INTERNATIONAL LITERACY ASSOCIATION READING RESEARCH QUARTERLY

Beyond Decoding: A Meta-Analysis of the Effects of Language Comprehension Interventions on K–5 Students' Language and Literacy Outcomes

Rebecca D. Silverman
Stanford University, California, USA

Erika Johnson
University of Iowa, Iowa City, USA

Kristin Keane
Saurabh Khanna
Stanford University, California, USA

ABSTRACT

The debate over the science of reading has focused primarily on decoding (i.e., connecting letters and sounds to read words) and whether to use phonics to teach it. However, research on reading has included much more than decoding. Language comprehension, which allows readers to derive meaning from text, is an equally critical component of reading. Research has suggested that explicit instruction on the components of language comprehension—vocabulary and semantics, morphology, and syntax—can support language and reading comprehension. To inform the field on the science of reading as it pertains to language comprehension, in this meta-analysis of recent language comprehension interventions (n = 43) in U.S. elementary schools, the authors examined whether effects vary depending on participant and intervention characteristics. Findings suggest positive effects on custom measures of vocabulary, listening comprehension, and reading comprehension but not on standardized measures of these outcomes. Results also indicate positive effects for English learners and promise for multicomponent interventions and those that include technology. Much more research is needed on how best to support language comprehension for underserved populations (e.g., students from low-income backgrounds) and how interventions can be optimized to support generalizable language and literacy outcomes. Implications for policy and practice are discussed.

The debate over the science of reading has focused primarily on decoding (i.e., the ability to connect letters and sounds to read words) and whether to use explicit phonics to teach it. Yet, language comprehension (i.e., the ability to make meaning from oral or written language) is an equally critical component of reading that has received far less attention. Language comprehension can be acquired through exposure to language in the environment but also can be supported through explicit instruction. Despite studies showing positive effects of such instruction (National Reading Panel, 2000), research has suggested that language comprehension instruction is not robust in elementary schools (e.g., Sparapani, Carlisle, & Connor, 2018). Thus, to inform research, policy, and practice, in this article, we review the science of reading beyond decoding as it pertains to language comprehension instruction.

In discussions of the science of reading, researchers have often invoked the simple view of reading, which describes reading comprehension as a

INTERNATIONAL LITERACY ASSOCIATION

Reading Research Quarterly, 55(S1)
pp. S207–S233 | doi:10.1002/rrq.346

product of decoding and linguistic comprehension (Gough & Tunmer, 1986). Although other factors are also important (e.g., background knowledge, fluency, motivation), researchers have agreed that decoding and linguistic comprehension, also referred to as language comprehension or listening comprehension, are both essential for reading comprehension (Florit & Cain, 2011). Developmental research has suggested that the relative importance of language comprehension to reading comprehension increases over time as students become more efficient decoders and can better attend to meaning in text (Adlof, Catts, & Lee, 2010). Misinterpreting this research could lead to an early focus on decoding and a later focus on language comprehension. However, language comprehension, which develops cumulatively across the life span, must be a focus in early and later elementary school so students acquire the robust language skills that they will need to understand texts as they learn to decode them (Cunningham & Stanovich, 1997).

Language comprehension relies on vocabulary and semantic awareness, as well as knowledge of how morphology and syntax affect meaning (Scarborough, 2001). Facility with these skills predicts reading comprehension, and students who struggle with these skills tend to have difficulty with reading comprehension in elementary school and beyond (Adlof et al., 2010). Whereas some students acquire language comprehension without much intervention, other students require explicit support to develop the language comprehension skills needed to understand and learn from complex texts (Connor et al., 2011). Therefore, intervention research, which evaluates the effects of various approaches to instruction, has provided critical evidence to inform the science of reading as it pertains to language comprehension. Building on previous reviews of research, in this article, we report on the effects of language comprehension instruction in elementary school, with attention to whether effects differ by participant and intervention characteristics.

Until recently, syntheses of intervention research have focused on whether vocabulary intervention leads to gains in language comprehension and reading comprehension (Elleman, Lindo, Morphy, & Compton, 2009; Marulis & Neuman, 2010; Stahl & Fairbanks, 1986). Whereas these reviews showed that vocabulary intervention has positive effects on custom (i.e., researcher-developed) measures, the effects are not evident on standardized measures, suggesting that effects do not generalize. Other reviews have investigated morphology intervention with similar findings (Bowers, Kirby, & Deacon, 2010; Carlisle, 2010; Goodwin & Ahn, 2013). Researchers have acknowledged that vocabulary or morphology intervention alone is not enough, and understanding how the components of language work together may be key to more impactful intervention (Kieffer, Petscher, Proctor, & Silverman, 2016; Proctor, Silverman, Harring, Jones, & Hartranft, 2020; Wright & Cervetti, 2017). In this context, Rogde, Hagen, Melby-Lervåg, and Lervåg (2019) conducted a comprehensive review of the effects of linguistic comprehension interventions in preschool through secondary school that included attention to vocabulary and/or narrative or grammatical development on generalized measures of language and reading comprehension. In the meta-analysis, *generalized measures* referred to those that assess via content not specifically targeted in the intervention. The researchers found small effects on generalized linguistic comprehension (Hedges' $g = 0.16$, $p < .01$) and negligible effects on generalized reading comprehension ($g = 0.05$, $p = .13$). The findings of this review suggest that much more work is needed on how to best support language comprehension.

To inform research, policy, and practice, we built on Rogde et al.'s (2019) study by narrowing our focus to studies conducted in kindergarten through fifth-grade settings in the United States and published after 2010 when the Common Core State Standards, which include a strong focus on language, were introduced (National Governors Association Center for Best Practices & Council of Chief State School Officers, 2010). Whereas Rogde et al. disaggregated effects for grades K–5, they did not disaggregate the K–5 findings by study location or date. Because effects were greater for studies set in Europe, and the mean date of publication was 2010 (Rogde et al., 2019), we determined that a closer look at recent studies in U.S. elementary school settings was needed. Furthermore, given that interventions could be differentially effective for different populations of students (Connor et al., 2011) and vary according to intervention characteristics (Elleman et al., 2009), we aimed to understand the current research on how effects differ across participant and intervention characteristics.

Regarding how intervention effects could vary for different populations, there is some evidence that it is harder to effect change in language and reading comprehension as students progress through the grades (Pearson, Palincsar, Biancarosa, & Berman, 2020). As students grow and develop, they learn more and more about how language works, but that also means that any difficulties they may have with language comprehension become compounded over time. Thus, it is important to compare intervention effects across lower and upper elementary school. Additionally, there is some evidence that intervention effects may differ for English learners (ELs) versus non-ELs (Foorman, Herrera, & Dombek, 2018), as well as for students from higher or lower socioeconomic backgrounds or considered at risk or not at risk for experiencing reading difficulties (Marulis & Neuman, 2010). Exploring effects across these different participant populations is needed to determine how to best support diverse learners.

In previous reviews examining how intervention characteristics influence effects, duration and grouping strategies have been examined, although there is inconsistent evidence that either makes a difference (e.g., Elleman et al., 2009; Hall & Burns, 2018). Reviews have examined other instructional characteristics as well. For example, Bowers et al. (2010) found that interventions addressing multiple language components were more effective than those focusing on a single component of language, although Goodwin and Ahn (2013) did not find differential effects for multiple- or single-component interventions. For another example, Elleman et al. (2009) found that vocabulary interventions with higher levels of discussion were more effective than interventions with lower levels, but including strategy instruction in addition to or instead of explicit instruction on particular words did not make a difference. Although few reviews have examined the role of writing in language comprehension intervention, research on reciprocity across language, reading, and writing has suggested that it may be worthwhile to examine effects of interventions that include writing (Kim, Petscher, Wanzek, & Al Otaiba, 2018). Some research has examined combining language comprehension and content area instruction, but so far reviews have not disentangled whether this makes a difference (e.g., Wright & Cervetti, 2017). The review of vocabulary and comprehension interventions by the National Reading Panel (2000) suggested that interventions that include technology may be beneficial, but the researchers noted that much more research on this topic was needed. Furthermore, a review of preschool interventions identified that focusing on professional development may be a viable way to support language and literacy (Markussen-Brown et al., 2017), but the researchers did not summarize research in elementary school. Understanding how these intervention characteristics are related to effects on language and literacy outcomes is essential to inform research, policy, and practice on K–5 language comprehension instruction.

The Present Study

In an attempt to expand the focus of discussions on the science of reading to include attention to language comprehension and its instruction, our aim in this meta-analysis was to review the research base on language comprehension intervention and to examine how effects vary by participant and intervention characteristics. Three research questions guided this study:

1. What are the effects of language comprehension interventions on K–5 students' language and literacy outcomes?
2. Do these effects differ for particular populations of students?
3. Do these effects differ according to specific intervention characteristics?

Method

Literature Search and Study Selection Criteria

We used established guidelines (Cooper, Hedges, & Valentine, 2009) to obtain a corpus of studies eligible for inclusion.

Search Procedures

First, we conducted a systematic search using the ERIC and PsycINFO databases. We searched for peer-reviewed studies published between January 2010 and January 2020 using the following search terms: (*teaching methods* OR *intervention* OR *instruction*) AND (*oral language OR language skills* OR *listening comprehension* OR *semantics OR syntax* OR *morphology* (*languages*) OR *vocabulary OR grammar* OR *reading comprehension* OR *pragmatics* OR *language usage*) NOT (*higher education* OR *secondary education*). This search yielded 12,344 studies.

Review Procedures

We reviewed abstracts and full articles and included studies meeting these five criteria:

1. Studies were written in English, set in the United States, and published in a peer-reviewed journal.
2. Studies used quasi-experimental or experimental designs with at least one treatment and one control or comparison group. Effect sizes or means and standard deviations were reported by condition. Control or comparison conditions could include business as usual or an alternate treatment.
3. Studies included language comprehension (syntax, vocabulary/semantics, or morphology), listening comprehension, or reading comprehension outcome measures.
4. Studies were set in K–5 general education contexts (i.e., not special education, after school, or summer school) where English was the language of instruction.[1]
5. Studies included sustained language comprehension instruction, defined as five or more sessions and at least 50% of instruction targeting language comprehension.[2] Studies focused on instruction and not, for example, vocabulary acquisition or affordances of different media.

Based on this search, we identified 45 studies meeting the inclusion criteria. We combined data from initial and follow-up studies, resulting in 43 unique studies. Many of these studies evaluated multiple interventions or investigated effects for specific samples of students (e.g., students at different grade levels) separately.

Coding Procedures

We read full articles, recording sample sizes, means, and standard deviations (or reported effect sizes and standard errors), and coded studies for design, outcome, participant, and intervention characteristics. If information was not available in an article, we contacted the author(s) to try to obtain this information.

Designs

We coded for three types of experimental designs: (1) randomized control trials with at least two groups, (2) quasi-experimental designs with at least two groups, and (3) within-subjects designs with at least two conditions. We noted whether the control or comparison condition(s) were business as usual or an alternative treatment.

Outcomes

We coded for whether studies included measures of vocabulary or semantics, morphology, syntax, listening comprehension, decoding, or reading comprehension. We also coded for whether measures were custom or standardized. For custom measures, we further coded for whether measures were proximal (i.e., included words or passages directly taught in the intervention) or distal (i.e., included words or passages not directly taught in the intervention).

Participants

We coded for sample size, grade levels, and demographics such as race/ethnicity, income status as measured by the proportion of students receiving free or reduced-price lunch, language status (i.e., EL vs. non-EL), and whether students with disabilities were included. We also noted whether studies examined differences in effects by participant characteristics.

Intervention Characteristics

First, we coded for duration (total hours) and whether the intervention was conducted in whole-class, small-group, partner, or one-on-one contexts. Next, we coded for whether interventions were focused on one component of language (e.g., morphology) or multiple components. We also coded for which specific aspects of language comprehension were addressed (i.e., vocabulary or semantics, morphology, syntax) and whether interventions also included attention to decoding-related skills (e.g., phonological awareness, phonics) or general listening or reading comprehension skills (e.g., questioning, retelling). Additionally, we coded for whether interventions included strategy instruction, discussion, or writing and whether interventions incorporated content area instruction or technology. We also noted whether the primary focus of the intervention was professional development for teachers.

Reliability

The first three authors coded abstracts and articles. During the abstract review, three rounds of double or triple coding took place with adjudication. We coded 10% of the abstracts with over 90% reliability and then coded the remainder of the abstract corpus individually. In the article review stage, 75% of the studies were double-coded with 95% agreement. All disagreements were adjudicated until 100% agreement was reached across the whole team.

Analytic Procedures

We calculated effect sizes for each study as standardized mean differences (i.e., the difference between the treatment and control group means on an outcome variable, divided by the pooled standard deviations). Because effect sizes from small samples can be biased, we adjusted each effect size with a small sample correction, $[1 - (3/(4(n_T + n_C) - 9)]$, resulting in Hedges' unbiased effect sizes (Hedges, 1982).

$$\text{Effect Size} = \frac{(m_T - m_C)}{\sqrt{\frac{(n_T-1)S_T^2+(n_C-1)S_C^2}{n_T+n_C-2}}}\left(1-\frac{3}{4(n_T+n_C)-9}\right)$$

For studies detailing pretest mean scores and standard deviations, we calculated the effect sizes as the mean pre–post change in the treatment group minus the mean pre–post change in the control group, divided by the pooled pretest standard deviation (S.B. Morris, 2008). This information was not available for Gersten, Dimino, Jayanthi, Kim, and Santoro's (2010) and Jayanthi et al.'s (2018) studies, so we used the reported effect sizes. Note that in many cases, the researchers used a different method for deriving effect sizes and/or controlled for other variables in deriving them. Thus, effect sizes in this meta-analysis may not match reporting in the original studies.

When a study reported multiple outcomes, we aggregated the effect sizes. This was essential to uphold the assumption of statistical independence. Ignoring these dependencies can result in biased estimates, as well as spurious heavier weights to studies reporting multiple outcomes. Prior meta-analyses have either selected a single effect size from those available or used simple unweighted means for aggregation. We instead considered within-study correlations while aggregating effect sizes (Borenstein, 2009) and conducted sensitivity analyses for within-study correlations (r = .2–.8) before deciding on one effect size for each study.

We synthesized effect sizes for each construct based on a random-effects model using the restricted maximum likelihood method because differences in methods and sample characteristics between studies were likely introducing variability in the true effect sizes across studies. We then conducted a test of homogeneity for each synthesis using the *Q*-statistic to ascertain whether the variability in effect sizes occurred for reasons beyond sampling error (Hedges & Olkin, 1985). For vocabulary, listening comprehension, and reading comprehension, we synthesized effect sizes for custom and standardized measures combined and separately.

To examine whether effects differed by participant and intervention characteristics, we conducted moderator analyses on outcomes for which there was high heterogeneity (Higgins & Green, 2008). We used mixed-effects models or random-effects metaregressions to model moderating effects. The results of metaregressions are reported as slopes (along with their statistical significance). A slope is interpreted as the change in standardized mean difference when the variable under consideration changes by one unit. Because the moderator analysis was exploratory in nature, we considered any effects at $p < .10$ to be statistically significant (Atal, Porcher, Boutron, & Ravaud, 2019).

Results

Although design characteristics were not central to our research questions, we investigated these and found that most studies in our corpus were randomized control trials (61%, $n = 27$) and that half (50%, $n = 22$) used business as usual versus an alternative treatment. Results showed no differences in effects across randomized control trials and quasi-experimental designs, although for vocabulary, effects were higher for within-subjects designs. There were no differences by whether control/comparison conditions were business as usual or an alternative treatment for vocabulary and listening comprehension, although unexpectedly, effects were higher for studies with an alternative treatment rather than business as usual for reading comprehension. Because this was not central to our research questions, we did not unpack this finding, but future meta-analyses should consider this further. Below we summarize results for each of our research questions. See Tables 1–3 for overviews of design, outcome, participant, and intervention characteristics and see Tables 4–5 for effect size estimates.

Effects of Language Comprehension Interventions

A range of outcome measures was identified in our corpus. Most studies ($n = 38$) included a vocabulary outcome, whereas only two studies included a morphology outcome, and only one study included a syntax outcome. Two studies included a relatively new measure of a construct referred to as *academic language*. Many studies included either listening comprehension ($n = 10$) or reading comprehension ($n = 16$) outcomes or both ($n = 5$), and several studies ($n = 7$) included a decoding measure. Notably, all custom vocabulary measures were proximal, and all custom listening comprehension measures were distal.

We examined effects of language comprehension interventions across measures. Intervention effects were large and statistically significant for vocabulary ($g = 0.85$, $p < .01$), but this effect was seen on custom measures ($g = 1.27$, $p < .01$), not standardized measures ($g = 0.03$, $p = 0.17$; see Figure 1). Although effects were modest, the trend of positive effects was consistent across listening comprehension ($g = 0.10$, $p < .01$) and reading comprehension ($g = 0.19$, $p < .01$); however, again, statistically significant effects were found on custom measures ($g = 0.19$ and 0.68, respectively) but not on standardized measures ($g = 0.03$ and 0.08, respectively; see Figures 2 and 3). Additionally, positive and statistically significant effects were found on morphology ($g = 1.14$, $p < .01$) and academic language ($g = 0.08$, $p = .04$), but given the limited number of studies, including these outcome results should be considered with caution. No effects were seen on syntax ($g = 0.01$, $p = .99$) or decoding ($g = 0.05$, $p = .73$). Delayed effects were positive and statistically significant on vocabulary ($g = 0.73$, $p < .01$) and reading comprehension ($g = 0.35$, $p < .01$), although few studies examined delayed effects. Because there were only a few studies with morphology, syntax, academic language, and decoding, we did not examine moderators for these outcomes. Heterogeneity was not statistically significant for listening comprehension, so we did not explore moderators for this outcome either. For vocabulary and reading comprehension, heterogeneity was statistically significant, so we conducted moderator analyses for these outcomes.[3]

Effects by Participant Characteristics

In total, 37,149 participants were studied, with sample sizes ranging from 10 to 16,471. A disproportionate number of studies focused solely on grades K–2 ($n = 29$). Across studies that reported race/ethnicity data, the samples were diverse (38% Caucasian, 35% Hispanic, 24% African American, 4% Asian, and 5% other), but 42% of the studies did not report these data. The majority of studies (59%, $n = 26$) included high or mid-high poverty samples as measured via students' eligibility for free or reduced-price lunch. Several studies ($n = 16$) included ELs, and 10 studies had entirely EL samples. Finally, 13 studies included students with disabilities, and 11 studies specifically focused on students considered at risk. Approximately one third of the studies investigated whether effects differed according to participant characteristics (i.e., initial level, EL status, risk status).

Results suggest that effects across studies did not differ according to whether the study focused on grades

TABLE 1
Design and Outcome Characteristics

Study	Design	Control(s)	Proximal custom measure(s)	Distal custom measure	Standardized measure(s)
Apel, Brimo, Diehm, and Apel (2013)	RCT	BAU		M	D, R
Apthorp et al. (2012)	RCT	BAU	R, V		V
Arthur and Davis (2016)	QED	BAU	V		
August, Artzi, Barr, and Francis (2018)	WSD	ALT	V		
Baker et al. (2013)	RCT	ALT	V	L	L
Brimo (2016)	QED	BAU	M		
Connor et al. (2018)	RCT	BAU			D, L, R, S, V
Connor et al. (2019)	RCT	ALT, BAU	V		
Coyne et al. (2010)	QED	BAU	V		V
Coyne et al. (2019)	RCT	ALT	V		V
Crevecoeur, Coyne, and McCoach (2014)[a]	QED	BAU	V		V
Dalton, Proctor, Uccelli, Mo, and Snow (2011)	RCT	ALT	R, V		R, V
Daunic et al. (2013)	QED	BAU			R, V
Filippini, Gerber, and Leafstedt (2012)	RCT	ALT	V		D
Gersten, Dimino, Jayanthi, Kim, and Santoro (2010)	RCT	BAU			V
Goldstein et al. (2017)	RCT	ALT			
Graham, Graham, and West (2015)	RCT	BAU	R, V	R	
Hassinger-Das, Jordan, and Dyson (2015)	RCT	ALT, BAU	V		V
Huang (2015)	QED	ALT			V
Jayanthi et al. (2018)	RCT	BAU			R, V
Jones et al. (2019)	RCT	BAU	V		A, R
Language and Reading Research Consortium, Jiang, and Davis (2017)	RCT	BAU	V	L	
Mancilla-Martinez (2010)	QED	BAU	V		
McKeown and Beck (2014)	WSD	ALT	V		
Morris et al. (2012)	RCT	ALT			D, R
Nelson, Vadasy, and Sanders (2011)	RCT	ALT	V		D, V
Neuman and Kaefer (2018)	RCT	BAU	V		V
Nielsen and Friesen (2012)	QED	BAU	V		L, V
Powell and Driver (2015)	RCT	ALT, BAU	V		
Proctor et al. (2011)	QED	BAU	V		R, V
Proctor, Silverman, Harring, Jones, and Hartranft (2020)	QED	BAU			A, R
Puhalla (2011)	RCT	ALT	V		
Pullen, Tuckwiller, Konold, Maynard, and Coyne (2010)	QED	ALT	V		

(continued)

TABLE 1
Design and Outcome Characteristics (*continued*)

Study	Design	Control(s)	Proximal custom measure(s)	Distal custom measure	Standardized measure(s)
Silverman, Kim, Hartranft, Nunn, and McNeish (2017)	QED	BAU	V		R, V
Silverman, Martin-Beltran, et al. (2017)	QED	BAU	V	L	R, V
Simmons et al. (2010)	RCT	ALT, BAU	V		R, V
Tong, Irby, Lara-Alecio, Yoon, and Mathes (2010)	RCT	BAU			L, R, V
Vadasy, Nelson, and Sanders (2013)[b]	RCT	ALT	V		V
Vadasy and Sanders (2015)	RCT	ALT	V		
Vadasy and Sanders (2016)	RCT	ALT	V		D, V
Vadasy, Sanders, and Logan Herrera (2015)	RCT	BAU	V		R, V
Vadasy, Sanders, and Nelson (2015)	RCT	ALT	V		D, V
Wood et al. (2018)	RCT	ALT	V		V
Wright and Gotwals (2017)	QED	BAU	V		
Zipoli, Coyne, and McCoach (2011)	WSD	ALT	V		

Note. A = academic language; ALT = alternative treatment comparison group; BAU = business-as-usual control group; D = decoding; L = listening comprehension; M = morphology; QED = quasi-experimental design; R = reading comprehension; RCT = randomized control trial; S = syntax; V = vocabulary; WSD = within-subjects design.
[a]Follow-up study of Coyne et al. (2010). [b]Follow-up study of Nelson et al. (2011).

K–2, 3–5, or both. Given the spotty reporting of race/ethnicity and whether students with disabilities were included in the sample, we were unable to examine whether effects differed for these populations. To examine whether effects differed by income status, we compared effects for studies with mid-high or high proportions of students receiving free or reduced-price lunch versus studies with mid-low or low proportions. Findings suggest that studies with higher proportions of students from low-income families tended to have lower effects on vocabulary outcomes ($g = -0.77$, $p = .04$). We also compared effects for studies with high proportions of ELs (≥25%) versus low proportions (<25%). We found no difference across outcomes, but we also conducted a secondary analysis of studies that included both ELs and non-ELs and presented means for both populations ($n = 5$). For vocabulary, effects were 0.65 ($p < .01$) for non-ELs and 0.72 ($p < .01$) for ELs, and for reading comprehension, effects were 0.05 ($p > .05$) for non-ELs and 0.55 ($p < .05$) for ELs. Although we were unable to compare effects of studies that did or did not include students with disabilities, we examined effects for the few studies that included samples in which all students were considered at risk ($n = 4$). These studies only included custom vocabulary outcomes. The effect size for these studies ($g = 0.77$, $p < .01$) was slightly smaller than the effect size across all studies ($g = 0.85$, $p < .01$).

Effects by Intervention Characteristics

There was substantial variation across studies on a number of intervention characteristics. Duration (hours spent in intervention) ranged from 1 to 100, with a mean of 27 and a median of 20. There were almost equal numbers of whole-class instruction ($n = 18$) versus small-group, partner, or one-on-one instruction ($n = 19$). All but two interventions included attention to vocabulary. In contrast, only 14 included morphology instruction, and only seven included syntax instruction. A majority of the studies ($n = 35$) were multicomponent. Many studies ($n = 13$) included instruction attending to other aspects of comprehension (e.g., comprehension strategies), and several ($n = 8$) included decoding-related instruction. Whereas a number of studies included strategy instruction ($n = 8$) or discussion ($n = 8$), only a few ($n = 4$) included writing. Multiple studies ($n = 14$) were set within non–English language arts contexts, and many studies ($n = 11$) included technology. In only three studies was the intervention primarily focused on professional development.

In analyses exploring whether effects differed according to the above characteristics, differential effects were rare, likely because of limited power or heterogeneity in the sample. For vocabulary, interventions that were primarily whole group were more effective than those that were small group, partner, or one-on-one ($g = 0.76$, $p < .01$), although this could be confounded by the fact that many interventions with small-group, partner, or

TABLE 2
Participant Characteristics

Study	Total sample size and *n* by group[a]	Grade level(s)	Percentage Caucasian	Percentage African American	Percentage Hispanic	Percentage Asian	Percentage other	Proportion receiving FRPL[b]	Percentage EL[c]	Included students with disabilities	Subgroup analysis
Apel, Brimo, Diehm, and Apel (2013)	Kindergarten: 43 Grade 1: 54 Grade 2: 54	K–2	16	75	3	1	5	Mid-high	0	Yes	
Apthorp et al. (2012)	Kindergarten: 3,119 Grade 1: 3,445 Grade 3: 3,316	K, 1, 3[d]	NR	NR	NR	NR	NR	Mid-high	NR	NR	
Arthur and Davis (2016)	Kindergarten: 55 Grade 1: 52 Grade 2: 46 Grade 3: 60	K–3	82	7	NR	5	6	NR	NR	Yes	Grade
August, Artzi, Barr, and Francis (2018)	187	2	NR	NR	NR	NR	NR	High	100	NR	
Baker et al. (2013)	225	1	NR	NR	NR	NR	NR	Mid-high	NR	NR	Risk status
Brimo (2016)	10	3	NR	NR	NR	NR	NR	NR	0	Yes	
Connor et al. (2018)	Grade 3: 401 Grade 4: 225	3, 4	NR	NR	NR	NR	NR	Mid-low	NR	NR	Initial level
Connor et al. (2019)	603	3–5	10	9	67	NR	NR	Mid-high	NR	NR	
Coyne et al. (2010)	124	K	19	21	56	NR	4	High	39	NR	Initial level
Coyne et al. (2019)	2,347	K	22	19	34	NR	10	NR	NR	NR	Initial level
Crevecoeur, Coyne, and McCoach (2014)	122	K	NR	NR	NR	NR	NR	High	39	NR	EL status
Dalton, Proctor, Uccelli, Mo, and Snow (2011)	106	5	NR	NR	NR	NR	NR	Mid-low	36 (BIL)	NR	EL status
Daunic et al. (2013)	57	K	33	33	5	NR	NR	NR	0	[e]	
Filippini, Gerber, and Leafstedt (2012)	71	1	NR	NR	NR	NR	NR	High	87	NR	Initial level
Gersten, Dimino, Jayanthi, Kim, and Santoro (2010)	468	1	NR	NR	NR	NR	NR	NR	24 (LM)	NR	
Goldstein et al. (2017)	241	1–3	15	75	5	NR	5	High	<1 (LEP)	Yes	Initial level

(continued)

TABLE 2
Participant Characteristics (*continued*)

Study	Total sample size and *n* by group[a]	Grade level(s)	Percentage Caucasian	Percentage African American	Percentage Hispanic	Percentage Asian	Percentage other	Proportion receiving FRPL[b]	Percentage EL[c]	Included students with disabilities	Subgroup analysis
Graham, Graham, and West (2015)	375	4	74	18	6	1	1	NR	NR	NR	
Hassinger-Das, Jordan, and Dyson (2015)	124	K	18	18	63	NR	2	NR	55	NR	
Huang (2015)	40	2	68	10	20	3	0	High	NR	NR	
Jayanthi et al. (2018)	1,680	1	42	11	38	NR	9	Mid-high	25 (LEP)	NR	Initial level
Jones et al. (2019)	Year 1: 2,808 Year 2: 2,558	Year 1: 4, 5 Year 2: 4, 5	Year 1: 30 Year 2: 29	Year 1: 42 Year 2: 41	Year 1 21 Year 2: 24	Year 1: 3 Year 2: 2	Year 1: 3 Year 2: 3	High	Year 1: 9 Year 2: 7	Yes	
Language and Reading Research Consortium, Jiang, and Davis (2017)	Kindergarten: 155 Grade 1: 139 Grade 2: 155 Grade 3: 150	K–3	86	8	12	4	2	Low	NR	Yes	
Mancilla-Martinez (2010)	49	5	6	2	90	2	NR	High	92 (LM)	NR	
McKeown and Beck (2014)	131	K	75	25	NR	NR	NR	Mid-low	0	NR	
Morris et al. (2012)	279	2, 3	48	45	NR	NR	NR	Low	0	Yes	
Nelson, Vadasy, and Sanders (2011)	185	K	NR	NR	NR	NR	NR	NR	100	NR	
Neuman and Kaefer (2018)	265	K	2	24	38	5	17	Mid-high and high	14	Yes	EL status
Nielsen and Friesen (2012)	28	K	11	71	11	7	NR	High	18	NR	
Powell and Driver (2015)	98	1	59	23	14	NR	4	Mid-low	<1	Yes	
Proctor et al. (2011)	240	5	NR	NR	NR	NR	NR	Mid-high	49 (BIL)	NR	
Proctor, Silverman, Harring, Jones, and Hartranft (2020)	239	4, 5	NR	NR	NR	NR	NR	Mid-high	100	NR	Initial level
Puhalla (2011)	66	1	94	5	1	0	0	Low	0	No	

(continued)

TABLE 2
Participant Characteristics (*continued*)

Study	Total sample size and *n* by group[a]	Grade level(s)	Percentage Caucasian	Percentage African American	Percentage Hispanic	Percentage Asian	Percentage other	Proportion receiving FRPL[b]	Percentage EL[c]	Included students with disabilities	Subgroup analysis
Pullen, Tuckwiller, Konold, Maynard, and Coyne (2010)	224	1	39	NR	NR	NR	NR	Mid-low	NR	Yes	Risk status
Silverman, Kim, Hartranft, Nunn, and McNeish (2017)	Kindergarten: 504 Grade 4: 534	K, 4	62	30	NR	NR	7	Mid-high	8	NR	Initial level
Silverman, Martin-Beltran, et al. (2017)	Kindergarten: 196 Grade 4: 239	K, 4	4	25	63	NR	5	High	34	NR	EL status
Simmons et al. (2010)	903	4	13	16	70	NR	1	Mid-high	NR	NR	
Tong, Irby, Lara-Alecio, Yoon, and Mathes (2010)	196	K–2	0	0	100	0	0	High	100	NR	Gender
Vadasy, Nelson, and Sanders (2013)	140	1	NR	NR	NR	NR	NR	NR	100	NR	
Vadasy and Sanders (2015)	69	K	NR	NR	NR	NR	NR	NR	100	Yes	Initial level
Vadasy and Sanders (2016)	100	K	NR	NR	NR	NR	NR	Mid-high	100	Yes	Initial level
Vadasy, Sanders, and Logan Herrera (2015)	1,232	4, 5	66	11	3	14	5	Mid-low	100	Yes	Initial level
Vadasy, Sanders, and Nelson (2015)	324	K	NR	NR	NR	NR	NR	Mid-high	100	Yes	Initial level
Wood et al. (2018)	288	K, 1	NR	NR	NR	NR	NR	High	100	No	Initial level
Wright and Gotwals (2017)	147	K	27	NR	NR	NR	NR	Mid-high	NR	NR	
Zipoli, Coyne, and McCoach (2011)	80	K	NR	NR	NR	NR	NR	High	NR	[e]	
Mean across studies	**826**		**38**	**24**	**35**	**4**	**5**		**48**		

Note. EL = English learner; FRPL = free or reduced-price lunch; NR = not reported.
[a]In this table, we indicate the sample sizes used in analysis by grade level or year; these may differ from the total sample sizes listed in the articles because, in some cases, there were different sample sizes by grade level, year, and outcome. [b]Using National Center for Education Statistics standards for proportion FRPL, low = 25.0% or fewer of the students are eligible, mid-low = 25.1–50.0%, mid-high = 50.1–75.0%, and high = more than 75.0%. [c]In this article, we use the term *English learner* to refer to students whose home language is not English only; some studies referred to these students as bilingual (BIL), language-minority learners (LMs), or limited English proficient (LEP), and we indicated in parentheses when a different term was used in the study. [d]Apthorp et al. included grade 4 as well, but data on this group were not provided in the article, and we were unable to get these data from the authors. [e]Excluded at least some students with severe disabilities.

TABLE 3
Intervention Characteristics

Study	Number of hours	Instructional group size	Language comprehension components	Decoding-related instruction	Other comprehension instruction	Strategy instruction	Discussion	Writing	Non–English language arts content focus	Technology	Professional development
Apel, Brimo, Diehm, and Apel (2013)	13	Small group	M								
Apthorp et al. (2012)	20–40	Whole class	V								
Arthur and Davis (2016)	6–12	Whole class	S, V		+						
August, Artzi, Barr, and Francis (2018)	25	Whole class	M, V								
Baker et al. (2013)	38	Whole class	V		+						
Brimo (2016)	12.5	Small group	M								
Connor et al. (2018)	20–24	Small group	T1: V T2: V T3: S, V T4: V		+				Science		
Connor et al. (2019)	4.5	Small group, one-on-one	T1: M, V T2: M, V			+	+			+	
Coyne et al. (2010)	18	Whole class, small group	V								
Coyne et al. (2019)	44	Small group	V								
Crevecoeur, Coyne, and McCoach (2014)	18	Whole class, small group	V								
Dalton, Proctor, Uccelli, Mo, and Snow (2011)	20	One-on-one	T1: M. V T2: M, V		+					+	
Daunic et al. (2013)	4–5	Small group	V		+		+				
Filippini, Gerber, and Leafstedt (2012)	7.25	Small group	T1: M, V T2: V	+							

(continued)

TABLE 3
Intervention Characteristics (*continued*)

Study	Number of hours	Instructional group size	Language comprehension components	Decoding-related instruction	Other comprehension instruction	Strategy instruction	Discussion	Writing	Non–English language arts content focus	Technology	Professional development
Gersten, Dimino, Jayanthi, Kim, and Santoro (2010)	NR	Whole class	V								+
Goldstein et al. (2017)	72	One-on-one	V			+				+	
Graham, Graham, and West (2015)	9	Whole class	V			+			Social studies		
Hassinger-Das, Jordan, and Dyson (2015)	12	Small group	V						Math		
Huang (2015)	80	Whole class	S, V							+	
Jayanthi et al. (2018)	NR	Whole class	V			+					+
Jones et al. (2019)	95	Whole class, small group	V				+		Multiple		
Language and Reading Research Consortium, Jiang, and Davis (2017)	50	Whole class	M, S, V		+					+	
Mancilla-Martinez (2010)	25	Whole class	V				+	+	Multiple		
McKeown and Beck (2014)	2.5	Whole class	V								
Morris et al. (2012)	70	Small group	M, S, V	+		+					
Nelson, Vadasy, and Sanders (2011)	33	Small group	M, V	+							
Neuman and Kaefer (2018)	20	Whole class	V						Science	+	
Nielsen and Friesen (2012)	18	Small group	V		+						

(*continued*)

TABLE 3
Intervention Characteristics (*continued*)

Study	Number of hours	Instructional group size	Language comprehension components	Decoding-related instruction	Other comprehension instruction	Strategy instruction	Discussion	Writing	Non–English language arts content focus	Technology	Professional development
Powell and Driver (2015)	2.5–3.75	One-on-one	V						Math		
Proctor et al. (2011)	26.67	One-on-one	M, S, V		+					+	
Proctor, Silverman, Harring, Jones, and Hartranft (2020)	20	Small group	M, S, V		+		+	+	Multiple	+	
Puhalla (2011)	5	Whole class, small group	V						Science		
Pullen, Tuckwiller, Konold, Maynard, and Coyne (2010)	<2	Whole class, small group	V								
Silverman, Kim, Hartranft, Nunn, and McNeish (2017)	10	Whole class, partner	V			+	+		STEM	+	
Silverman, Martin-Beltran, et al. (2017)	20	Whole class, partner	M, V		+	+	+		STEM	+	
Simmons et al. (2010)	27	Whole class	M, V			+			Social studies		
Tong, Irby, Lara-Alecio, Yoon, and Mathes (2010)	>100[a]	Whole class, small group	V	+	+				Multiple		+
Vadasy, Nelson, and Sanders (2013)	33	Small group	V	+							
Vadasy and Sanders (2015)	0.5–1	One-on-one	V	+							
Vadasy and Sanders (2016)	14	One-on-one	V	+							

(continued)

TABLE 3
Intervention Characteristics (*continued*)

Study	Number of hours	Instructional group size	Language comprehension components	Decoding-related instruction	Other comprehension instruction	Strategy instruction	Discussion	Writing	Non–English language arts content focus	Technology	Professional development
Vadasy, Sanders, and Logan Herrera (2015)	35	Whole class	V		+						
Vadasy, Sanders, and Nelson (2015)	40	Small group	V	+							
Wood et al. (2018)	12.5–25	Small group, one-on-one	M, V							+	
Wright and Gotwals (2017)	30	Whole class	V		+		+	+	Science		
Zipoli, Coyne, and McCoach (2011)	18	Whole class, small group	V								

Note. M = morphology; NR = not reported; S = syntax; STEM = science, technology, engineering, and mathematics; T1 = treatment 1; T2 = treatment 2; T3 = treatment 3; T4 = treatment 4; V = vocabulary.
[a]The intervention lasted multiple years, and the article provided only an approximate indication of total hours.

one-on-one instruction targeted at-risk students. Duration did not surface as a predictor of effects. Results showed that multicomponent interventions tended to have higher effects than single-component interventions on vocabulary (g = 0.50, p = .07), interventions that included morphology (mostly in addition to vocabulary) tended to have higher effects on vocabulary (g = 0.66, p < .01), and interventions that included syntax (mostly in addition to vocabulary) tended to have higher effects on reading comprehension (g = 0.36, p = .03). For reading comprehension, there was also an indication that incorporating technology may be beneficial (g = 0.31, p = .04). Differences in effects were not detected for other characteristics.

Bias and Sensitivity Analyses

For each included study, we assessed the risk of bias along five dimensions: selection, performance, detection, attrition, and reporting bias. Studies were classified as high, low, or unclear risk on each of these dimensions (Higgins & Green, 2008). The overall risk of bias along each dimension was calculated by aggregating across studies and weighting each study by the inverse of its effect size variance. We also conducted sensitivity analyses in three ways to check the robustness of synthesized effect sizes: excluding studies with a high selection bias risk, constraining outlier effect sizes, and resynthesizing effect sizes using a correlated effects metaregression model that allows dependence between effect sizes. The results in Table 6 show that both the effect size magnitude and its statistical significance for each approach broadly mirror the trends outlined in the Results section.

Further, we explored the possibility of publication bias. If studies reporting strong effects are more likely to be published, the estimated mean population effects will be upward biased. Studies with smaller samples are at a greater risk of producing statistically nonsignificant results and, thus, being rejected for publication. Thus, we assessed small sample bias by plotting the observed effect sizes versus the standard errors. We plotted all independent effects residualized for variance attributable to measurement type (custom vs. standardized), study components (single vs. multiple), and participant grouping (whole class vs. small group, partner, or individual). A visual analysis of the funnel plots reveals that most estimates lie within prediction limits on the basis of sampling error (see Figures 4–6). Two studies for vocabulary (McKeown & Beck, 2014; Simmons et al., 2010) and one study for reading comprehension (Dalton, Proctor, Uccelli, Mo, & Snow, 2011) showed larger positive effects than expected. We accounted for these aberrations using sensitivity analyses involving resynthesizing effect sizes after Winsorizing such outliers. We also confirmed the presence of symmetry in each funnel plot through Egger's tests (Egger, Smith, Schneider, & Minder, 1997). The Egger's test results showed that the intercept values fall within the 95%

TABLE 4
Effects Across Language and Literacy Outcomes

Outcomes	k	Hedges' g and 95% confidence interval	p	Heterogeneity
Immediate outcomes				
Vocabulary (combined)	68	0.85, [0.58, 1.12]	<.01	Q = 2,339.94; df = 67; p < .01; T^2 = 1.26; I^2 = 99.28
Vocabulary (custom)	44	1.27, [0.90, 1.64]	<.01	Q = 1,578.70; df = 43; p < .01; T^2 = 1.51; I^2 = 99.33
Vocabulary (standardized)	24	0.03, [−0.01, 0.08]	.17	Q = 31.82; df = 23; p = .10; T^2 < 0.01; I^2 = 15.87
Listening comprehension (combined)	12	0.10, [0.02, 0.18]	.02	Q = 12.65; df = 11; p = .32; T^2 < 0.01; I^2 = 21.15
Listening comprehension (custom)	10	0.19, [0.05, 0.32]	<.01	Q = 5.50; df = 5; p = .36; T^2 < 0.01; I^2 = 10.50
Listening comprehension (standardized)	6	0.03, [−0.03, 0.09]	.36	Q = 2.47; df = 5; p = .78; T^2 = 0; I^2 = 0
Reading comprehension (combined)	20	0.19, [0.04, 0.34]	.01	Q = 102.09; df = 19; p < .01; T^2 = 0.09; I^2 = 90.52
Reading comprehension (custom)	3	0.68, [0.13, 1.23]	.02	Q = 22.36; df = 2; p < .01; T^2 = 0.21; I^2 = 90.86
Reading comprehension (standardized)	17	0.08, [−0.00, 0.17]	.06	Q = 44.74; df = 16; p < .01; T^2 = 0.02; I^2 = 64.68
Morphology (custom)	4	1.14, [0.37, 1.90]	<.01	Q = 10.43; df = 3; p = .02; T^2 = 0.44; I^2 = 77.35
Syntax (standardized)	2	0.01, [−0.14, 0.17]	.88	Q = .77; df = 1; p = .38; T^2 = 0; I^2 = 0
Academic language (standardized)	3	0.08, [0.00, 0.16]	.04	Q = 2.25; df = 2; p = .32; T^2 = <0.01; I^2 = 30.44
Decoding (standardized)	10	0.05, [−0.24, 0.35]	.73	Q = 42.35; df = 9; p < .01; T^2 = 0.17; I^2 = 82.93
Delayed outcomes				
Vocabulary (combined)	9	0.73, [0.19, 1.28]	<.01	Q = 82.08; df = 8; p < .01; T^2 = 0.60; I^2 = 92.99
Vocabulary (custom)	6	1.05, [0.27, 1.83]	<.01	Q = 47.86; df = 5; p < .01; T^2 = 0.82; I^2 = 92.05
Vocabulary (standardized)	3	0.17, [−0.15, 0.48]	.31	Q = 4.85; df = 2; p = .09; T^2 = 0.05; I^2 = 63.83
Reading comprehension (combined)	2	0.35, [0.15, 0.55]	<.01	Q = .03; df = 1; p = .85; T^2 = 0; I^2 = 0

confidence interval of expected values for vocabulary (p = .12), listening comprehension (p = .19), and reading comprehension (p = .63), suggesting no upward biasing in the current sample of effects. We further conducted a *p*-curve analysis that addressed any biases induced on account of *p*-hacking, which is a relatively novel approach for addressing publication bias (Simonsohn, Nelson, & Simmons, 2014). The *p*-curves for vocabulary and reading comprehension showed a significant right skew, which suggests the presence of true intervention effects for both constructs[4] (see Figures 7 and 8).

Discussion

In the debate over the science of reading, there is agreement that language comprehension is important to reading comprehension. Yet, there has been little discussion of how to support language comprehension in elementary school, when language comprehension competes with decoding for attention. To inform research, policy, and practice, discussions about the science of reading should include attention to language comprehension instruction. To forward that discussion, for the present meta-analysis, we reviewed studies on the effects of explicit language comprehension instruction in K–5 settings.

Intervention Effects

First, we investigated the effects of K–5 language comprehension interventions on a variety of language and literacy outcomes. Consistent with previous reviews (Elleman et al., 2009; Rogde et al., 2019), our

TABLE 5
Effects by Design, Outcome, Participant, and Intervention Characteristics

Characteristics		Vocabulary (combined)				Reading comprehension (combined)			
		k	β	Standard error	*p*	*k*	β	Standard error	*p*
Design characteristics									
Design	Quasi-experimental design	68	−0.24	0.29	.41	20	−0.07	0.17	.68
	Within-subjects design		**1.50**	**0.65**	**.02**			-	
	Randomized control trial								
Control or comparison	Alternative treatment comparison group	68	−0.01	0.30	.96	20	**0.65**	**0.18**	**<.01**
	Business-as-usual control group								
Outcome characteristics									
	Custom	68	**1.16**	**0.25**	**<.01**	20	**0.56**	**0.17**	**<.01**
	Standardized								
	Delayed	9	**0.73**	**0.28**	**<.01**	2	**0.35**	**0.10**	**<.01**
Participant characteristics									
Grade level	K–2	68	−0.19	0.30	.76	20	−0.15	0.21	.47
	Both		0.90	1.17	.44		0.02	0.378	.96
	3–5								
Income status	Lower income	54	**−0.77**	**0.38**	**.04**	17	−0.19	0.16	.28
	Higher income								
Language status	English learner	68	−0.46	0.28	.11	20	0.17	0.16	.30
	Non–English learner								

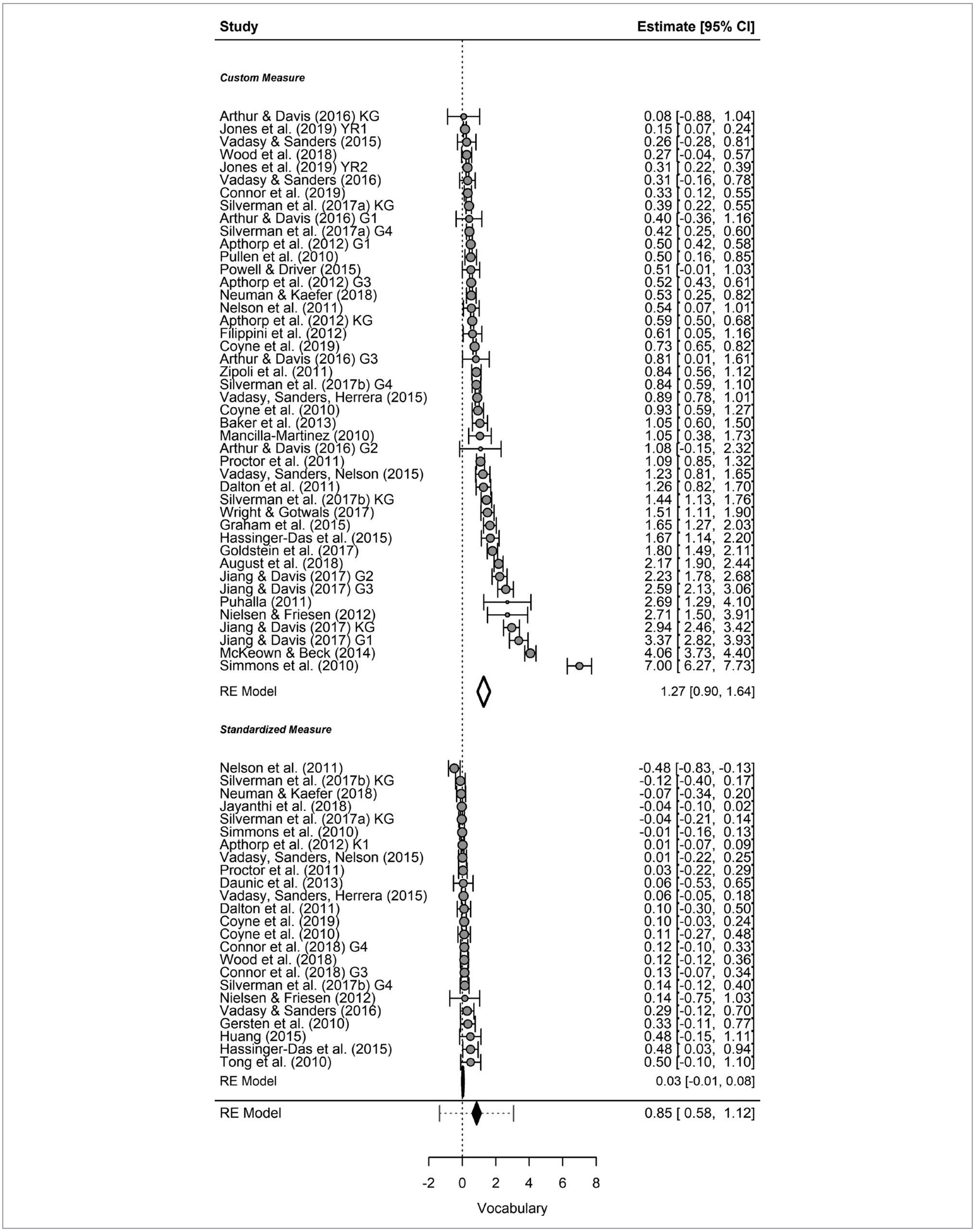
(continued)

TABLE 5
Effects by Design, Outcome, Participant, and Intervention Characteristics (*continued*)

Characteristics		Vocabulary (combined)				Reading comprehension (combined)			
		k	ß	Standard error	*p*	*k*	ß	Standard error	*p*
Intervention characteristics									
Duration	≥20 hours	68	−0.00	0.29	.99	20	−0.02	0.18	.90
	<20 hours								
	Hours	66	0.00	0.01	.80	20	−0.00	.00	.77
Grouping	Whole class	68	**0.76**	**0.27**	**<.01**	20	−0.17	0.18	.38
	Other								
Scope	Multicomponent	*68*	*0.50*	*0.28*	*.07*	20	0.16	0.16	.34
	Morphology	68	**0.66**	**0.16**	**<.01**	20	0.13	0.16	.42
	Syntax	68	0.53	0.35	.13	**20**	**0.36**	**0.17**	**.03**
	Phonological awareness or decoding	68	−0.56	0.41	.17	20	0.01	.29	.97
	Other comprehension	68	0.18	0.29	.54	**20**	**−0.30**	**.14**	**.04**
Instructional components	Strategies	68	0.47	0.39	.23	20	−0.18	.19	.33
	Discussion	68	−0.42	0.36	.25	20	−0.03	.16	.86
	Writing	68	0.07	0.59	.91	20	−0.11	.26	.68
	Technology	68	0.08	0.30	.79	**20**	**0.31**	**.15**	**.04**
	Content	68	0.24	0.31	.44	20	−0.20	.15	.20
	Professional development	68	−0.62	0.68	.36	20	−0.06	.29	.95

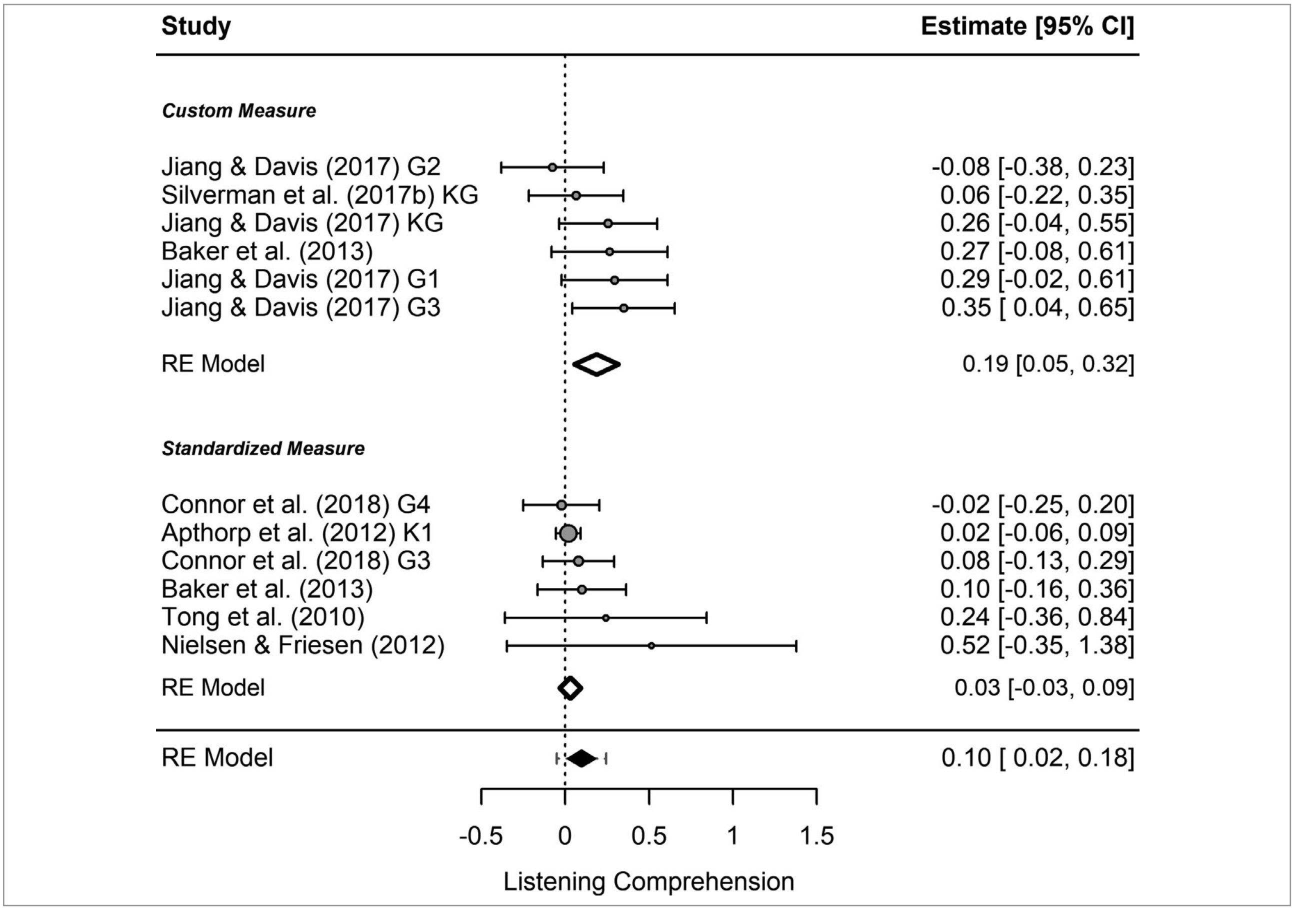
Note. The comparison category is at the bottom of each category. Bold indicates $p \leq .05$. Italics indicates $p \leq .10$.

FIGURE 1
Forest Plot for Vocabulary Outcomes



Note. CI = confidence interval; G1 = grade 1; G2 = grade 2; G3 = grade 3; G4 = grade 4; K1 = kindergarten and grade 1; KG = kindergarten; RE = random-effects; YR1 = year 1; YR2 = year 2.

FIGURE 2
Forest Plot for Listening Comprehension Outcomes



Note. CI = confidence interval; G1 = grade 1; G2 = grade 2; G3 = grade 3; G4 = grade 4; K1 = kindergarten and grade 1; KG = kindergarten; RE = random-effects.

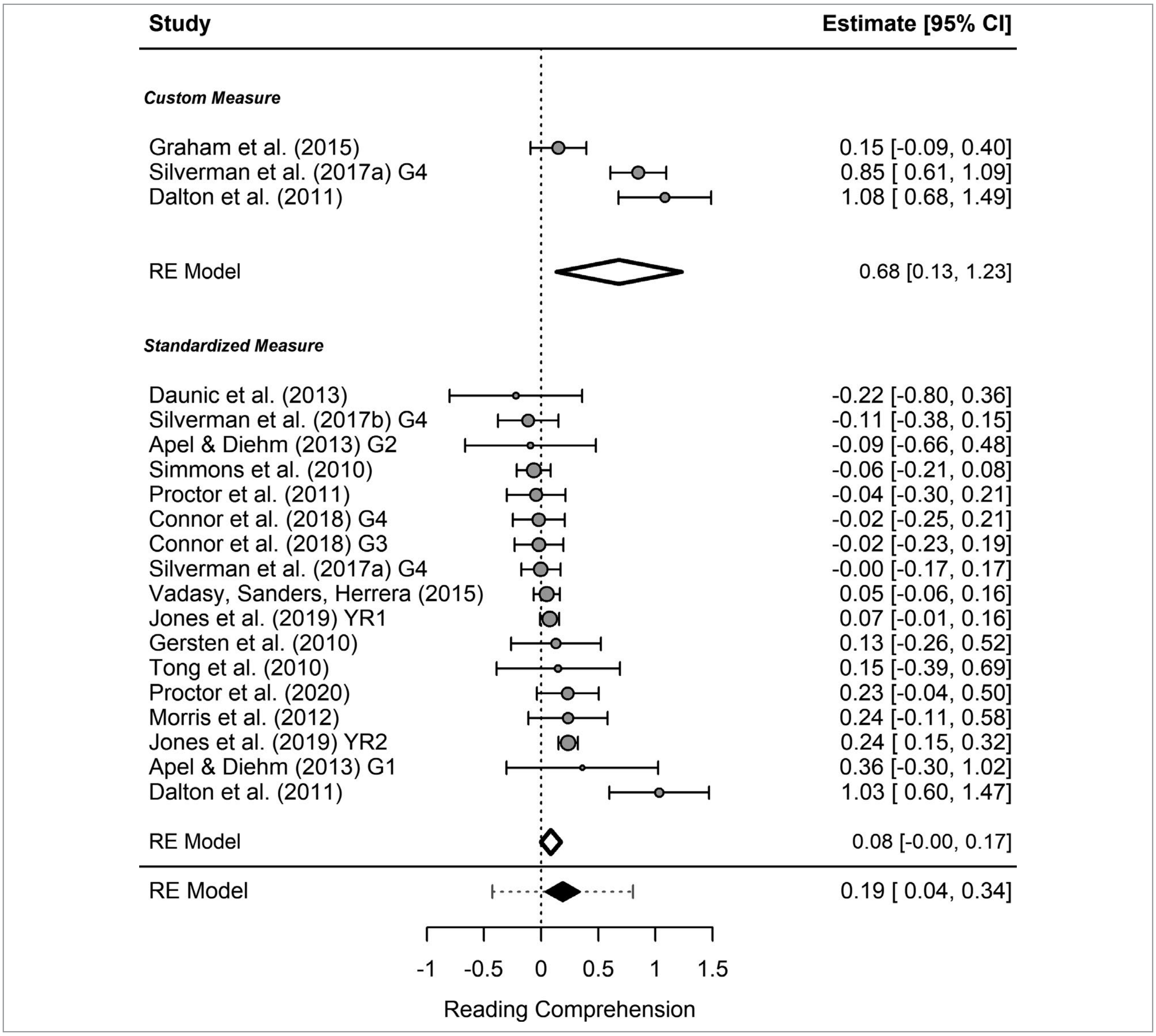
findings suggest positive and significant effects on custom measures of vocabulary, listening comprehension, and reading comprehension, although these effects were not seen on standardized measures. (Effects were found on a standardized measure of academic language, but only two studies included this measure, suggesting that more research is needed.) The custom measures of vocabulary were all proximal, indicating that interventions help students learn taught words. The custom measures of listening comprehension were all distal, suggesting some transfer of language comprehension beyond what was specifically taught. However, the findings suggest that current interventions are not strong enough to show effects on standardized measures, which are robust to relatively small changes in underlying ability. Of note, only one study in our corpus (Tong, Irby, Lara-Alecio, Yoon, & Mathes, 2010) lasted multiple years. This study showed positive effects on standardized vocabulary and reading comprehension. Given how challenging it is to realize effects on reading comprehension (Pearson et al., 2020), long-term interventions should be further investigated.

Participant Characteristics

Second, we considered whether effects differed for particular populations of students. Although effects did not differ across grades K–2 and 3–5, it is notable that most studies focused on K–2, with overrepresentation of kindergarten. More studies are needed to inform instruction in the upper elementary grades, when students encounter increasingly complex text. Overall, the reported participant sample was ethnically and racially diverse. However, because reporting was incomplete, we were unable to examine whether effects differed by race and ethnicity, which should be examined in the future.

We were able to explore effects by income status, EL background, and risk status. Studies with higher proportions of students from low-income backgrounds showed lower effects. Thus, research is needed on how to best support these students. Although no differences were seen for

FIGURE 3
Forest Plot for Reading Comprehension Outcomes



Note. CI = confidence interval; G1 = grade 1; G2 = grade 2; G3 = grade 3; G4 = grade 4; RE = random-effects; YR1 = year 1; YR2 = year 2.

TABLE 6
Sensitivity Analysis Results: Synthesized Effect Sizes for Each Approach

Outcome	Excluding studies with high risk of selection bias	Constraining outliers	Varying effect size dependence *r* = 0	*r* = .5	*r* = .8
Vocabulary	0.863 (.000)	0.726 (.000)	0.904 (.000)	0.904 (.000)	0.904 (.000)
Listening comprehension	0.099 (.028)	0.097 (.020)	0.083 (.110)	0.084 (.107)	0.085 (.105)
Reading comprehension	0.208 (.020)	0.129 (.010)	0.129 (.031)	0.129 (.031)	0.129 (.031)

Note. The *p*-values for corresponding effect sizes are in parentheses.

studies with higher (>25%) versus lower (<25%) EL participation, in studies explicitly comparing ELs with non-ELs, there were similar effects across these populations on vocabulary but greater effects for ELs on reading comprehension. These studies attended specifically to the strengths and needs of ELs (e.g., attention to cognates or translation),

FIGURE 4
Funnel Plot of Residualized Vocabulary Effect Sizes (*n* = 52)

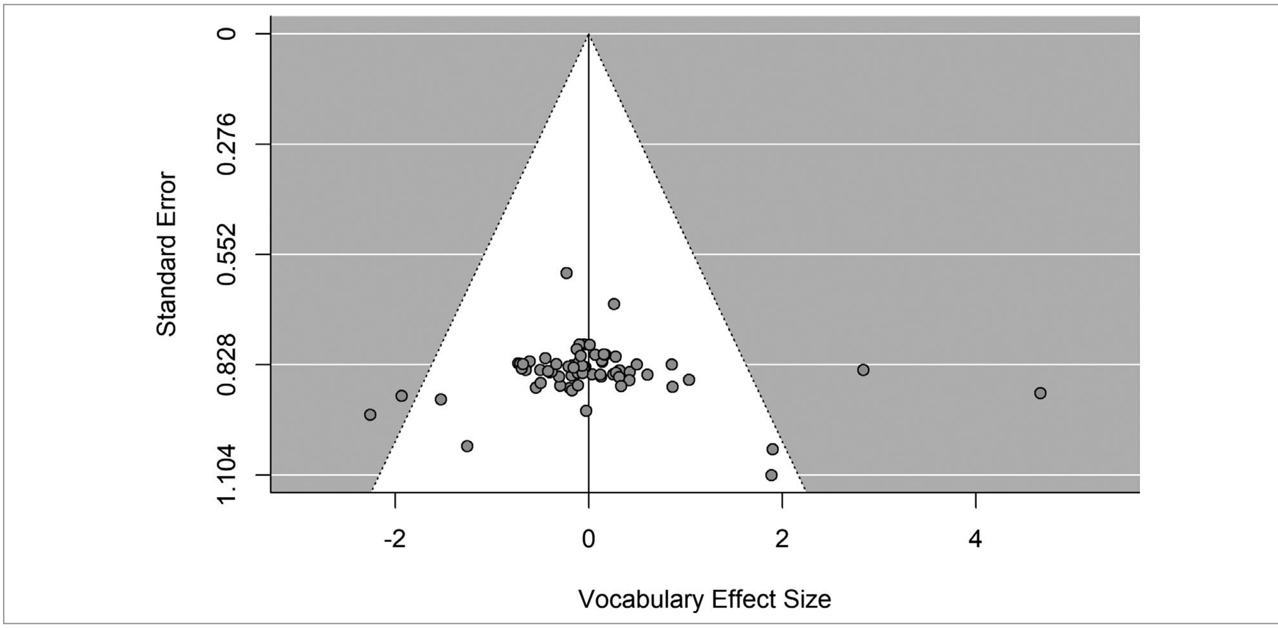
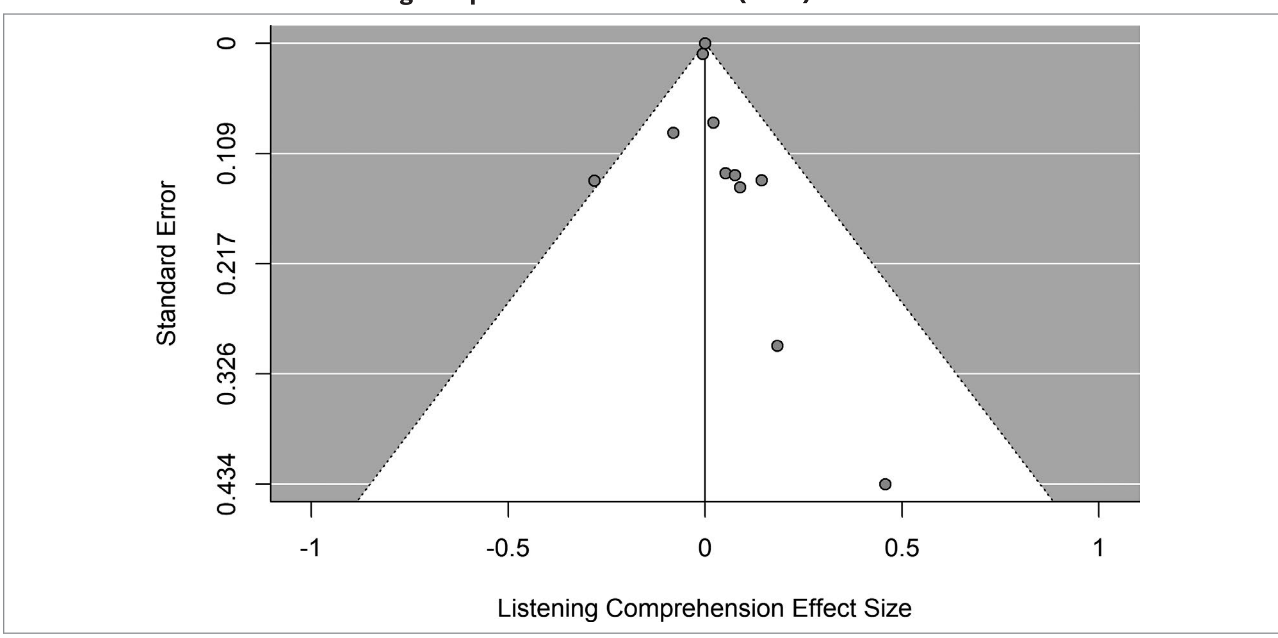


FIGURE 5
Funnel Plot of Residualized Listening Comprehension Effect Sizes (*n* = 11)



which may be why the interventions were equally or more effective for ELs as compared with non-ELs.

Findings from studies focused specifically on at-risk populations showed effects consistent with the rest of the corpus, potentially signaling that interventions focused on supporting at-risk students were sufficiently aligned with students' needs. Due to limited reporting, we could not examine whether effects differed when students with disabilities were included, so additional research conducted in inclusive contexts is needed. Several studies in the corpus investigated whether effects differed by initial level of language or literacy, but due to inconsistency in reporting, summarizing these effects was not possible. To provide clear indications of what kinds of interventions may be supportive for students from different backgrounds and abilities, future studies need to include more

FIGURE 6
Funnel Plot of Residualized Reading Comprehension Effect Sizes (*n* = 18)

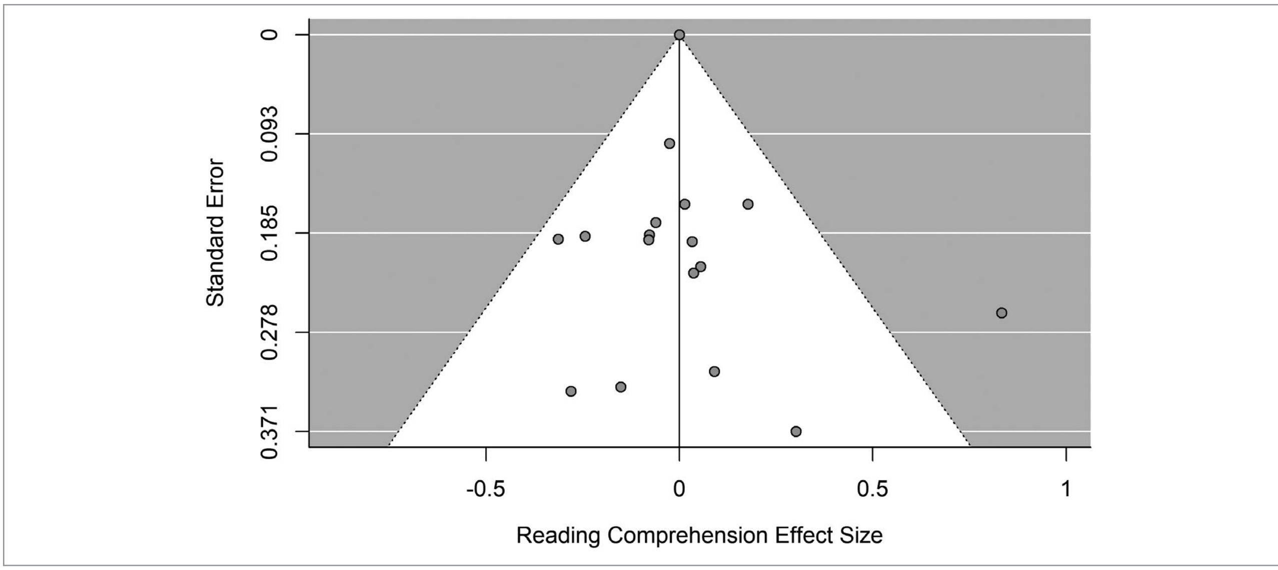
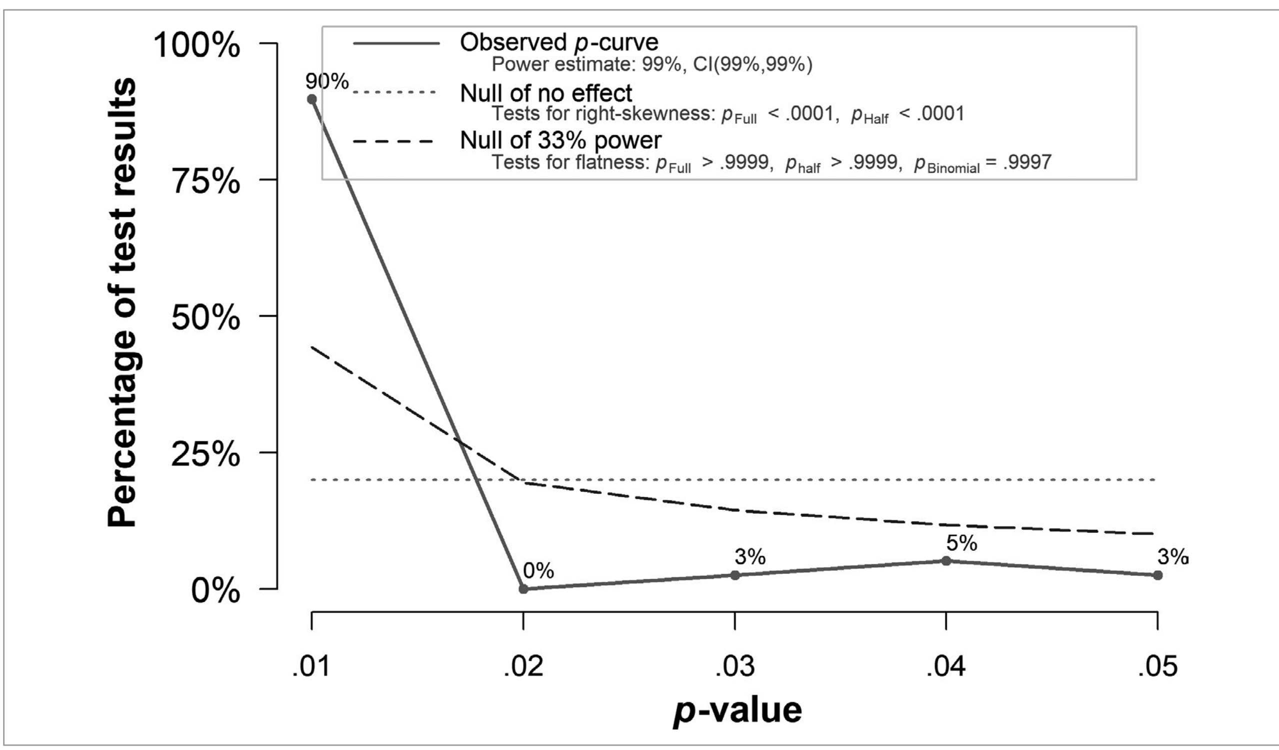


FIGURE 7
***p*-Curve Analysis for Vocabulary**



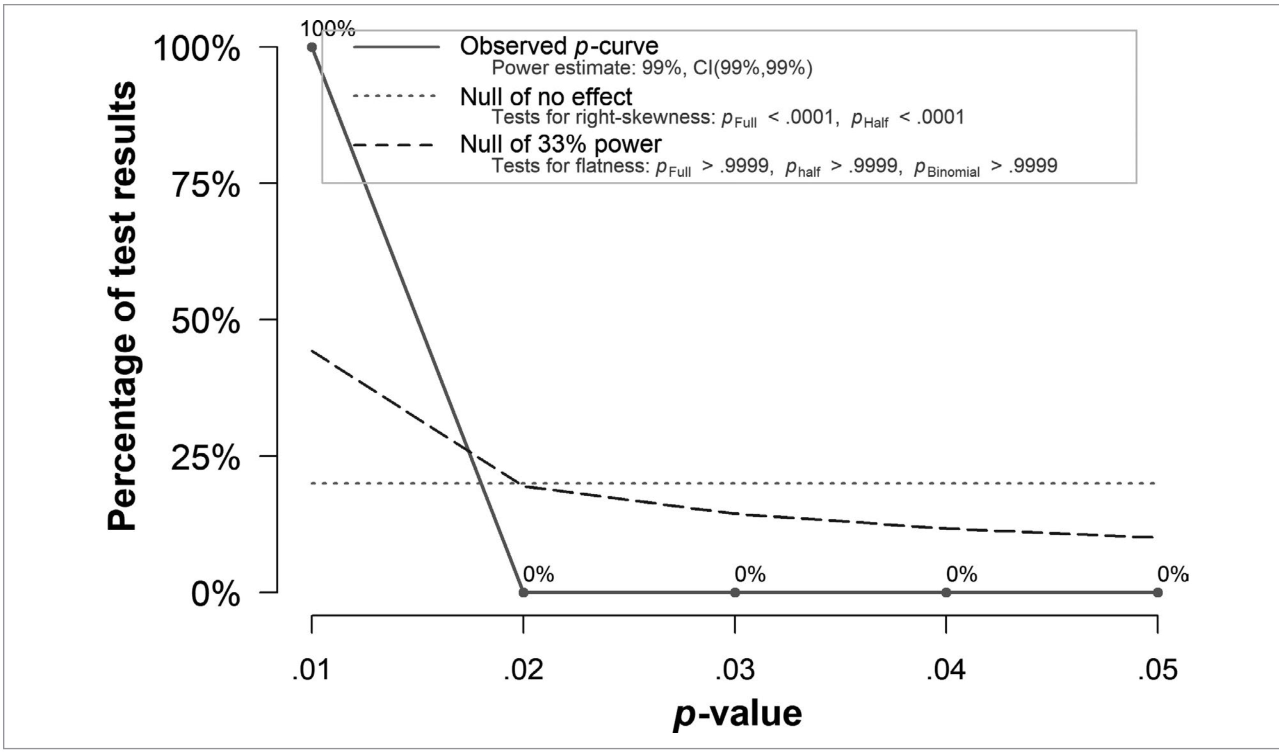
Note. CI = confidence interval. The observed *p*-curve includes 39 statistically significant ($p < .05$) results, of which 36 are $p < .025$. Twenty-nine additional results were entered but excluded from the *p*-curve because they were $p > .05$.

direct and systematic comparison of effects across diverse student populations.

Intervention Characteristics

Finally, we explored whether effects differed according to specific intervention characteristics. There was no indication that duration was associated with effects, and results for grouping were inconsistent. Considering that other meta-analyses have found that intervention effects are correlated with duration and group size (e.g., Hall & Burns, 2018), more research that systematically varies these intervention characteristics is needed.

FIGURE 8
***p*-Curve Analysis for Reading Comprehension**



Note. CI = confidence interval. The observed *p*-curve includes four statistically significant ($p < .05$) results, of which all four are $p < .025$. Sixteen additional results were entered but excluded from the *p*-curve because they were $p > .05$.

The majority of the studies in our corpus included attention to more than one component of language (i.e., vocabulary and semantics, morphology, and/or syntax), and results indicated that integrated instruction may be beneficial. Specifically, interventions including attention to both vocabulary and morphology showed positive effects on vocabulary, and interventions combining syntax and vocabulary showed positive effects on reading comprehension. In line with a view of language comprehension as a unified construct (Kieffer et al., 2016), intervention that targets individual components and brings them together with authentic opportunities to use language for comprehension and expression may be most effective. It is important to note that meta-analyses such as this one focus on quantitative studies that tend to take a more componential view of language. These should be complemented with reviews of qualitative studies, which can offer a more holistic view of language comprehension and use.

Particularly relevant to the debate on the science of reading, which has focused primarily on decoding, studies that included decoding-related instruction (e.g., phonological awareness, phonics) did not show differential effects. Whereas some researchers have suggested that combining decoding and language comprehension instruction may be beneficial (e.g., Rosenthal & Ehri, 2008), findings from this meta-analysis do not support that claim. Furthermore, language comprehension interventions did not show effects on decoding, suggesting that language comprehension and decoding instruction may not be reciprocal in elementary school. As in Scarborough's (2001) model, in which decoding and language comprehension develop separately at first and then come together over time in the process of reading comprehension, instruction in decoding and language comprehension may best be separated in elementary school as students are developing facility with each of these components. Given the relatively small number of studies that included decoding instruction ($n = 8$) or decoding outcomes ($n = 7$), much more research is needed to understand the relation between decoding and language comprehension instruction and outcomes.

Interestingly, although relatively few studies in our corpus included technology, results indicated that its use in language comprehension interventions may support reading comprehension, which aligns with the findings of previous reviews identifying affordances of using technology to support language and literacy (e.g., National Reading Panel, 2000). Theoretically, providing multiple representations of content and interactive experiences could enable students to develop deeper language comprehension than through teacher instruction alone (Kamil, Intrator, & Kim, 2000). The way technology is used in language comprehension interventions has not been consistent, however, and considering that the use of technology is sometimes associated with negative effects, such as when it is distracting (Kamil et al., 2000), much more research is needed.

Effects were not associated with other characteristics of content or context. This is likely because we did not have power to detect differential effects or because the interventions were too homogeneous to examine these effects. Future studies should examine the potential benefits of teaching strategies, including discussion and writing, or using non–English language arts (i.e., science, social studies, math) content to support language comprehension. Finally, only three interventions were part of larger professional development programs, which can support teachers in implementing and adapting language comprehension instruction (Markussen-Brown et al., 2017). Conceivably, any of these variables could make language comprehension intervention more effective, but much more research is needed to clarify which matter most.

Limitations

Findings from meta-analyses must be considered with caution, as such analyses generalize from the characteristics of existing studies. Across studies, effects may be confounded with the outcomes examined and the participant sample and intervention characteristics included. For example, many studies that included discussion, strategies, or writing in the intervention examined effects on standardized measures. Thus, null effects for these intervention characteristics could reflect how hard it is to affect change on standardized measures rather than whether these instructional approaches are supportive of language comprehension. Further, because many studies did not provide or disaggregate findings by demographic characteristics, we were unable to explore whether particular intervention characteristics were more or less supportive for students from different populations. Hopefully, future research will allow for more fine-grained analyses because such analyses are needed to inform policy and practice.

Implications for Policy and Practice

Given that findings from this meta-analysis indicate that much more research on K–5 language comprehension intervention is needed, policymakers must prioritize and allocate resources toward this research. The large-scale U.S. Institute of Education Sciences investment in the Reading for Understanding initiative was one step in the right direction, but this initiative uncovered that moving the needle on reading comprehension is not easy and requires commitment and sustenance (Pearson et al., 2020). Further investment in research on language comprehension interventions intended to serve diverse populations is needed (Pearson et al., 2020). Beyond supporting research, policymakers can also provide guidance for teacher preparation and professional development, as well as curriculum and intervention design, to reflect what is known about language comprehension and the kind of instruction needed to support it.

Although findings from this review should be regarded cautiously, as we compiled information from across a wide range of studies, results suggest a few directions to guide practitioners. Specifically, our findings suggest that language comprehension, which research has established is an essential component of reading comprehension, can be taught effectively. Instruction that attends to multiple components of language (e.g., vocabulary and semantics, morphology, syntax) may be helpful, and instruction that uses technology to facilitate language comprehension (rather than detract from it) may be supportive for students in K–5 settings. Practitioners of K–5 students must weigh competing priorities while planning instruction for an already crowded literacy block. With the debate on the science of reading focused on decoding, the importance of effectively integrating attention to language comprehension has been sidelined. We recommend that practitioners stay abreast of the research on language comprehension as it evolves and, hopefully, becomes increasingly relevant to instructional decision making.

Conclusion

Language comprehension is an important component of reading that develops alongside decoding and becomes evermore important in the reading process as students move through the elementary grades. Thus, in conversations about the science of reading, language comprehension, and how to support its development among diverse populations, should not be overlooked.

NOTES

[1] Studies were included if at least 75% of the sample was in K–5 and/or the data were disaggregated by grade level.

[2] Studies focused on other aspects of language (e.g., narrative development, pragmatics) or primarily targeting comprehension strategies or text structures were excluded.

[3] Heterogeneity was significant for combined (i.e., custom and standardized measures) of vocabulary and custom vocabulary measures separately. Heterogeneity was also significant for custom and standardized reading comprehension combined and separately. For parsimony, we report on effects on combined vocabulary and reading comprehension, although results were fairly consistent across types of measures. Findings by type of measure are available upon request.

[4] We could not generate a *p*-curve for listening comprehension because only one study reported effects with high statistical significance ($p < .05$).

REFERENCES

References marked with an asterisk indicate studies included in the meta-analysis.

Adlof, S.M., Catts, H.W., & Lee, J. (2010). Kindergarten predictors of second versus eighth grade reading comprehension impairments. *Journal of Learning Disabilities*, *43*(4), 332–345. https://doi.org/10.1177/0022219410369067

*Apel, K., Brimo, D., Diehm, E., & Apel, L. (2013). Morphological awareness intervention with kindergartners and first- and second-grade students from low socioeconomic status homes: A feasibility study. *Language, Speech, and Hearing Services in Schools*, *44*(2), 161–173. https://doi.org/10.1044/0161-1461(2012/12-0042)

*Apthorp, H., Randel, B., Cherasaro, T., Clark, T., McKeown, M., & Beck, I. (2012). Effects of a supplemental vocabulary program on word knowledge and passage comprehension. *Journal of Research on Educational Effectiveness*, *5*(2), 160–188. https://doi.org/10.1080/19345747.2012.660240

*Arthur, A.M., & Davis, D.L. (2016). A pilot study of the impact of double-dose robust vocabulary instruction on children's vocabulary growth. *Journal of Research on Educational Effectiveness*, *9*(2), 173–200. https://doi.org/10.1080/19345747.2015.1126875

Atal, I., Porcher, R., Boutron, I., & Ravaud, P. (2019). The statistical significance of meta-analyses is frequently fragile: Definition of a fragility index for meta-analyses. *Journal of Clinical Epidemiology*, *111*, 32–40. https://doi.org/10.1016/j.jclinepi.2019.03.012

*August, D., Artzi, L., Barr, C., & Francis, D. (2018). The moderating influence of instructional intensity and word type on the acquisition of academic vocabulary in young English language learners. *Reading and Writing*, *31*(4), 965–989. https://doi.org/10.1007/s11145-018-9821-1

*Baker, S.K., Santoro, L.E., Chard, D.J., Fien, H., Park, Y., & Otterstedt, J. (2013). An evaluation of an explicit read aloud intervention taught in whole-classroom formats in first grade. *The Elementary School Journal*, *113*(3), 331–358. https://doi.org/10.1086/668503

Borenstein, M. (2009). Effect sizes for continuous data. In H. Cooper, L.V. Hedges, & J.C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (2nd ed., pp. 221–235). New York, NY: Russell Sage Foundation.

Bowers, P.N., Kirby, J.R., & Deacon, S.H. (2010). The effects of morphological instruction on literacy skills: A systematic review of the literature. *Review of Educational Research*, *80*(2), 144–179. https://doi.org/10.3102/0034654309359353

*Brimo, D. (2016). Evaluating the effectiveness of a morphological awareness intervention: A pilot study. *Communication Disorders Quarterly*, *38*(1), 35–45. https://doi.org/10.1177/1525740115604592

Carlisle, J.F. (2010). Effects of instruction in morphological awareness on literacy achievement: An integrative review. *Reading Research Quarterly*, *45*(4), 464–487. https://doi.org/10.1598/RRQ.45.4.5

*Connor, C.M., Day, S.L., Zargar, E., Wood, T.S., Taylor, K.S., Jones, M.R., & Hwang, J.K. (2019). Building word knowledge, learning strategies, and metacognition with the Word-Knowledge e-Book. *Computers & Education*, *128*, 284–311. https://doi.org/10.1016/j.compedu.2018.09.016

Connor, C.M., Morrison, F.J., Fishman, B., Giuliani, S., Luck, M., Underwood, P.S., … Schatschneider, C. (2011). Testing the impact of child characteristics × instruction interactions on third graders' reading comprehension by differentiating literacy instruction. *Reading Research Quarterly*, *46*(3), 189–221. https://doi.org/10.1598/RRQ.46.3.1

*Connor, C.M., Phillips, B.M., Kim, Y.-S.G., Lonigan, C.J., Kaschak, M.P., Crowe, E., … Al Otaiba, S. (2018). Examining the efficacy of targeted component interventions on language and literacy for third and fourth graders who are at risk of comprehension difficulties. *Scientific Studies of Reading*, *22*(6), 462–484. https://doi.org/10.1080/10888438.2018.1481409

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

*Coyne, M.D., McCoach, D.B., Loftus, S., Zipoli, R., Ruby, M., Crevecoeur, Y.C., & Kapp, S. (2010). Direct and extended vocabulary instruction in kindergarten: Investigating transfer effects. *Journal of Research on Educational Effectiveness*, *3*(2), 93–120. https://doi.org/10.1080/19345741003592410

*Coyne, M.D., McCoach, D.B., Ware, S., Austin, C.R., Loftus-Rattan, S.M., & Baker, D.L. (2019). Racing against the vocabulary gap: Matthew effects in early vocabulary instruction and intervention. *Exceptional Children*, *85*(2), 163–179. https://doi.org/10.1177/0014402918789162

*Crevecoeur, Y.C., Coyne, M.D., & McCoach, D.B. (2014). English language learners and English-only learners' response to direct vocabulary instruction. *Reading & Writing Quarterly*, *30*(1), 51–78. https://doi.org/10.1080/10573569.2013.758943

Cunningham, A.E., & Stanovich, K.E. (1997). Early reading acquisition and its relation to reading experience and ability 10 years later. *Developmental Psychology*, *33*(6), 934–945. https://doi.org/10.1037/0012-1649.33.6.934

*Dalton, B., Proctor, C.P., Uccelli, P., Mo, E., & Snow, C.E. (2011). Designing for diversity: The role of reading strategies and interactive vocabulary in a digital reading environment for fifth-grade monolingual English and bilingual students. *Journal of Literacy Research*, *43*(1), 68–100. https://doi.org/10.1177/1086296X10397872

*Daunic, A., Corbett, N., Smith, S., Barnes, T., Santiago-Poventud, L., Chalfant, P., … Gleaton, J. (2013). Brief report: Integrating social-emotional learning with literacy instruction: An intervention for children at risk for emotional and behavioral disorders. *Behavioral Disorders*, *39*(1), 43–51. https://doi.org/10.1177/019874291303900106

Egger, M., Smith, G.D., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *The British Medical Journal*, *315*, 629–634. https://doi.org/10.1136/bmj.315.7109.629

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), 1–44. https://doi.org/10.1080/19345740802539200

*Filippini, A.L., Gerber, M.M., & Leafstedt, J.M. (2012). A vocabulary-added reading intervention for English learners at-risk of reading difficulties. *International Journal of Special Education*, *27*(3), 14–26.

Florit, E., & Cain, K. (2011). The simple view of reading: Is it valid for different types of alphabetic orthographies? *Educational Psychology Review*, *23*, 553–576. https://doi.org/10.1007/s10648-011-9175-6

Foorman, B.R., Herrera, S., & Dombek, J. (2018). The relative impact of aligning Tier 2 intervention materials with classroom core reading materials in grades K–2. *The Elementary School Journal*, *118*(3), 477–504. https://doi.org/10.1086/696021

*Gersten, R., Dimino, J., Jayanthi, M., Kim, J.S., & Santoro, L.E. (2010). Teacher study group: Impact of the professional development model on reading instruction and student outcomes in first grade classrooms. *American Educational Research Journal*, *47*(3), 694–739. https://doi.org/10.3102/0002831209361208

*Goldstein, H., Ziolkowski, R.A., Bojczyk, K.E., Marty, A., Schneider, N., Harpring, J., & Haring, C.D. (2017). Academic vocabulary learning in first through third grade in low-income schools: Effects of automated supplemental instruction. *Journal of Speech, Language, and Hearing Research*, *60*(11), 3237–3258. https://doi.org/10.1044/2017_JSLHR-L-17-0100

Goodwin, A.P., & Ahn, S. (2013). A meta-analysis of morphological interventions in English: Effects on literacy outcomes for school-age children. *Scientific Studies of Reading*, *17*(4), 257–285. https://doi.org/10.1080/10888438.2012.689791

Gough, P.B., & Tunmer, W.E. (1986). Decoding, reading, and reading disability. *Remedial and Special Education*, *7*(1), 6–10. https://doi.org/10.1177/074193258600700104

*Graham, L., Graham, A., & West, C. (2015). From research to practice: The effect of multi-component vocabulary instruction on increasing vocabulary and comprehension performance in social studies. *International Electronic Journal of Elementary Education*, *8*(1), 147–160.

Hall, M.S., & Burns, M.K. (2018). Meta-analysis of targeted small-group reading interventions. *Journal of School Psychology*, *66*, 54–66. https://doi.org/10.1016/j.jsp.2017.11.002

*Hassinger-Das, B., Jordan, N.C., & Dyson, N. (2015). Reading stories to learn math: Mathematics vocabulary instruction for children with early numeracy difficulties. *The Elementary School Journal*, *116*(2), 242–264. https://doi.org/10.1086/683986

Hedges, L.V. (1982). Estimation of effect size from a series of independent experiments. *Psychological Bulletin*, *92*(2), 490–499. https://doi.org/10.1037/0033-2909.92.2.490

Hedges, L.V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. Orlando, FL: Academic.

Higgins, J., & Green, S. (2008). *Cochrane handbook for systematic reviews of interventions*. London, UK: The Cochrane Collaboration; Chichester, UK: John Wiley & Sons.

*Huang, S. (2015). Mixed-method research on learning vocabulary through technology reveals vocabulary growth in second-grade students. *Reading Psychology*, *36*(1), 1–30. https://doi.org/10.1080/02702711.2013.808723

*Jayanthi, M., Dimino, J., Gersten, R., Taylor, M.J., Haymond, K., Smolkowski, K., & Newman-Gonchar, R. (2018). The impact of teacher study groups in vocabulary on teaching practice, teacher knowledge, and student vocabulary knowledge: A large-scale replication study. *Journal of Research on Educational Effectiveness*, *11*(1), 83–108. https://doi.org/10.1080/19345747.2017.1327625

*Jones, S.M., LaRusso, M., Kim, J., Kim, H.Y., Selman, R., Uccelli, P., … Snow, C. (2019). Experimental effects of Word Generation on vocabulary, academic language, perspective taking, and reading comprehension in high-poverty schools. *Journal of Research on Educational Effectiveness*, *12*(3), 448–483. https://doi.org/10.1080/19345747.2019.1615155

Kamil, M.L., Intrator, S.M., & Kim, H.S. (2000). The effects of other technologies on literacy and literacy learning. In M.L. Kamil, P.B. Mosenthal, P.D. Pearson, & R. Barr (Eds.), *Handbook of reading research* (Vol. 3, pp. 771–788). Mahwah, NJ: Erlbaum.

Kieffer, M.J., Petscher, Y., Proctor, C.P., & Silverman, R.D. (2016). Is the whole greater than the sum of its parts? Modeling the contributions of language comprehension skills to reading comprehension in the upper elementary grades. *Scientific Studies of Reading*, *20*(6), 436–454. https://doi.org/10.1080/10888438.2016.1214591

Kim, Y.-S.G., Petscher, Y., Wanzek, J., & Al Otaiba, S. (2018). Relations between reading and writing: A longitudinal examination from grades 3 to 6. *Reading and Writing*, *31*(7), 1591–1618. https://doi.org/10.1007/s11145-018-9855-4

*Language and Reading Research Consortium, Jiang, H., & Davis, D. (2017). Let's Know! Proximal impacts on prekindergarten through grade 3 students' comprehension-related skills. *The Elementary School Journal*, *118*(2), 177–206. https://doi.org/10.1086/694220

*Mancilla-Martinez, J. (2010). Word meanings matter: Cultivating English vocabulary knowledge in fifth-grade Spanish-speaking language minority learners. *TESOL Quarterly*, *44*(4), 669–699. https://doi.org/10.5054/tq.2010.213782

Markussen-Brown, J., Juhl, C.B., Piasta, S.B., Bleses, D., Højen, A., & Justice, L.M. (2017). The effects of language and literacy-focused professional development on early educators and children: A best-evidence meta-analysis. *Early Childhood Research Quarterly*, *38*(1), 97–115. https://doi.org/10.1016/j.ecresq.2016.07.002

Marulis, L.M., & Neuman, S.B. (2010). The effects of vocabulary intervention on young children's word learning: A meta-analysis. *Review of Educational Research*, *80*(3), 300–335. https://doi.org/10.3102/0034654310377087

*McKeown, M.G., & Beck, I.L. (2014). Effects of vocabulary instruction on measures of language processing: Comparing two approaches. *Early Childhood Research Quarterly*, *29*(4), 520–530. https://doi.org/10.1016/j.ecresq.2014.06.002

*Morris, R.D., Lovett, M.W., Wolf, M., Sevcik, R.A., Steinbach, K.A., Frijters, J.C., & Shapiro, M.B. (2012). Multiple-component remediation for developmental reading disabilities: IQ, socioeconomic status, and race as factors in remedial outcome. *Journal of Learning Disabilities*, *45*(2), 99–127. https://doi.org/10.1177/0022219409355472

Morris, S.B. (2008). Estimating effect sizes from pretest-posttest-control group designs. *Organizational Research Methods*, *11*(2), 364–386. https://doi.org/10.1177/1094428106291059

National Governors Association Center for Best Practices & Council of Chief State School Officers. (2010). *Common Core State Standards for English language arts and literacy in history/social studies, science, and technical subjects*. Washington, DC: Authors.

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

*Nelson, J.R., Vadasy, P.F., & Sanders, E.A. (2011). Efficacy of a Tier 2 supplemental root word vocabulary and decoding intervention with kindergarten Spanish-speaking English learners. *Journal of Literacy Research*, *43*(2), 184–211. https://doi.org/10.1177/1086296X11403088

*Neuman, S.B., & Kaefer, T. (2018). Developing low-income children's vocabulary and content knowledge through a shared book reading program. *Contemporary Educational Psychology*, *52*, 15–24. https://doi.org/10.1016/j.cedpsych.2017.12.001

*Nielsen, D.C., & Friesen, L.D. (2012). A study of the effectiveness of a small-group intervention on the vocabulary and narrative development of at-risk kindergarten children. *Reading Psychology*, *33*(3), 269–299. https://doi.org/10.1080/02702711.2010.508671

Pearson, P.D., Palincsar, A.S., Biancarosa, G., & Berman, A.I. (Eds.). (2020). *Reaping the rewards of the Reading for Understanding initiative*. Washington, DC: National Academy of Education.

*Powell, S.R., & Driver, M.K. (2015). The influence of mathematics vocabulary instruction embedded within addition tutoring for first-grade students with mathematics difficulty. *Learning Disability Quarterly*, *38*(4), 221–233. https://doi.org/10.1177/0731948714564574

*Proctor, C.P., Dalton, B., Uccelli, P., Biancarosa, G., Mo, E., Snow, C., & Neugebauer, S. (2011). Improving comprehension online: Effects of deep vocabulary instruction with bilingual and monolingual fifth graders. *Reading and Writing*, *24*(5), 517–544. https://doi.org/10.1007/s11145-009-9218-2

*Proctor, C.P., Silverman, R.D., Harring, J.R., Jones, R.L., & Hartranft, A.M. (2020). Teaching bilingual learners: Effects of a language-based reading intervention on academic language and reading comprehension in grades 4 and 5. *Reading Research Quarterly*, *55*(1), 95–122. https://doi.org/10.1002/rrq.258

*Puhalla, E.M. (2011). Enhancing the vocabulary knowledge of first-grade children with supplemental booster instruction. *Remedial and Special Education*, *32*(6), 471–481. https://doi.org/10.1177/0741932510362495

*Pullen, P.C., Tuckwiller, E.D., Konold, T.R., Maynard, K.L., & Coyne, M.D. (2010). A tiered intervention model for early vocabulary instruction: The effects of tiered instruction for young students at risk for reading disability. *Learning Disabilities Research & Practice*, *25*(3), 110–123. https://doi.org/10.1111/j.1540-5826.2010.00309.x

Rogde, K., Hagen, Å.M., Melby-Lervåg, M., & Lervåg, A. (2019). The effect of linguistic comprehension instruction on generalized language and reading comprehension skills: A systematic review. *Campbell Systematic Reviews*, *15*(4), article e1059. https://doi.org/10.1002/cl2.1059

Rosenthal, J., & Ehri, L.C. (2008). The mnemonic value of orthography for vocabulary learning. *Journal of Educational Psychology*, *100*(1), 175–191. https://doi.org/10.1037/0022-0663.100.1.175

Scarborough, H.S. (2001). Connecting early language and literacy to later reading (dis)abilities: Evidence, theory, and practice. In S.B. Neuman & D.K. Dickinson (Eds.), *Handbook of early literacy research* (pp. 97–110). New York, NY: Guilford.

*Silverman, R., Kim, Y.-S., Hartranft, A., Nunn, S., & McNeish, D. (2017). Effects of a multimedia enhanced reading buddies program on kindergarten and grade 4 vocabulary and comprehension. *The Journal of Educational Research*, *110*(4), 391–404. https://doi.org/10.1080/00220671.2015.1103690

*Silverman, R.D., Martin-Beltran, M., Peercy, M.M., Hartranft, A.M., McNeish, D.M., Artzi, L., & Nunn, S. (2017). Effects of a cross-age peer learning program on the vocabulary and comprehension of English learners and non-English learners in elementary school. *The Elementary School Journal*, *117*(3), 485–512. https://doi.org/10.1086/690210

*Simmons, D., Hairrell, A., Edmonds, M., Vaughn, S., Larsen, R., Willson, V., … Byrns, G. (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*(2), 121–156. https://doi.org/10.1080/19345741003596890

Simonsohn, U., Nelson, L.D., & Simmons, J.P. (2014). *p*-Curve and effect size: Correcting for publication bias using only significant results. *Perspectives on Psychological Science*, *9*(6), 666–681. https://doi.org/10.1177/1745691614553988

Sparapani, N., Carlisle, J.F., & Connor, C.M. (2018). Observations of vocabulary activities during second- and third-grade reading lessons. *Education in Science*, *8*, article 198. https://doi.org/10.3390/educsci8040198

Stahl, S.A., & Fairbanks, M.M. (1986). The effects of vocabulary instruction: A model-based meta-analysis. *Review of Educational Research*, *56*(1), 72–110. https://doi.org/10.3102/00346543056001072

*Tong, F., Irby, B.J., Lara-Alecio, R., Yoon, M., & Mathes, P.G. (2010). Hispanic English learners' responses to longitudinal English instructional intervention and the effect of gender: A multilevel analysis. *The Elementary School Journal*, *110*(4), 542–566. https://doi.org/10.1086/651195

*Vadasy, P.F., Nelson, J.R., & Sanders, E.A. (2013). Longer term effects of a Tier 2 kindergarten vocabulary intervention for English learners. *Remedial and Special Education*, *34*(2), 91–101. https://doi.org/10.1177/0741932511420739

*Vadasy, P.F., & Sanders, E.A. (2015). Incremental learning of difficult words in story contexts: The role of spelling and pronouncing new vocabulary. *Reading and Writing*, *28*(3), 371–394. https://doi.org/10.1007/s11145-014-9529-9

*Vadasy, P.F., & Sanders, E.A. (2016). Attention to orthographic and phonological word forms in vocabulary instruction for kindergarten English learners. *Reading Psychology*, *37*(6), 833–866. https://doi.org/10.1080/02702711.2015.1116477

*Vadasy, P.F., Sanders, E.A., & Logan Herrera, B. (2015). Efficacy of rich vocabulary instruction in fourth- and fifth-grade classrooms. *Journal of Research on Educational Effectiveness*, *8*(3), 325–365. https://doi.org/10.1080/19345747.2014.933495

*Vadasy, P.F., Sanders, E.A., & Nelson, J.R. (2015). Effectiveness of supplemental kindergarten vocabulary instruction for English learners: A randomized study of immediate and longer-term effects of two approaches. *Journal of Research on Educational Effectiveness*, *8*(4), 490–529. https://doi.org/10.1080/19345747.2015.1033120

*Wood, C., Fitton, L., Petscher, Y., Rodriguez, E., Sunderman, G., & Lim, T. (2018). The effect of e-book vocabulary instruction on Spanish-English speaking children. *Journal of Speech, Language, and Hearing Research*, *61*(8), 1945–1969. https://doi.org/10.1044/2018_JSLHR-L-17-0368

Wright, T.S., & Cervetti, G.N. (2017). A systematic review of the research on vocabulary instruction that impacts text comprehension. *Reading Research Quarterly*, *52*(2), 203–226. https://doi.org/10.1002/rrq.163

*Wright, T.S., & Gotwals, A.W. (2017). Supporting kindergartners' science talk in the context of an integrated science and disciplinary literacy curriculum. *The Elementary School Journal*, *117*(3), 513–537. https://doi.org/10.1086/690273

*Zipoli, R.P., Jr., Coyne, M.D., & McCoach, D.B. (2011). Enhancing vocabulary intervention for kindergarten students: Strategic integration of semantically related and embedded word review. *Remedial and Special Education*, *32*(2), 131–143. https://doi.org/10.1177/0741932510361262

Submitted May 1, 2020
Final revision received June 24, 2020
Accepted June 25, 2020

REBECCA D. SILVERMAN (corresponding author) is an associate professor in the Graduate School of Education at Stanford University, California, USA; email rdsilver@stanford.edu. She conducts mostly quantitative research on the role of language comprehension in reading development throughout childhood.

ERIKA JOHNSON is an assistant professor in the College of Education at the University of Iowa, Iowa City, USA; email erika-johnson@uiowa.edu. She qualitatively examines reading instruction with linguistically diverse student populations.

KRISTIN KEANE is a doctoral candidate in the Graduate School of Education at Stanford University, California, USA; email knkeane@stanford.edu. She is interested in mixed-methods approaches to reading research in the elementary grades.

SAURABH KHANNA is a doctoral candidate in the Graduate School of Education at Stanford University, California, USA; email khanna90@stanford.edu. He is interested in using quantitative methods to inform educational policy.