intervention on preference for self-regulated learning, finding main ideas in expository texts, and reading comprehension. *Journal of Educational Psychology, 106,* 799–814. doi:10.1037/a0036035.

Swanson, E., Hairrell, A., Kent, S., Ciullo, S., Wanzek, J. A., & Vaughn, S. (2012). A synthesis and meta-analysis of reading interventions using social studies content for students with learning disabilities. *Journal of Learning Disabilities, 47,* 178–195. doi:10.1177/0022219412451131.

Swanson, H. L., Hoskyn, M., & Lee, C. (1999). *Interventions for students with learning disabilities: A meta-analysis of treatment outcomes*. New York: Guilford.

Swanson, E., Vaughn, S., Wanzek, J., Petscher, Y., Heckert, J., Cavanaugh, C., et al. (2011). A synthesis of read-aloud interventions on early reading outcomes among preschool through third graders at risk for reading difficulties. *Journal of Learning Disabilities, 44,* 258–275. doi:10.1177/0022219410378444.

Swanson, E., Wanzek, J., Vaughn, S., Roberts, G., & Fall, A. M. (2015). Improving reading comprehension and social studies knowledge among middle school students with disabilities. *Exceptional Children, 81,* 426–442. doi:10.1177/0014402914563704.

The National Center for Education Statistics (2015). *The condition of education 2015* (NCES 2015-144). Institute of Education Sciences, U.S. Department of Education, Washington DC: U.S. Department of Education.

Tipton, E. (2015). Small sample adjustments for robust variance estimation with meta-regression. *Psychological Methods, 20,* 375–393. doi:10.1037/met0000011.

U.S. Department of Education, Institute of Education Sciences, What Works Clearinghouse. (2009). *English Language Learner intervention report: Success for All*™. Retrieved from http://whatworks.ed.gov

U.S. Department of Education, Institute of Education Sciences, What Works Clearinghouse. (2013). *Beginning reading intervention report: Accelerated Reader*™. Retrieved from http://whatworks.ed.gov

Valentine, J. C., & Cooper, H. (2008). A systematic and transparent approach for assessing the methodological quality of intervention effectiveness research: The study design and implementation assessment device (Study DIAD). *Psychological Methods, 13,* 130–149. doi:10.1037/1082-989X.13.2.130.

Van Keer, H., & Verhaeghe, J. P. (2005). Comparing two teacher development programs for innovating reading comprehension instruction with regard to teachers' experiences and student outcomes. *Teaching and Teacher Education, 21,* 543–562. doi:10.1016/j.tate.2005.03.002.

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., & Fletcher, J. M. (2012). Response to intervention with secondary school students with reading difficulties. *Journal of Learning Disabilities, 45,* 244–256. doi:10.1177/0022219412442157.

Vaughn, S., & Fuchs, L. S. (2003). Redefining learning disabilities as inadequate response to instruction: The promise and potential problems. *Learning Disabilities Research & Practice, 18,* 137–146. doi:10.1111/1540-5826.00070.

Vaughn, S., Klingner, J., Swanson, E. A., Boardman, A. G., Roberts, G., Mohammed, S., et al. (2011). Efficacy of collaborative strategic reading with middle school students. *American Educational Research Journal, 48,* 938–964. doi:10.3102/0002831211410305.

Vaughn, S., Roberts, G., Klingner, J., Swanson, E. A., Boardman, A. G., Stillman-Spisak, S. J., et al. (2009). Collaborative strategic reading: Findings from experienced implementers. *Journal of Research on Educational Effectiveness, 6,* 137–163. doi:10.1080/19345747.2012.741661.

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,* 31–50. doi:10.1007/s10648-014-9274-2.

Vaughn, S., Swanson, E. A., Roberts, G., Wanzek, J., Stillman-Spisak, S. J., Solis, M., et al. (2013). Improving reading comprehension and social studies knowledge in middle school. *Reading Research Quarterly, 48,* 77–93. doi:10.1177/0014402914563704.

Wanzek, J., Vaughn, S., Kent, S., Swanson, E. A., Roberts, G., Haynes, M., et al. (2014). The effects of team-based learning on social studies knowledge acquisition in high school. *Journal of Research on Educational Effectiveness, 7,* 183–204. doi:10.1080/19345747.2013.836765.

Wanzek, J., Vaughn, S., Scammacca, N., Gatlin, B., Walker, M. A., & Capin, P. (2015). Meta-analyses of the effects of tier 2 type reading interventions in grades K-3. *Educational Psychology Review, 28,* 1–26. doi:10.1007/s10648-015-9321-7.

Wanzek, J., Vaughn, S., Scammacca, N., Metz, K., Murray, C., Roberts, G., et al. (2013). Extensive reading interventions for students with reading difficulties after grade 3. *Review of Educational Research, 83,* 163–195. doi:10.3102/0034654313477212.

Willingham, D. T. (2007). Ask the cognitive scientist: The usefulness of brief instruction in reading comprehension strategies. *American Educator, 30*(4), 39–45.