Text Types and Their Relation to Efficacy in Beginning Reading Interventions

Alia Pugh[1], Devin Kearns[1], and Elfrieda H. Hiebert[2]
[1] University of Connecticut, Storrs, Connecticut, United States
[2] TextProject, Santa Cruz, California, United States

Author Notes

Alia Pugh https://orcid.org/0000-0002-5792-9991
Devin Kearns http://orcid.org/0000-0001-9703-0932
Elfrieda H. Hiebert https://orcid.org/0000-0002-2864-7549

We thank Leigh Ann Martin for her assistance as an additional rater of studies.
Study data can be made available upon request.
We have no known conflicts of interest or ethical concerns to disclose.

This is a preprint version of the article and it therefore may include typographical errors and minor differences from the version published in *Reading Research Quarterly*.

Correspondence concerning this article should be addressed to Alia Pugh, Department of Educational Psychology, Neag School of Education, University of Connecticut, 249 Glenbrook Road, Unit 3064, Storrs, CT 06269. Email: alia.pugh@uconn.edu

Abstract

Researchers disagree about the value of controlling the decodability of texts for students with reading difficulty, specifically what type of text they should read: *decodable texts* (words limited to taught patterns), *nondecodable texts* (those not limited by instruction), or *both*. We analyzed the effects of reading intervention for elementary-age students with reading difficulty (k = 119) to determine whether effects varied by the type of texts students read—decodable, nondecodable, or both—compared with interventions including no text reading. Inadequate information was available to code text type for 22 interventions including text reading; effect sizes were calculated for 97 studies. Effects for interventions with decodable or nondecodable reading did not differ from no-text interventions. For both-types interventions, the effect (g = 0.28) approached significance versus no-text, 95% CI [-0.09, 0.65]. Disaggregating effects by whether the measures were standardized or researcher-designed showed a significant both-types effect, g = 0.45, 95% CI [0.02, 0.89] relative to no-text. Disaggregating by whether outcomes were for word recognition or reading comprehension showed a positive both-types effect for word recognition outcomes; data were inadequate to examine comprehension. A possible confounding effect of time spent reading was tested but was uncorrelated with the intervention effect. The both-types finding suggests the possible value of varied reading experiences in intervention, but this analysis did not account for other factors that might be correlated with text type and the intervention effect. Further, more comprehensive reporting about text types is important for replication and meta-analytic review.

Keywords: reading interventions, replication, text analysis, decodable texts, leveled texts

Text Types and Their Relation to Efficacy in Beginning Reading Interventions

The great debate about the efficacy of reading different types of texts in beginning reading instruction has again become prominent among policymakers and researchers (Goldberg & Goldenberg, 2022). The two text types at the center of the current debate are decodable texts, which include controlled vocabulary that emphasizes taught phonetic patterns, and nondecodable texts, which do not include such vocabulary control (The Reading League, 2022; Schwartz, 2019). This issue is particularly salient when examining interventions for students with reading difficulty because interventions must maximize their efficiency by providing the reading experiences that will best accelerate their achievement. Unfortunately, little quasi-experimental or experimental research has involved systematic examination of the relation between the types of texts students read in intensive reading interventions and the effect of that instruction on their achievement (*cf*. reviews by Cheatham & Allor, 2012 and Birch et al., 2022). The aim of this study, then, is to examine the relation between the texts used in beginning reading interventions and the effects of those interventions on student reading achievement.

Despite the limited literature, some influential studies provide evidence that the types of texts used in beginning reading instruction and interventions do appear to influence student outcomes (Juel & Roper/Schneider, 1985; Lesgold et al., 1985). Juel and Roper/Schneider compared the achievement of first-grade students who received the same type of instruction but read different types of texts. Students who read more decodable texts appeared more likely to use a decoding strategy when they read other texts at three testing points compared with students who read nondecodable texts, who appeared to use an orthographic strategy (i.e., identifying words by recognizing unique letter strings) to read them. This difference was especially evident in the first trimester of first grade, when the decodability of texts was most disparate across text types. Yet even in the third trimester, when decodability ratings of texts were the same for both groups, the students who read texts from the initially less decodable basal program were still more likely to use an orthographic strategy than students from the other group, although there was no longer a significant difference in reading ability. The authors concluded that the texts used for instruction contributed more to which word identification strategies students used than the instructional method. These data provided early and enduring evidence that the types of texts students read have a central role in their reading acquisition.

However, there have been few studies since that have examined how text types affect intervention outcomes. One way to address this lack of evidence is to examine the overall effects of interventions themselves and examine whether there is an association between the type of text used in the intervention and student response to instruction. With this goal of clarifying the contributions of different text types to interventions, we examined studies from the 1990s to the present that tested the value of reading interventions on the reading achievement of students with reading difficulty in kindergarten through Grade 3 using community-sourced data from three recent meta-analyses.

Uses of Decodable and Nondecodable Texts and Their Role in Instruction

The issue of decodability in text has been a topic of extensive and intensive discussion as the decades-long reading wars have shifted the emphasis multiple times (Pearson, 2004). Widely used beginning reading programs have undergone numerous overhauls. In some iterations, they have included more decodable texts—that is, texts that contain controlled vocabularies designed to provide extensive practice with specific grapheme-phoneme correspondences (GPCs). The goal of such vocabulary control is to improve students' ability to decode on the premise that the decoding ability promotes independent access to future texts.

In other iterations, reading programs have included more nondecodable texts designed to provide exposure to a wide range of texts and the creativity of language that is often associated with texts written by authors whose goals are to communicate ideas rather than promote controlled practice—often regarded as authentic. Some of the texts that are labeled as authentic are described as leveled books because of the use of qualitative features (e.g., themes, text structures, vocabulary) to assign complexity levels. Books that are also found in libraries and not only in instructional programs—labeled as trade books—also fall under the authentic text rubric. The goal in these nondecodable-weighted iterations of reading programs has been to promote access and exposure to unique ideas and various cultural contexts on the premise that engaging in wide reading will enhance motivation and the use of meaning-based cues to identify unfamiliar words (Pearson, 2004).

The Relation Between Reading Decodable Texts and Reading Acquisition

The findings of Juel and Roper/Schneider's (1985) study suggest that students are more likely to use decoding strategies over the long term if their teachers provide instruction using decodable texts that correspond with the content of code-based lessons rather than selecting leveled trade books that may or may not match lesson content. The consequent inference is that reading decodable texts improves reading skills.

Whether data support that idea is unclear. One challenge is that there is no empirical basis for determining when a text is adequately decodable. Decodable texts do have some distinguishing features. Mesmer (2005) suggested that decodable texts can be identified by two criteria. The first is via a rating of phonetic regularity, in which GPCs are simple, follow common letter-sound associations, and occur frequently in text. The other is via lesson-to-text match (LTTM; Beck, 1997; Stein et al., 1999), in which a high percentage of words in instructional texts follow taught phonetic patterns or were included as sight words in previous lessons. There is no established definition of the "high" in "high percentage," although states have adopted their own standards without evidence (Foorman et al., 2004): California mandated that texts include 75% decodable words to be considered decodable; Texas mandated 80%. By contrast, researchers have explored the association between the level of decodability and student achievement to establish a research-based recommendation.

One study that explored decodability was by Jenkins et al. (2004). The authors tested the effects of two supplemental interventions on the reading outcomes of first graders at risk of reading difficulty—compared with students who participated in typical classroom instruction and received intervention as determined by the schools. For their alternative supplemental interventions, Jenkins et al. created 100 lessons focused on GPC learning and word reading practice and used them in both conditions in a 25-week period. The supplemental interventions differed by the level of decodability for the texts students read in three different phases. For the first, second, and last third of instruction, the more-decodable texts were 85%, 72%, and 80% decodable using the LTTM approach, while the less-decodable texts were 11%, 40%, and 69% decodable. Students in both intervention conditions had better reading outcomes than students in the comparison condition on standardized measures of word recognition, decoding, passage reading fluency and accuracy, spelling, and reading comprehension. However, the students in the intervention groups did not show differences in reading progress for any of the measures. The effect of text decodability on reading outcomes is evidently ambiguous. The fact that there was no difference between intervention groups supports no clear inference—particularly because the study included less decodable texts that were actually in some cases quite decodable.

By contrast, Mesmer (2005) provided stronger evidence in favor of using decodable texts to encourage readers to apply taught letter sound knowledge in decoding words. Mesmer conducted a 14-day study in which first graders with understanding of the alphabetic principle learned a limited set of GPCs in two phonics conditions. In both conditions, students segmented sounds, read and spelled words with taught sounds, and read books. In one condition, students read texts with 40% LTTM, and in the other condition the texts had an 8% LTTM. Students in the 40% LTTM group had greater word reading accuracy (they applied letter-sound knowledge in reading text) and made fewer errors (substitution errors featured two or more phonemes that matched the printed word) at posttest than students in the 8% LTTM group. Despite the differences in the intensity of the intervention and the levels of decodability in the Mesmer and Jenkins et al. (2004) studies, the data do support the possibility that practice reading texts with more decodable words provides some benefit to the reader. (Notably, these empirical data provide no support for the California and Texas mandate levels.)

To summarize the relevant data, Cheatham and Allor (2012) attempted to isolate the effects of decodable texts on reading outcomes and found only seven eligible studies in their review. Of these, three included an intervention and four were descriptive studies of students reading decodable texts. Although results were mixed across both study types, the salient finding of the descriptive studies was that, for the most part, children read more accurately and quickly with better prosody when instruction had a high level of LTTM. Specific to interventions for students with reading difficulty, there was no consistent finding that decodable books led to significantly better performance, but correlations among reading and decodability, novel word reading, and number of graphically similar errors suggested that students who read decodable books used more decoding skills. In short, the evidence whether reading decodable texts supports reading achievement for students with reading difficulty is variable. One conclusion is that there may be a small positive effect for reading decodable texts, but this conclusion is tentative at best.

The Efficacy of Nondecodable Texts in Early Reading Acquisition

Studies that have focused on the relation of nondecodable text to reading performance are similarly limited and have provided only weak evidence of the efficacy of such texts in beginning reading instruction and intervention. Hoffman et al. (2001) found that when end-of-year first graders read texts at kindergarten, early-first grade, and final-first grade levels under three conditions, only the high-achieving readers read with at least 90% accuracy, a criterion identified by Clay (1991) as necessary for the adequate comprehension of text leveled by the Reading Recovery system. In contrast, middle-performing students achieved a mean accuracy level of 90% only with kindergarten-level texts, with 81% and 79% accuracy on first grade texts; low-performing student mean percentages were 70, 58, and 49, which were well short of the word-level accuracy required to comprehend the texts.

Several studies have been conducted on the effects of the Leveled Literacy Intervention program (LLI; Fountas & Pinnell, 2008) on beginning reading skills. LLI is an intervention program that is implemented with texts leveled in a system called Guided Reading Levels (GRLs). This system uses letters to indicate difficulty of a text; kindergarten texts begin at level A and adult-level texts are at level Z. Texts are qualitatively assessed on features such as genre, themes and ideas, sentence complexity, word frequency, and illustrations. Each GRL is assigned a fiction text and a nonfiction text that are used as benchmark assessments (Fountas & Pinnell, 2012).

Ransford-Kaldon et al. (2010) found that kindergarteners, first graders, and second graders successfully read at higher GRL benchmark levels after a semester of reading LLI texts, but results on subtests of Dynamic Indicators of Basic Early Literacy Skills (DIBELS) were mixed. While kindergarteners in the LLI program outperformed control group students in nonword reading fluency, and first graders in the LLI program outperformed control students in nonword reading fluency, letter naming fluency, and oral reading measures, no significant differences were found between groups in initial sound fluency or phoneme segmentation fluency for either grade. Further, intervention group kindergarteners did not outperform control group kindergarteners in letter naming fluency, and no significant differences were found between second grade groups in any of the measures. In a subsequent study, Ransford-Kaldon et al. (2013) found that kindergarteners and first graders moved to higher reading levels, but second graders did not. In addition, treatment groups did not outperform comparison groups on standardized measures of reading.

While both studies resulted in higher reading levels for most treatment groups, the reading skills did not reliably generalize to assessment contexts. This reality, along with the lack of comparison of growth after reading different text types, brings into question the utility of such a program in a wider academic context, in which students are eventually expected to read a variety of fiction and nonfiction texts. This limitation generally accords with the conclusions of Murray et al. (2014) based on their study comparing two first-grade reading interventions for at-risk readers, the code-based My Sidewalks on Reading Street (Juel et al., 2008) and the LLI program (Fountas & Pinnell, 2008). The authors found dramatic differences between the interventions in terms of percentages of polysyllabic words and decodable words using an LTTM approach, raising questions about the value of programs that do not include at least some decodable texts. Murray and colleagues consequently encouraged further examination of text characteristics among intervention programs and of the effects of reading programs with different text types on reading-related skills for students at risk of reading difficulties.

Comparisons of the Effects of the Two Text Types

Recent intervention studies comparing the effects of different text types, either to each other or a control group, have focused on subjects with different age levels and had mixed results. One study (Hiebert & Fisher, 2016) revealed that first-grade English language learners who read phonetically regular texts in addition to receiving regular phonics lessons gained more words correct per minute in an informal reading inventory than those in the control group. In an intervention comparing multi-criteria texts to leveled texts for second graders, Cheatham et al. (2014) found no significant differences in word reading for developing decoders. In a study of children aged 4 to 5, Price-Mohr and Price (2020) found that in tasks of reading comprehension, students who read texts written to include low percentages of decodable words performed significantly better than students who read texts with high percentages of decodable words. For word reading tasks, the difference in performance between the two groups was marginally significant in favor of reading low-decodability texts. In summary, although the research suggests that text-reading practice help students learn to read, there is little evidence to support a focus on either decodable texts or leveled texts.

Rationale for the Current Study

Because of the scarcity of evidence for the effect of text type on the reading achievement of students with reading difficulty, we examined the effects of studies of reading interventions with students from kindergarten through third grade and compared their effects based on whether they included decodable texts, nondecodable texts, or both. Many factors influence the success of reading interventions, such as instructional method, group size, and teacher experience, and we acknowledge the limitation in a study that excludes such factors. We were concerned, however, that examining all potential mediators would lead to an atheoretical analysis that would increase the risk of type I error. We therefore opted to limit the scope of our study to this understudied aspect of reading research to disambiguate the effect of text reading from other intervention effects.

As we began our review of studies, it became apparent that many studies do not report much detail about the characteristics of the texts that their studies included. This is likely the result of editors' admonitions to limit the length of manuscripts, but this does not mitigate the concern that limited information makes reporting difficult and subsequent replication impossible for independent researchers. Given the importance of replication in research (Cook et al., 2016), access to complete information about study materials and procedures is essential. As a result, after exploring the first 10 studies or so, we added a research question to describe the level of information about texts available from the manuscript and its supplemental materials.

Research Questions

Research Question 1: Availability of Information About Texts

For the reason described above, our first question concerned the availability of information to conduct analyses of texts used in interventions. Answering this question is important because the role of text used for successful interventions can be measured with meta-analyses and syntheses only if adequate data are available. Perhaps more important, the possibility of replicating an intervention depends on highly descriptive information about studies. Research Question 1 was as follows: To what extent do authors describe the texts used in reading interventions?

We hypothesized that there would be considerable variability in the extent to which information about the texts was available.

Research Question 2: Variability of Effects Related to Text Types

Our second question, dependent on the first, addressed the question of whether intervention effects differed based on the type of texts used. Answering this question is important to establish a starting point in terms of the field's knowledge about the role of connected text and to motivate future research about that role. Research Question 2 was as follows: Do the findings of reading intervention studies vary based on the type of text used in each study?

Regarding the role of text, we expected that interventions including text would have stronger effects on student skills than interventions without texts (that is, interventions in which students read isolated words or sentences). However, we considered that the effects of text type might vary by the type of instructional outcome. Most of the studies described above examined the effect of text type on word recognition and fluency outcomes. As a result, we decided to examine study outcomes for two separate categories—foundational skills and comprehension skills—to determine whether any text effect would be confined to one type of outcome. One argument for the use of nondecodable text is that they facilitate reading comprehension because they are semantically rich and address meaningful interesting topics. This raised the possibility that the types of texts students read might have different effects for word recognition and reading comprehension outcomes. We addressed this possibility by categorizing measures by their outcome type and examining whether differential effects were present.

In addition, we examined whether the effects were present for researcher-designed or standardized assessments. Interventions often fail to produce significant improvements for standardized reading comprehension outcomes but can show positive effects on more proximal measures (e.g., Fuchs et al., 2018). To address this possibility, we considered the role of measure source in the analysis.

There were no clear hypotheses, but the existing data provided limited support for the possibility that reading decodable texts improves reading outcomes more than reading nondecodable texts—especially on word recognition outcomes.

Method

Identification of Meta-Analyses

The purpose of the current study was to examine the use of texts in the literature focused on the needs of early elementary students who were at risk for or had been identified as having reading disabilities. To facilitate this process, and in alignment with principles of open science—that is, examining and expanding on high-quality previous research—we opted to collect effect size information from published meta-analyses that examined the effects of reading interventions on the reading achievement of students with reading difficulty in kindergarten through third grade. We hoped that, compared with a traditional meta-analysis, this novel method would provide a more efficient means of collecting data on our population of interest and allow for swifter dissemination of the results. Because newer meta-analyses would include the most recent studies and capture older studies that earlier meta-analyses examined, we selected the three most recently published meta-analyses the inclusion criteria for which matched our topic and population of interest. These included Wanzek et al. (2016), Wanzek et al. (2018), and Gersten et al. (2020). This method allowed for an additional layer of peer review via the meta-analyses and avoided duplication of data analysis efforts (e.g., effect size calculation) already undertaken by skilled meta-analysts.

Inclusion Criteria for Studies Within the Meta-Analyses

The three meta-analyses had similar inclusion criteria, indicating that they represented a cohesive but expansive examination of the literature. The inclusion criteria across all meta-analyses were that the studies (a) were published in English-language peer-reviewed journals (Gersten et al. [2020] further specified that studies must have been conducted only in the U.S. between 2002 and 2017), (b) included an effect size for measures of word recognition or comprehension, (c) included participants who were at risk for or diagnosed with reading or learning disabilities, (d) included students from Grades 4 to 6 only if at least 50% of participants were from the target grade levels (see below) or disaggregated data from the target grade levels was available, and (e) provided enough data for effect sizes to be calculated.

The meta-analyses differed in several respects. Wanzek et al. (2016) and Wanzek et al. (2018) included only school-based studies, whereas Gersten et al. (2020) also allowed nonschool clinics as program sites. Both Wanzek et al. reviews included studies of students in kindergarten through third grade; Gersten et al. limited their studies to students in the first through third grade.

Wanzek et al. (2016) included studies with at least 15 and as many as 99 intervention sessions; Gersten et al. specified that the intervention duration should be at least 8 hours.

In contrast to the wider criteria applied by the two Wanzek et al. studies, Gersten et al. (2020) excluded reading instruction that was provided to an entire class of students for grade-appropriate skills (i.e., in Tier 1 general education), interventions designed to support students who had not responded to prior interventions, and pre-reading or phonological awareness interventions in which the instruction did not include activities designed to help students read words or pseudowords. These differences in sampling criteria did not limit the cohesiveness of the studies included.

The differences in inclusion criteria, however, resulted in different numbers of reviewed publications and studies: Wanzek et al. (2016) reviewed 69 publications describing 76 studies; Wanzek et al. (2018) reviewed 24 publications describing 39 studies; and Gersten et al. (2020) reviewed 25 publications describing 37 studies. The number of studies differed from the number of publications (i.e., individual papers) because some studies included multiple intervention groups (resulting in separate contrasts with the comparison condition) or different cohorts in which each was treated as a different study, or multiple projects were reported in one publication. In this paper, we refer to any comparison of an intervention group and a control group with at least one associated effect size as a study. Thus, some publications, as in the meta-analyses, include multiple studies.

Overall, the three meta-analyses provided unique but complementary information, and the study samples all fell within our population of interest. The minor differences in the inclusion criteria were unimportant because the studies collectively allowed us to examine the effect of text exposure on student outcomes in response to instruction.

Additional Criteria for This Study

Some of the study contrasts (n = 11) involved the comparison of multiple intervention groups rather than a researcher-implemented intervention against a no-intervention comparison. Because we were primarily interested in the effects of text exposure in reading interventions in comparison to no-intervention groups, we removed these studies from consideration. There were also studies (n = 21) that appeared in multiple meta-analyses. After removing duplicates, 120 studies described in 99 publications were included in our review.

Coding of Studies

We coded the studies according to two variables that were important for our analysis of texts: (a) the amount of descriptive information available about the texts used in the interventions and (b) the types of texts the researchers used (when this information was available).

Descriptive Information About Texts

To answer Research Question 1 (RQ1), we coded studies based on the availability of information to identify the texts used in the intervention. There were three possible scores for the study reporting to the extent that it would facilitate inclusion in a synthesis or replication of the study:

- 0: There was not enough information about the texts used in the study to determine which text type was used. A common case was that the researchers reported that students read from self-selected or teacher-selected books. Sometimes the report mentioned that books were “leveled,” but authors did not define the range of levels or which leveled book collections were used.

- 1: Texts were partly described in that the researchers provided information that limited the scope of the possible texts but did not provide enough information for complete replicability. For instance, some researchers reported that students read books within a collection but did not specifically state the titles or levels of the books.
- 2: Texts were described with enough detail to specifically identify which ones were used. A common case involved researchers using a published set of books with a limited scope (e.g., *Bob Books*).

Interventions in which the researchers reported that students read no texts were coded *NA* (*not applicable*) to reflect that the text type was irrelevant. Text type was *none* (see *Text Type Codes* below).

We provided up to three ratings for descriptions of intervention texts provided in each study. A rating was determined for each type of text used in the intervention (see *Text Type Codes* below); each study could include up to two text types. We then determined a total description rating for each study based on the rating for each text type. If a study only included one text type (including *none* and *not reported*; n = 110), the rating for the description of that text type was repeated as the total intervention text description rating. The same was true if a study included two text types with the same description rating (e.g., a decodable text with a rating of 2 and a nondecodable text with a rating of 2 would merit a total rating of 2 for that study; n = 6). Studies with mixed text description ratings (e.g., a decodable text with a rating of 2 and a nondecodable text with a rating of 0; n = 4) required a total text description rating of 1; in all such cases, only partial replication of the study would be possible with the given text descriptions.

Text Type Codes

For Research Question 2 (RQ2), we coded the text type with an emphasis on the categories of *decodable* and *nondecodable*. We also coded for no text and missing information; therefore, even studies that were rated 0 on the 0 – 2 scale for RQ1 were coded for text type. These were the codes:

- *Decodable*: Texts were controlled in that the words used in them were limited to those containing patterns and words students had already been taught.
- *Nondecodable:* Texts were not controlled for difficulty in terms of the sound–spelling correspondences they contained. The researchers might have specified that students read leveled books, which are often defined as books bundled based on a publisher’s estimation that they have the same level of difficulty. Texts might have also included trade books with no assigned reading level and other types of texts such as newspapers.
- *None:* There was no connected text for students to read. If researchers did not include reading beyond the word or sentence level in the intervention, we recorded the text type as *none*. These studies scored *NA* for descriptive information.
- *Not reported:* There was not enough information available to ascertain whether texts were used, or it was possible to discern that texts were used but without additional details. These studies scored 0 for descriptive information.

Some interventions included both decodable and nondecodable texts. In these instances, we identified studies as fitting a *decodable and nondecodable* category. Authors of two studies reported using nondecodable texts and additional texts whose types we could not determine. We identified these studies in a *nondecodable and not reported* category. Most interventions included more than one text. We did not code each text separately but rather grouped all individual texts by text type (e.g., all trade books were grouped under the category *nondecodable*).

Multiple Intervention Conditions

Some studies included multiple intervention conditions, each of which was tested relative to a comparison condition. Other studies included multiple cohorts or grade bands of students in which each was treated by the researchers as a separate study. In such cases, we relied on the approach used for the meta-analysis data sources. Wanzek et al. (2016) reported aggregated mean effect sizes for multiple-intervention studies, weighted by sample size of groups. In these cases, we reported studies as single studies. In contrast, Wanzek et al. (2018) and Gersten et al. (2020) reported effect sizes for each intervention group against the comparison condition. For these, we coded each intervention group as a separate study. If the meta-analyses reported effects according to cohort or grade band, or if an original article reported multiple studies, we also treated these as separate studies.

Interrater Reliability

We conducted two tests of interrater reliability for coding of text type and text description. The first author compiled a practice set of 20 studies from the sample, including relevant excerpts from the published papers. The first set of five studies served as an example of how the coding worked and how researchers might describe intervention texts in their publications. In two of the studies, the authors described multiple text types. The second set of five studies served as an initial practice set for the second and third authors, who independently coded each study for text type and description and compared their ratings to those of the first author. The final set of ten studies served as a first test of interrater reliability, in which all three authors independently coded the texts in the studies. Interrater reliability across the three authors was 80%.

The second test of interrater reliability used a random selection of 24 studies (a 20% sample; two studies included multiple text types) that had not been included in the practice sets or first reliability test. The first author coded all studies in the set, and the third author coded half of the set. A research assistant with a Master of Arts in literacy, language, and learning disabilities coded the second half of the set. Interrater reliability between the first and third authors was 100%. Interrater reliability between the first author and the research assistant was 89%.

Effect Sizes

We drew effect sizes for most of the individual studies from Wanzek et al. (2016) and Gersten et al. (2020). Because effect sizes in Wanzek et al. (2018) were aggregated by intervention duration and group size, we requested individual effect sizes from the corresponding author. The effect size data we received did not include sample sizes, so we drew these from the original publications. In studies with different sample sizes for multiple outcome measures, we used the mean sample size as N.

Hedges's g was used to calculate effect size in all three meta-analyses. Effect sizes for one treatment group from Vadasy et al. (2002) were not provided in the Wanzek et al. (2018) effect size data, so we used a Hedges's g calculator (Zach, 2021) to determine effect sizes for that study's outcome measures. Effect sizes for one intervention group in Hurry and Sylva (2007), were also not provided in Wanzek et al. (2016). Because it was unclear which outcome measures were used to calculate the effect sizes for the intervention group that was included in the meta-analysis, we were unable to calculate the missing effect sizes ourselves. We therefore dropped this study from our analysis in RQ2, leaving 119 studies in the effect size analysis.

Categorization of Outcome Measures

Wanzek et al. (2016) calculated effect sizes for standardized and researcher-designed measures of foundational reading skills (including phonological awareness, phonics, word recognition, and fluency) and standardized and researcher-designed measures of vocabulary and reading comprehension. As described above, this method of organization provides useful contrasts, especially in terms of word recognition and comprehension, so we used it to organize the effect sizes from the two other meta-analyses. Gersten et al. (2020) included effect sizes for standardized and researcher-designed measures of word and pseudoword reading, passage fluency, and reading comprehension. We categorized the first two measure types as foundational skills; the latter categorization was in line with Wanzek et al.'s (2016) vocabulary and reading comprehension category. Studies in Wanzek et al. (2018) included standardized and researcher-designed measures of phonological awareness, phonics, word recognition, fluency, vocabulary, and comprehension, in line with the other two meta-analyses, but also included measures of letter and digit identification, spelling, and writing. For consistency, we omitted the effect sizes from the latter measures from our analyses. Both studies by Wanzek et al. included measures of pseudoword reading within the phonics category, so this was consistent with measures included in the Gersten et al. meta-analysis.

Wanzek et al. (2016) aggregated all outcome measures of the same type when calculating effect sizes, but Gersten et al. (2020) reported effect sizes for individual outcome measures. After requesting individual effect sizes from Wanzek et al. (2018), we received a mix of single-measure and aggregated effect sizes. In line with Wanzek et al. (2016), we took the average of all individual measure effect sizes for each of the four measure categories for studies from the other two meta-analyses. We then calculated overall mean effect sizes (N = 119) and mean effect sizes for each of the four measure categories (N = 218), all of which were weighted by sample size and organized by text type.

Robust Variance Estimation

The studies reported here contained different numbers of effect sizes following Wanzek et al. (2016) as described above. Studies ranged from reporting one mean effect size to as many as four—one in each of the four categories of measure type and skill type. As a result, reporting effect sizes without controlling for the nesting of effect sizes within study would bias the results toward studies reporting more effect sizes. Moreover, RVE provides confidence intervals for the effect sizes to allow the most accurate reporting of the reliability of the effects. To answer our questions, we used moderators of interest as covariates in models in which the effect size (g) was the dependent variable. We used the robumeta command for Stata (Hedberg, 2014).

Outliers

Across all effect sizes (N = 218) within the four outcome measure categories, we found several figures that could be considered outliers. We conducted Rosner's test of outliers within the standardized foundational (n = 107), researcher-designed foundational (n = 42), and standardized language and comprehension (*n* = 62) measure categories. The number of effect sizes within the researcher-designed language and comprehension measure category was only seven and thus too small to consider outliers. We found no outliers among the standardized language and comprehension measure effect sizes and three outliers each among the standardized foundational and researcher-designed foundational measure effect sizes. The number of effect sizes within each of the two latter measure categories was further reduced by text type groupings, however. For groups in which the number of effect sizes was less than 30, we determined that the sample sizes were too small to consider outliers. For studies in which *n* was greater than 30 (standardized measure effect sizes from decodable-only studies and nondecodable-only studies, *n* = 33 for both groups), we winsorized effect sizes to the 5th and 95th percentiles. Because the participant sample sizes of the studies with the potential outliers in these groups (all *N*s < 60) were well below the mean *N* of 178, winsorizing these figures resulted in negligible changes to group mean effect sizes. We therefore present unaltered effect size data in the next section.

Results

RQ1: Availability of Information About Texts

Table 1 shows the frequencies and percentages of text description ratings by text type for RQ1. Of the 120 studies, 106 (88%) included at least one text in the intervention and were rated for text descriptions, while 14 studies (12%) did not used connected text. These studies did not receive a rating on the 0 – 2 scale and were coded *not applicable*. Among the remaining 106 studies, 29% (*n* = 31) cited the primary sources of intervention texts, allowing for reasonable replication of the intervention. More than one in five studies (*n* = 23) included enough information for partial replication. Of these, 19 studies received a rating of 1 on the 0 – 2 scale for all text descriptions and four studies received different ratings for descriptions of multiple text types (e.g., a rating of 2 for decodable texts and a rating of 0 for nondecodable texts; this category is labeled *discrepant* in Table 1), which is equivalent to a rating of 1. About half of the studies (*n* = 52) included so little information from the researchers that replicating the interventions with the same texts would be impossible. Of these 52 studies, most (*n* = 31) provided enough information about the text type that they could be coded *nondecodable* for RQ2, but the remainder were coded *not reported*. The combination of a rating of 0 for RQ1 and *not reported* for RQ2 occurred in 21 of the 120 studies—that is, more than one in six studies reported essentially no information about the texts used in the interventions.

Table 1

Frequencies of Text Description Ratings by Text Type

	Text type								
Rating	D	ND	D & ND	NR	ND & NR[a]	None	Total	%	% without N/A[b]
0		29		21	2		52	43	49
1	12	5	2				19	16	18
2	22	7	2				31	26	29
Discrepant (1)			4				4	3	< 1
Not applicable						14	14	12	
Total	34	41	8	21	2	14	120	100	98

Note. 0 = Descriptions of intervention texts not adequate for study replication; 1 = Descriptions allow for partial replication; 2 = Descriptions allow for replication of intervention; Discrepant = Within a study, more than one text type was used, and the ratings differed by text type. This is equivalent to a total rating of 1. Not applicable = Text description not applicable to studies that did not use connected text in intervention. D = Decodable; ND = Nondecodable; NR = Not reported.

[a] At least one text was identifiable as nondecodable text and at least one text was not described well enough to be coded as either text type.

[b] N = 107. Percentages do not sum to 100 due to rounding.

RQ2: Variability in Effects Related to Text Types

Table 2 shows the weighted effect size means by text type and measure category. Table 3 shows effect sizes for each of the four outcome measure categories as well as the mean effect size for each study. Codes from RQ1 and RQ2 are also included for each study.

Table 2

Mean Effect Size by Text Type, Weighted by Sample Size

	All measures			Standardized				Researcher-designed			
Text type	*k*	%	Overall	*n*	Found.	*n*	Comp.	*n*	Found.	*n*	Comp.
Decodable (D) only	34	29	0.41	33	0.45	25	0.34	11	0.65	1[a]	1.17
Nondecodable (ND) only	40	34	0.39	33	0.38	19	0.37	16	0.46	5	1.10
D and ND	8	7	0.65	8	0.84	6	0.67	2[a]	0.80	1[a]	-0.05
Not reported (NR)	21	18	0.30	20	0.28	9	0.24	3[a]	0.93		
ND and NR[b]	2[a]	2	0.42	2[a]	0.42						
None	14	12	0.54	11	0.58	3[a]	0.67	10	0.50		
Totals[c]	119	102	0.41	107	0.44	62	0.38	42	0.57	7	0.94

Note. Found. = Foundational skills (phonological awareness, phonics, word recognition, fluency) measures; Comp. = Language and comprehension skills measures; Overall = Effect size across all four measure categories. Effect size *N* = 218. *k* = number of studies; *n* = number of effect sizes

[a] Small number of effect sizes; the values are reported to provide descriptive information.

[b] At least one text was identifiable as nondecodable text and at least one text was not described well enough to be coded as either text type.

[c] Total means are weighted by *k* or *n.* Percentages do not sum to 100 due to rounding.

Table 3
Effect Sizes

	Rating				Standardized		Researcher-designed	
Study	RQ1	RQ2	*N*	Mean	Found.	Comp.	Found.	Comp.
Al Otaiba et al. (2005)	1	D	49	0.37	0.37	0.36		
Allor & McCathren (2004) Cohort 1	0	ND	86	0.5		0.5		
Allor & McCathren (2004) Cohort 2	0	ND	157	0.11	0.35	-0.16	0.13	
Baker et al. (2000)	1	ND	84	0.4	0.44	0.35		
Barker & Torgesen (1995)	NA	None	54	0.31	0.45		0.16	
Berninger et al. (2003)	0	ND	125	0.68	0.68			
Berninger et al. (2006)	2	ND	93	0.35	0.35			
Blachman et al. (2004)	1	DND	69	0.65	0.77	0.53		
Brown et al. (2005)	1	DND	83	1.17	1.62	1.14	0.74	
Burns et al. (2003) K	0	NR	20	0.77	0.77			
Burns et al. (2003) Gr1-3	0	NR	236	0.34	0.19	0.49		
Case et al. (2014)	2	D	123	0.61	0.55		0.66	
Center et al. (1995)	0	ND	43	1.01	1.2	0.87	0.97	

	Rating				Standardized		Researcher-designed	
Study	RQ1	RQ2	*N*	Mean	Found.	Comp.	Found.	Comp.
Chapman et al. (2001)	0	ND	46	-0.81	-1.05*	-0.47	-0.92*	
Coyne, Little, et al. (2013)	2	D	162	-0.06	-0.06			
Coyne, Simmons, et al. (2013)	2	D	103	0.4	0.4			
Denton et al. (2014) Explicit instruction	2	D	112	0.42	0.47	0.38		
Denton et al. (2014) Guided reading	1	ND	103	0.23	0.35	0.11		
Denton, Nimon, et al. (2010)	0	ND	422	0.47	0.42	0.51		
Denton, Solari, et al. (2010)	1	D	53	0.02	0.06	0.05	-0.05	
Ehri et al. (2007)	2	D	186	0.67	0.83	0.55	0.64	
Fawcett et al. (2001)	0	NR	87	0.74	0.74			
Fien et al. (2015)	2	D	239	0.38	0.38			
Foy (2009)	0	ND	53	0.73	0.73			
Fuchs et al. (2006)	NA	None	33	1.25	1.64		0.86	
Fuchs et al. (2008)	0	NR	64	0.4	0.4			
Gilbert et al. (2013)	0	ND	212	0.19	0.19			

	Rating				Standardized		Researcher-designed	
Study	RQ1	RQ2	*N*	Mean	Found.	Comp.	Found.	Comp.
Gillon (2000)	NA	None	46	1.04	1.3	0.66	1.17	
Graham, et al. (2002)	NA	None	54	0.57	0.6		0.53	
Gunn et al. (2000)	2	D	204	0.37	0.45	0.3		
Gunn et al. (2005)	2	D	245	0.34	0.35	0.32		
Hagan-Burke et al. (2011)	2	D	206	0.24	0.24			
Hurry & Sylva (2007)	NA	None	135	0.15	0.11		0.19	
Jacob et al. (2016)	0	ND	1166	0.1	0.1	0.1		
Jenkins et al. (2004) Less decodable	2	ND	60	0.62	0.64	0.69	0.54	
Jenkins et al. (2004) More decodable	2	D	59	0.7	0.66	0.86	0.57	
Kerins et al. (2010)	NA	None	20	-0.04	-0.04			
Kyle et al. (2013)	NA	None	31	0.48	0.3		0.65	
Lane (1999)	0	NR	26	0.87	0.87			
Lane et al. (2009) Group 1	0	ND	41	0.96	1.24		0.68	
Lane et al. (2009) Group 2	0	ND	42	0.27			0.27	

	Rating				Standardized		Researcher-designed	
Study	RQ1	RQ2	*N*	Mean	Found.	Comp.	Found.	Comp.
Lane et al. (2009) Group 3	0	ND	43	0.47			0.47	
Lane et al. (2009) Group 4	0	ND	46	0.79	1.02		0.56	
Lee et al. (2012)	0	NR	881	0.08	0.1	0.06		
Lennon & Slesinski (1999)	0	ND	134	0.79	0.9	0.67		
Little et al. (2012)	2	D	90	0.46	0.3	0.63		
Luftig (2003)	1	D	36	1.17				1.17
Marston et al. (1995)	2	D	100	0.37	0.37			
Mathes & Babyak (2001)	1	DND	49	0.69	0.89	0.49		
Mathes et al. (2003)	0	NR	89	0.72	0.87	0.57		
Mathes et al. (2005) Proactive Reading	1	D	151	0.47	0.36	0.57		
Mathes et al. (2005) Responsive Reading	0	ND	151	0.23	0.18	0.29		
May et al. (2016)	0	ND	6888	0.42	0.41	0.42		
McCarthy et al. (1995)	0	ND	38	0.87	0.7		1	0.91
McMaster et al. (2005)	1	D	56	0.23	0.29	0.2	0.19	

	Rating				Standardized		Researcher-designed	
Study	RQ1	RQ2	*N*	Mean	Found.	Comp.	Found.	Comp.
Meier & Invernizzi (2001)	1	ND	55	0.64	0.64			
Miller (2003) Cohort 1	0	ND	48	0.73	0.73			
Miller (2003) Cohort 2	0	ND	64	0.95	0.95			
Morris et al. (2000)	2	ND	86	0.74	0.76	0.74	0.74	
Morris et al. (2012)	2	D	279	0.42	0.45	0.38		
Nelson et al. (2005)	NA	None	36	0.76	0.76			
Nicolson et al. (1999)	0	NR	62	0.6	0.6			
Nielsen & Friesen (2012)	1	ND	28	1		0.37		1.62
O'Connor (2000)	1	D	44	0.63	0.65		0.6	
O'Connor et al. (2010) Difficult condition	0	NR	43	0.6	0.79	0.41		
O'Connor et al. (2010) Independent condition	0	NR	40	0.56	0.61	0.51		
O'Shaughnessy & Swanson (2000)	NA	None	45	0.93	0.73	0.71	1.36	
Osborn et al. (2007)	0	NDNR	306	0.44	0.44			
Papadopoulos et al. (2003)	NA	None	40	0.67	0.67			

	Rating				Standardized		Researcher-designed	
Study	RQ1	RQ2	*N*	Mean	Found.	Comp.	Found.	Comp.
Pericola Case et al. (2010)	1	DND	30	1.01	1.2	0.87	0.97	
Puhalla (2011)	2	ND	44	1.61				1.61
Pullen et al. (2004)	0	ND	47	0.55	0.57		0.53	
Rashotte et al. (2001)	0	NR	48	1.35	0.79	0.97	2.3*	
Reutzel et al. (2012)	1	ND	80	1.1		1.1		
Rimm-Kaufman et al. (1998)	0	NR	42	0.07	0.07			
Ryder et al. (2008)	2	D	24	2.03	1.90*	1.06	3.13*	
Santa & Høien (1999)	0	ND	49	0.82	0.73		0.91	
Savage & Carless (2005)	NA	None	104	0.43			0.43	
Savage et al. (2003)	NA	None	104	0.2			0.2	
Scanlon et al. (2005) Phonological emphasis	0	NR	117	0.46	0.57	0.35		
Scanlon et al. (2005) Text emphasis	0	NR	114	0.42	0.43	0.41		
Schwartz (2005)	0	ND	74	0.5	0.94	0.14	0.41	
Simmons et al. (2011)	2	D	206	0.29	0.69	-0.12		

	Rating				Standardized		Researcher-designed	
Study	RQ1	RQ2	*N*	Mean	Found.	Comp.	Found.	Comp.
Smith et al. (2016) Group 1	0	NR	743	0.16	0.16			
Smith et al. (2016) Group 2	0	NR	729	0.28	0.28			
Smith et al. (2016) Group 3	0	NR	749	0.21	0.21			
Torgesen et al. (1999) Embedded phonics	2	D	68	0.48	0.82	0.14		
Torgesen et al. (1999) PA + synthetic phonics	1	DND	65	0.44	0.45	0.43		
Torgesen et al. (1999) Reg. classroom support	0	NR	69	0.17	0.26	0.08		
Torgesen et al. (2010)	1	D	108	0.53	0.58	0.46	0.55	
Vadasy & Sanders (2008a)	2	D	86	0.45	0.47	0.43		
Vadasy & Sanders (2008b)	2	ND	162	0.04	0.09	-0.01		
Vadasy & Sanders (2009)	2	ND	202	0.2	0.17	0.2	0.24	
Vadasy & Sanders (2010)	2	D	148	0.7	0.66	0.62	0.81	
Vadasy & Sanders (2011)	2	D	89	0.36	0.31	0.2	0.57	
Vadasy et al. (1997)	0	ND	40	0.48	0.48			

	Rating				Standardized		Researcher-designed	
Study	RQ1	RQ2	*N*	Mean	Found.	Comp.	Found.	Comp.
Vadasy et al. (2000)	1	D	46	0.76	0.68		0.83	
Vadasy et al. (2002) Sound Partners only	2	D	29	0.72	1.39	0.05		
Vadasy et al. (2002) Thinking Partners only	2	ND	26	0.14	0.18			0.1
Vadasy et al. (2002) Sound Partners + Thinking Partners	2	DND	42	0.33	0.72			-0.05
Vadasy et al. (2005) Reading practice	2	D	38	0.78	0.76	0.81		
Vadasy et al. (2005) Word study	NA	None	38	0.66	0.69	0.63		
Vadasy et al. (2006a)	2	D	67	0.45	0.61	0.28		
Vadasy et al. (2006b) Study 1	0	ND	31	0.66	0.74	0.54	0.71	
Vadasy et al. (2006b) Study 2	1	DND	21	0.11	0.09	0.12		
Vadasy et al. (2007)	2	DND	43	0.26	0.26			
Vandervelden & Siegel (1997)	NA	None	29	0.81			0.81	
Vaughn et al. (2006)	1	D	39	0.59	0.64	0.53		
Vauras et al. (1999)	0	ND	44	1				1

	Rating				Standardized		Researcher-designed	
Study	RQ1	RQ2	*N*	Mean	Found.	Comp.	Found.	Comp.
Vellutino & Scanlon (2002)	0	NDNR	118	0.38	0.38			
Vellutino et al. (2006)	0	NR	113	0.66			0.66	
Vellutino et al. (2008)	0	NR	113	0.64	0.65		0.62	
Vernon-Feagans et al. (2012)	0	ND	135	-0.15	-0.15			
Wang & Algozzine (2008)	1	D	139	0.22	0.27	0.17		
Wanzek & Vaughn (2008) Study 1	1	D	50	-0.05	-0.03	-0.07		
Wanzek & Vaughn (2008) Study 2	1	D	36	-0.28	0.08	-0.64		
Wise et al. (1999)	0	ND	153	0.77	0.71		0.83	
Wright & Jacobs (2003)	0	ND	60	-1.06	-1.06*			
Zvoch & Stevens (2013)	0	NR	93	0.87	0.87			

Note. *N* = number of participants. Found. = Foundational skills (phonological awareness, phonics, word recognition, fluency) measures; Comp. = Language and comprehension skills measures.
*Potential outlier, as discussed in the *Outliers* section.

Preliminary Analysis: Overall Effect of Intervention

Before turning to our second research question on text types, we conducted an analysis of the overall effect of intervention. The mean effect size for all studies (K = 119; N = 218) was g = 0.46, 95% CIs [0.40, 0.52], p < .001. This indicated an overall large effect size (Kraft, 2020) for the reading interventions described in this study.

Effect of Text Type Used in Interventions and Reading Outcomes

Reading Texts vs. Reading No Texts. The first analysis compared the effect for studies in which students read text to those in which they did not. Compared to studies in which students read texts, the effect for studies without text reading were not significantly different, $g = 0.09$, 95% CIs [-0.12, 0.36], $p = .36$. Notably, the difference in magnitude favored the studies without text reading.

Reading Decodable vs. Nondecodable vs. Decodable and Nondecodable vs. No Texts. The second analysis was limited to those studies for which it was possible to code the studies as having (a) decodable texts, (b) nondecodable texts, or (c) both. Studies that included text reading but insufficient information to determine its type were excluded. This analysis included 97 studies and 183 effect sizes. For this analysis, the reference category was for studies that did not include texts. There were no significant differences for studies including text reading—regardless of type—relative to studies without text reading: (a) decodable, $g = -0.12$, 95% CIs [-0.34, 0.10], $p = .27$, (b) nondecodable, $g = 0.10$, 95% CIs [-0.33, 0.14], $p = .41$, and (c) both, $g = 0.28$, 95% CIs [-0.09, 0.65], $p = 0.13$. The confidence intervals for decodable and nondecodable texts revealed no suggestion of an effect. However, the magnitude for the *both* category showed a tendency toward significance with an upper boundary more than seven times greater than the lower boundary. The effect is not reliable, but its magnitude and CIs are noteworthy and provide a rationale for later speculative but possibly interesting interpretation. The effect for both text types was also compared with the effects for decodable and nondecodable texts, but the effects were not reliable, $p = .11$ and $p = .15$, respectively.

Measure Type and Outcome Type. The third analysis involved the disaggregation of the effects by skill type (foundational or comprehension) and measure type (standardized or researcher-designed). For the analysis of foundational skills outcomes ($k = 92$, $n = 126$), the effects for decodable and nondecodable texts were not significant, $p = .41$ and $p = .21$, respectively. The effect for reading both types of texts was significant, however, $g = 0.40$, 95% CIs [0.02, 0.78], $p = .04$. The effect for both vs. decodable texts was $g = 0.49$, 95% CIs [-0.03, 1.02], $p = .07$ and vs. nondecodable texts was $g = 0.55$, 95% CIs [-0.01, 1.09], $p = .05$.

For analysis of measure types, there was a main effect of standardized measures for studies whose text type could be determined ($k = 96$, $n = 184$), $g = -0.19$, 95% CIs [-0.33, -0.05], $p = .01$, in favor of researcher-designed measures. For analysis of researcher-designed measures ($k = 45$, $n = 46$), there were no reliable effects vs. no text reading for decodable, nondecodable, or both, *p*s = .63, .99, and .74, respectively. For analysis of standardized measures for interventions with text reading ($k = 90$, $n = 140$), the effects for studies with decodable and nondecodable texts alone were not noteworthy, *p*s = .20 and .26, respectively. For studies in which students read both types of texts, the effect was significant, $g = 0.45$, 95% CIs [0.02, 0.89], $p = .04$. The comparisons between both and decodable texts also revealed a significant effect, $g = 0.63$, 95% CIs [0.02, 1.24], $p = .04$. The comparison between both and nondecodable texts approached significance, $g = 0.62$, 95% CIs [-0.11, 1.25], $p = .054$.

Given our interest in understanding whether text type and outcome type interacted, we conducted a post hoc analysis of effect sizes for standardized tests of foundational skills only ($k = 87$, $n = 87$). The pattern was almost identical to those observed elsewhere; no evidence of an effect for reading only decodable or nondecodable texts and a significant effect for reading both, $g = 0.51$, 95% CIs [0.00, 1.02], $p = .05$. We were particularly interested to explore the relation between the text type and the researcher-designed outcome measures. However, we could not isolate the researcher-designed measure of reading comprehension, as we had hoped to do. There were only seven effect sizes and cell sizes were too small to evaluate the effects across conditions. Similarly, there were not enough effect sizes to conduct a stable analysis of the relation between foundational skills and researcher-designed measures or comprehension skills and standardized measures.

Summary

For our first question, we observed that there was considerable variability in the clarity with which researchers described the texts in the studies. Only 29% of the studies included enough information to replicate the text reading procedures used in the original studies. In addition, nearly 50% of the studies included only minimal or no information about the texts students read.

For our second question, we found that reading neither decodable texts nor nondecodable texts had a significant effect on student achievement compared with doing no reading and to each other. However, there was convincing evidence across the comparisons suggesting that interventions including both decodable text reading and nondecodable text reading had larger effects than studies without both types of text reading. Data suggested that this effect was especially robust for standardized word recognition outcomes. It was not possible to explore the interesting possibility that nondecodable texts might have had a stronger effect for reading comprehension than decodable texts given small cell sizes. Moreover, the other results suggest that reading both types of texts is likely to produce the greatest improvement.

Discussion

Since Chall (1967) attributed the “great debate” to the topic of text selection in beginning reading, there has been significant discussion about texts that best support beginning reading acquisition. Currently, decodable texts governed by the LTTM perspective and leveled texts based on guided reading levels (Fountas & Pinnell, 2022) represent the two sides of the debate. The features of these text types differ in important ways (Murray et al., 2014). Evidence, however, of the effects of either on students’ reading acquisition has been limited (e.g., Cheatham et al., 2014; Price-Mohr & Price, 2020).

One potential source of information about the effects of text types on reading acquisition can be found in the numerous intervention studies that have been conducted over the past three decades. We used effect sizes to determine if the efficacy of interventions differed as a function of inclusion of text reading and, in cases in which texts were included, of the text type used. Two patterns emerged from our analysis: the first had to do with the effect sizes of different text types on reading acquisition and the second involved the replicability of the text reading components of interventions.

Effects of Text Types

Decodable Texts Relative to Nondecodable Texts

The first finding of note is the similar effect sizes and lack of significant differences among intervention outcomes that used either decodable or nondecodable texts. We had anticipated that the use of decodable texts would result in higher effect sizes than nondecodable texts. This prediction emanated from a consistent finding of reviews that the programs that dominated in the nondecodable group lack a discernible decoding or word recognition curriculum (Cunningham et al., 2005; Hiebert & Tortorelli, 2022; Murray et al., 2014). In particular, we had expected that the effect for decodable texts would be higher on foundational measures, because most interventions focused on decoding support. For researcher-developed measures of foundational skills, the effect for decodable texts was higher than for nondecodable texts, but that was not the case for either standardized foundational or comprehension measures, for which effect sizes for the two text types were in a similar range.

We offer three potential explanations for this pattern of similar outcomes across the two text types. First, any beginning text has a modicum of decodability, as Cheatham and Allor (2012) have observed. The presence of this pattern is demonstrated by one of the studies in the decodable group included in this meta-analyses (Ehri et al., 2007). In that study, a program of nondecodable texts was reordered to fit with a phonics curriculum. Both Ehri et al. and Menon and Hiebert (2005), who similarly reordered the texts in a nondecodable program according to prominent vowel patterns in specific texts, reported significant effects for the group receiving the curriculum-based texts relative to those in the comparison group. Further, the distinction in degree of decodability between decodable and nondecodable texts decreases as the curriculum moves to more complex vowel patterns, as illustrated in the decodability ratings of the Jenkins et al. (2004) study.

A second explanation for the findings may be found in the unit used to establish the decodability of text. In the LTTM model (Beck, 1997; Stein et al., 1999), decodability is based on the individual letter–sound correspondences. A word is judged decodable if all its GPCs have been taught in previous lessons. In the decodable texts of the 1970s, such as those studied by Juel and Roper/Schneider (1985), words with target vowel patterns were the unit of decodability. The words within a similarly sized sample of texts for the first trimester of the Juel and Roper/Schneider study (Economy: Harris et al., 1972) and one in the 2010s (SPIRE: Clark-Edmands, 2013) illustrate the nature of the differences between these two kinds of decodable texts. The decodable text of the 1970s had 81 unique words with short vowel patterns, which were repeated an average of 18 times. The 2010s program had 331 unique words with a mean repetition of six in the 2010s program. That is, today's beginning readers who receive current decodable texts are expected to recognize four times the number of words with target patterns than their peers in the 1970s. Further, beginning readers of the 1970s were likely to have three times as many encounters with a target word than beginning readers in current programs. In describing alphabetic progress, Ehri (2005) explained the need for word repetition. The question of the degree to which beginning readers, especially at-risk students in interventions, can generalize knowledge of individual letter-sound patterns to the levels required in current decodable texts demands attention. To help define what the task of reading entails for early readers, researchers (e.g., Mesmer et al., 2012, Hiebert & Mesmer, 2013) have called for research into primary grade text characteristics, such as levels of decodability, word repetitions, pacing, and sequence.

The third explanation is that teacher training may be contributing disproportionately to reading outcomes. One-to-one interventions are overrepresented among the studies that used only nondecodable texts: 63% of them had a group size of one, compared to 26% of decodable-only studies. Five of the nondecodable-only studies used the Reading Recovery program, a one-to-one tutoring program in which teachers undergo a training period of one academic year (Reading Recovery, 2022). Further, more than 60% of nondecodable-only studies with one-to-one tutoring reported an initial interventionist training period of at least one full day, and more than a quarter reported a yearlong training period. Among decodable-only studies with a group size of one, no training periods exceeded 13 hours, and only a third of these studies included at least a full day of training for interventionists. As a result of more comprehensive training, teachers and tutors may have greater knowledge of reading acquisition and more skills for delivering interventions, contributing more to outcomes than text choice.

Moreover, as Hollands et al. (2016) note, reading programs that are training-intensive, such as Reading Recovery, or that are otherwise costly, are less likely to be faithfully implemented outside of research settings. Because the researchers typically provide funding for the program under examination, resources are more likely to be allocated as intended by the publisher, supporting high-fidelity implementation. Without such funding, schools may not deliver the programs as intended, potentially reducing the effects of the program as implemented. In the present study, it may be that some studies conflate the effect of programs that used nondecodable texts with the effect of high-fidelity implementation. Future studies should include investigations of such differences in teacher training intensity and program fidelity.

Decodable and Nondecodable Combined

The effect sizes among studies in which students read both text types also merit further consideration. The overall mean effect size among these studies was 0.65, and for standardized measures, the effect size was 0.84 for foundational skills and 0.67 for comprehension measures. These effect sizes were the largest of any comparisons. The significant effects of reading both decodable and nondecodable texts on foundational skills and on standardized measures of both foundational and comprehension skills suggest benefits to using multiple text types.

One explanation for the large effects may be the result of the opportunities provided by decodable texts in applying GPCs and those provided by leveled texts in applying high-frequency words. Approximately half of the words in texts overall are accounted for by the 100 most frequent words (Zeno et al., 1995) and a slightly higher percentage (56%) for trade books aimed at beginning readers (Hiebert et al., 2018). Recognition of the 100 most frequent words requires considerable flexibility for beginning readers because 25 have irregular vowel patterns, 12 are multisyllabic, and 14 have a vowel in which a single phoneme is represented by two graphemes (e.g., long-vowels, diphthongs). Of the 35 words for which the correspondence between phonemes and graphemes is one-to-one, about a third have at least one complex consonant cluster (e.g., *wh*, *sh*).

When using the LTTM model, teachers are advised to treat high-frequency words, including those with regular but not-yet-taught GPCs, as sight words. Yet it is unclear how beginning readers distinguish between words for which they are expected to apply decoding skills and those that they have been guided to remember by sight. Leveled texts may provide a context, however, in which young readers can read high-frequency words without encountering the many unique words that are typical of LTTM-aligned decodable texts—that is, readers may be exposed to more word repetitions and fewer unique words in leveled texts (Foorman et al., 2004). Descriptive analyses of leveled texts have suggested a relatively higher percentage of highly frequent and familiar words than texts based on the LTTM model (Hiebert & Tortorelli, 2022). This exposure to high-frequency and other words not typically included in decodable texts may also provide opportunities to generalize decoding and comprehension skills—the goal of reading instruction in general.

Our analysis counterintuitively suggested a significant main effect favoring standardized measures over researcher-designed measures. Despite the tendency for researcher-designed measures to target content from the intervention, the results suggest that students performed better on standardized measures. This could indicate that reading skills learned in this set of interventions (where text type could be determined) were successfully applied to content not targeted in the interventions. In other words, exposure to multiple types of texts might have resulted in more generalization than exposure to a controlled set of texts that only showed effects when the outcomes were designed to align with the types of text.

Interventions Without Text Reading

In a relatively small group of studies (12%), interventions were conducted with no texts. In that this group of studies had average effect sizes that exceeded those that used either decodable or nondecodable texts, this pattern merits attention.

It can be assumed that interventions without connected text are aimed at younger readers to support pre-reading skills. The percentage of kindergartners and first graders in the studies in which text was present was similar to those without connected text (79% for text studies and 75% for no-text studies). If age serves as a proxy for limited reading skill and therefore suggests that younger students have more to gain from reading interventions, we would expect a higher percentage of interventions with connected text to include students at Grade 3 and up. That this was not the case suggests that more research is needed to isolate the effects of word-level instruction on reading skills.

The entry levels of students rather than their ages may contribute to the effectiveness of interventions in which texts were not used. Many studies, however, lacked specificity in their descriptions of students' entry levels, and the researchers opted instead to describe students in the bottom quartile of pretest measures. For students who have not mastered foundational skills, an intervention with such a focus would be expected to produce favorable results.

These results would be expected for foundational skills, which formed the primary outcome in studies without texts. In the two studies for which a standardized comprehension measure was administered, effect sizes were large (an average of 0.65; Kraft, 2020). The small number of studies with comprehension measures as well as the nature of the comprehension measures—sentence tasks—make this finding one that requires replication. Even so, the finding bears consideration when developing theoretical frameworks for the relative amount of time spent on word study relative to text reading at different points in reading acquisition.

Effect of Text Reading Time

We considered the possibility that the effect of text type might be confounded by the amount of time spent reading and conducted a supplemental analysis to address this possibility. We coded the number of minutes spent reading where it was possible to calculate this value (k = 59; n = 119) and conducted several analyses. First, we examined the main effect of time. There was no overall effect of time, $g < 0$, $g > -0.01$, $p = .28$ (see also descriptive statistics in Table 4; note that we were able to calculate time for some studies for which the text type was not reported). This pattern was the same when excluding studies without any text reading, $k = 45$, $n = 91$; $p = .24$, or studies for which the text type could not be ascertained, $k = 50$, $n = 99$; $p = .42$. Second, we examined effects related to text type (*decodable*, *nondecodable,* and *both* using *none* as the reference category) and controlling for time. There was no main effect for time, $k = 45$, $n = 91$, $p = .31$. Finally, we attempted to analyze interaction effects, e.g., for decodable texts, whether the effect might be greater for studies in which students read decodable texts for longer periods of time. It was not possible to obtain reliable estimates for these models, so we could not examine this possibility.

Replicability of Text Reading Components of Studies

A pattern that became evident early in our analysis was the brevity of descriptions of texts for many of the researchers' reports of their interventions. In 19% of the 120 studies, descriptions of at least one text type used in interventions were insufficient for categorization. If texts are central to the act of reading, the specification of the texts used in interventions should be a primary concern. In the Defining Document of a consortium of organizations committed to the science of reading, replication was identified as one of three hallmarks of scientific research (The Reading League, 2022).

Furthermore, replicability, according to the National Academies of Sciences, Engineering, and Medicine (2019), is required to confirm previous results. For replication, the authors of the original publication should ensure that other researchers can follow the same methods as they obtain new data and determine if the results of the replication are consistent with the original study. According to the National Academy of Sciences, "Reports should include details appropriate for the type of research, such as a clear description of all methods, instruments, materials, procedures, measurements, and other variables involved in the study" (p. 108).

Articles for which the descriptions were adequate to determine the text type and specific texts used in the intervention included details such as text or program name and level, characterizations of words in the text (e.g., high-frequency words, words with taught phonics patterns), and characterizations of whole texts (e.g., natural language, decodable). In some cases (e.g., Jenkins et al., 2004), levels of decodability across text sets were provided. Other descriptions that could aid in replication, especially for interventions in which texts vary by student, could include means across texts for word frequency, type-token ratio, critical word factor (Hiebert & Fisher, 2007), or LTTM. Detailed descriptions of text features can aid in replication in future studies, even when specific intervention texts cannot be named. The ease with which online supplementary materials can be shared means the limitations of space, typical in a journal article, no longer need to restrict researchers in describing the texts they used in interventions. At the same time, particular instructional strategies can make it difficult to clearly designate all the texts used in certain interventions. For example, in some interventions, students choose books from a set or a class library. In such cases, we suggest describing the method by which texts were chosen in lieu of reporting specific texts. If texts are regarded to be a critical part of reading acquisition, researchers should provide greater specification about the texts used. Without such specification, replicability is hampered.

Table 4

Effects of Time Spent Reading

Text type	*k*	*n*	Mean (SD) effect size	Min reading time	Max reading time
Decodable (D) only	15	33	0.50 (0.48)	192	3919
Nondecodable (ND) only	17	33	0.49 (0.49)	226	2880
D and ND	4	9	0.66 (0.44)	360	1590
Not reported (NR)	13	14	0.61 (0.41)	540	1875

ND and NR[a]	8	2	0.66 (0.44)	780	1744
None	2	24	0.41 (0.04)	0	0
Totals	59	115			

Note. *k* = number of studies; *n* = number of effect sizes
[a] At least one text was identifiable as nondecodable text and at least one text was not described well enough to be coded as either text type.

Limitations

As in any research but particularly in meta-analyses that encompass many studies, limitations of study design and reporting should be recognized when interpreting results. First, effect size data was calculated for four separate categories of outcome measures and four text categories. This organization coupled with the incomplete text-type reporting of 23 studies means that the number of studies associated with each effect size was substantially lower than 119, ranging from only one study to 33 studies for each group. This limits our ability to generalize about the mean effect sizes we calculated for categories with few studies. In particular, although the interventions that used both decodable and nondecodable intervention texts generally resulted in higher mean effect sizes, this category included only eight studies, while the categories for studies using a single text type each exceeded 30 studies. Thus, the higher effect sizes could be at least in part due to small study sample sizes. Additionally, none of these eight studies reported the amount of time spent reading either decodable or nondecodable texts, so it is unclear whether effect sizes were affected by the dominance of one text type over the other.

Other variables that were not controlled among all studies will have had an impact on the outcomes of the interventions and on the effect sizes. Although upon further analysis we found no effect of time spent reading, other factors such as type of instruction, group size, or teacher expertise may have influenced the effects of the interventions. One salient consideration is the type of intervention. It is possible that interventions that included only nondecodable texts were more likely to include less focus on foundational skills than studies that included only decodable texts. This is possible, but we do not think it will change our interpretation. The primary reason concerns the definitions of the three categories: (a) reading *decodable* text, (b) reading *nondecodable* text, and (c) *both*. By definition, studies categorized as *decodable* and *both* involve interventions that have some emphasis on GPC learning. These studies would fit into an intervention category perhaps called “code-focused” that would be entirely collinear with the *decodable* and *both* text types.

The potentially challenging category is the *nondecodable* text reading category—because some studies might have involved code-focused instruction but had no reading involving decodable texts. However, we did not find that reading either nondecodable texts or decodable texts alone has a greater effect on student achievement than reading no texts at all. Therefore, for our research questions, the type of intervention is unlikely to be an important factor. Moreover, comprehensive coding of intervention characteristics is beyond the scope of this study. However, it may be an interesting topic for future research—with or without exploration of text-reading factors.

Our decision to use existing meta-analyses could be considered a limitation. Our rationale was that, since the existing studies covered the same interventions of interest to us, this strategy would allow us to depend on those studies for identifying appropriate samples and coding effect sizes—processes that require many steps we judged unnecessary for our work and that we hoped would accelerate the review process. We believe that our approach met our goal, although we still needed to examine the effects in every study ourselves to code them. Our approach might be an example for future researchers whose aims overlap with those of recent meta-analyses—and ours do, and we were fortunate to have at least one study published within a year of conducting our study to support our aims. The current approach might have resulted in missed studies, but we think the number of such studies is likely to be very small given the care taken by the authors of the prior studies. For areas of research such as the topic examined here—efficacy of text types—where evidence is limited, an examination of existing meta-analyses can provide much-needed direction and orientation for future studies. We are encouraged by the work reported here.

Another limitation was that we simply could not include some studies in our analysis because the descriptions of texts were inadequate for coding. The lack of description in numerous studies required us to designate texts as fitting one of two very simple categories. Further nuance—e.g., comparing studies by the level of decodability of the texts or distinguishing the use of nondecodable texts that appeared in published reading programs from nondecodable texts from trade books—was not possible. The lack of detail does not lessen the interest of these effects, but we hope that future meta-analyses may include such coding—on the basis that texts are more thoroughly described.

As a result of the simplicity of coding, variability within both categories may be substantial. Cheatham and Allor (2012) identified two main features that characterize decodable texts: (1) a substantial percentage of phonetically regular words and (2) a substantial percentage of words with letter–sound correspondences that match those included in instruction. However, as Cheatham and Allor have observed, no agreed thresholds exist for either the percentage of words that must be phonetically regular or the percentage of GPCs that must match prior lessons for a text to be designated as decodable. Further, there have been different views of decodable texts. For instance, the decodable texts used by Juel and Roper/Schneider (1985) and those in a decodable program that follows the LTTM model vary considerably in the number of unique words and unit of instruction (individual GPCs or rimes). In terms of nondecodable texts, variability within the category is likely even wider, as it includes all texts not considered decodable.

Conclusion

For reading instruction, texts comprise schools' largest budgetary expense for materials and resources. The type of beginning text also receives considerable attention from policymakers, including state mandates and specifications for text design. Yet the scientific basis for what features of texts are optimal at different points of students' progression to becoming proficient readers is surprisingly limited. Texts have typically been part of the many interventions designed to support students' progression. Yet the features of the texts in these interventions have infrequently been the focus of the careful design of lessons and activities related to phonological awareness and decoding. Our results suggest that students are best served by reading a combination of text types, but we believe that researchers must conduct additional studies that involve systematic manipulation of text type as the factor of interest. Without such studies, the findings reported here—though notable and probably important—are necessarily tentative. We urge readers to be circumspect in their interpretation. Moreover, the difficulty we experienced in attempting to identify text types in interventions specifically for reading is a concern for improving the evaluation and replication of research. Researchers have an obligation to provide thorough reports of the texts used in reading interventions.

References

Al Otaiba, S., Schatschneider, C., & Silverman, E. (2005). Tutor-assisted intensive learning strategies in kindergarten: How much is enough? *Exceptionality: The Official Journal of the Division for Research of the Council for Exceptional Children*, *13*(4), 195-208. https://doi.org/10.1207/s15327035ex1304_2

Allor, J., & McCathren, R. (2004). The efficacy of an early literacy tutoring program implemented by college students. *Learning Disabilities Research and Practice*, *19*(2), 116-129. https://doi.org/10.1111/j.1540-5826.2004.00095.x

Baker, S., Gersten, R., & Keating, T. (2000). When less may be more: A 2-year longitudinal evaluation of a volunteer tutoring program requiring minimal training. *Reading Research Quarterly, 35*(4), 494-519. https://doi.org/10.1598/RRQ.35.4.3

Barker, T. A., & Torgesen, J. K. (1995). An evaluation of computer-assisted instruction in phonological awareness with below average readers. *Journal of Educational Computing Research*, *13*(1), 89-103. https://doi.org/10.2190/TH3M-BTP7-JEJ5-JFNJ

Beck, I. L. (1997). Response to "Overselling phonics". *Reading Today*, *15*(2), 17.

Berninger, V. W., Abbott, R. D., Vermeulen, K., & Fulton, C. M. (2006). Paths to reading comprehension in at-risk second-grade readers. *Journal of Learning Disabilities*, *39*(4), 334-351. https://doi.org/10.1177/00222194060390040701

Berninger, V. W., Vermeulen, K., Abbott, R. D., McCutchen, D., Cotton, S., Cude, J., Dorn, S., & Sharon, T. (2003). Comparison of three approaches to supplementary reading instruction for low-achieving second-grade readers. *Language, Speech & Hearing Services in Schools*, *34*(2), 101-116. https://doi.org/10.1044/0161-1461(2003/009)

Birch, R., Sharp, H., Miller, D., Ritchie, D., & Ledger, S. (2022). *A systematic literature review of decodable and levelled reading books for reading instruction in primary school contexts: An evaluation of quality research evidence*. University of Newcastle Language Alliance for Language, Literature, and Literacy. https://www.newcastle.edu.au/__data/assets/pdf_file/0020/804611/Systematic-Literature-Review-UON-Decodable-and-Levelled-Books-PDF.pdf

Blachman, B. A., Schatschneider, C., Fletcher, J. M., Francis, D. J., Clonan, S. M., Shaywitz, B. A., & Shaywitz, S. E. (2004). Effects of intensive reading remediation for second and third graders and a 1-year follow-up. *Journal of Educational Psychology*, *96*(3), 444-461. https://doi.org/10.1037/0022-0663.96.3.444

Brown, K. J., Morris, D., & Fields, M. (2005). Intervention after grade 1: Serving increased numbers of struggling readers effectively. *Journal of Literacy Research*, *37*(1), 61-94. https://doi.org/10.1207/s15548430jlr3701_3

Burns, M. K., Senesac, B. V., & Symington, T. (2003). The effectiveness of the hosts program in improving the reading achievement of children at-risk for reading failure. *Reading Research and Instruction*, *43*(2), 87-103. https://doi.org/10.1080/19388070409558406

Case, L., Speece, D., Silverman, R., Schatschneider, C., Montanaro, E., & Ritchey, K. (2014). Immediate and long-term effects of tier 2 reading instruction for first-grade students with a high probability of reading failure. *Journal of Research on Educational Effectiveness*, *7*(1), 28-53. https://doi.org/10.1080/19345747.2013.786771

Center, Y., Wheldall, K., Freeman, L., Outhred, L., & Margaret. (1995). An evaluation of reading recovery. *Reading Research Quarterly, 30*(2), 240-263. https://doi.org/10.2307/748034

Chall, Jeanne S. (1967). *Learning to read: The great debate; an inquiry into the science, art, and ideology of old and new methods of teaching children to read, 1910-1965*. McGraw-Hill.

Chapman, J. W., Tunmer, W. E., & Prochnow, J. E. (2001). Does success in the Reading Recovery program depend on developing proficiency in phonological-processing skills? A longitudinal study in a whole language instructional context. *Scientific Studies of Reading*, *5*(2), 141-176. https://doi.org/10.1207/S1532799Xssr0502_2

Cheatham, J. P., & Allor, J. H. (2012). The influence of decodability in early reading text on reading achievement: A review of the evidence. *Reading & Writing*, *25*(9), 2223-2246. https://doi.org/10.1007/s11145-011-9355-2

Cheatham, J. P., Allor, J. H., & Roberts, J. K. (2014). How does independent practice of multiple-criteria text influence the reading performance and development of second graders? *Learning Disability Quarterly*, *37*(1), 3-14. https://doi.org/10.1177/0731948713494016

Clarke-Edmands, S. (2013). *S.P.I.R.E. decodable readers*. School Specialty.

Clay, M. M. (1991). *Becoming literate: The construction of inner control*. Heinemann.

Cook, B. G., Collins, L. W., Cook, S. C., & Cook, L. (2016). A replication by any other name: A systematic review of replicative intervention studies. *Remedial and Special Education*, *37*(4), 223-234. https://doi.org/10.1177/0741932516637198

Coyne, M. D., Little, M., Rawlinson, D., Simmons, D., Kwok, O., Kim, M., Simmons, L., Hagan-Burke, S., & Civetelli, C. (2013). Replicating the impact of a supplemental beginning reading intervention: The role of instructional context. *Journal of Research on Educational Effectiveness*, *6*(1), 1-23. https://doi.org/10.1080/19345747.2012.706694

Coyne, M. D., Simmons, D. C., Hagan-Burke, S., Simmons, L. E., Kwok, O., Kim, M., Fogarty, M., Oslund, E. L., Taylor, A. B., Capozzoli-Oldham, A., Ware, S., Little, M. E., & Rawlinson, D. M. (2013). Adjusting beginning reading intervention based on student performance: An experimental evaluation. *Exceptional Children*, *80*(1), 25-44. https://doi.org/10.1177/001440291308000101

Cunningham, J. W., Spadorcia, S. A., Erickson, K. A., Koppenhaver, D. A., Sturm, J. M., & Yoder, D. E. (2005). Investigating the instructional supportiveness of leveled texts. *Reading Research Quarterly*, *40*(4), 410-427. https://doi.org/10.1598/RRQ.40.4.2

Denton, C. A., Fletcher, J. M., Taylor, W. P., Barth, A. E., & Vaughn, S. (2014). An experimental evaluation of guided reading and explicit interventions for primary-grade students at-risk for reading difficulties. *Journal of Research on Educational Effectiveness*, *7*(3), 268-293. https://doi.org/10.1080/19345747.2014.906010

Denton, C. A., Nimon, K., Mathes, P. G., Swanson, E. A., Kethley, C., Kurz, T. B., & Shih, M. (2010). Effectiveness of a supplemental early reading intervention scaled up in multiple schools. *Exceptional Children*, *76*(4), 394-416. https://doi.org/10.1177/001440291007600402

Denton, C. A., Solari, E. J., Ciancio, D. J., Hecht, S. A., & Swank, P. R. (2010). A pilot study of a kindergarten summer school reading program in high-poverty urban schools. *The Elementary School Journal*, *110*(4), 423-439. https://doi.org/10.1086/651190

Ehri, L. C. (2005). Learning to read words: Theory, findings, and issues. *Scientific Studies of Reading*, *9*(2), 167-188. https://doi.org/10.1207/s1532799xssr0902_4

Ehri, L. C., Dreyer, L. G., Flugman, B., & Gross, A. (2007). Reading Rescue: An effective tutoring intervention model for language-minority students who are struggling readers in first grade. *American Educational Research Journal*, *44*(2), 414-448. https://doi.org/10.3102/0002831207302175

Fawcett, A. J., Nicolson, R. I., Moss, H., Nicolson, M. K., & Reason, R. (2001). Effectiveness of reading intervention in junior school. *Educational Psychology*, *21*(3), 299-312. https://doi.org/10.1080/01443410124914

Fien, H., Smith, J. L. M., Smolkowski, K., Baker, S. K., Nelson, N. J., & Chaparro, E. (2015). An examination of the efficacy of a multitiered intervention on early reading outcomes for first grade students at risk for reading difficulties. *Journal of Learning Disabilities*, *48*(6), 602-621. https://doi.org/10.1177/0022219414521664

Foorman, B. R., Francis, D. J., Davidson, K. C., Harm, M. W., & Griffin, J. (2004). Variability in text features in six grade 1 basal reading programs. *Scientific Studies of Reading*, *8*(2), 167-197. https://doi.org/10.1207/s1532799xssr0802_4

Fountas, I., & Pinnell, G. S. (2008). *Leveled literacy intervention*. Heineman.

Fountas, I. & Pinnell, G. S. (2012). *The F & P Text Level Gradient: Revision to recommended grade-level goals.* Heinemann. http://www.heinemann.com/fountasandpinnell/pdfs/WhitePaperTextGrad.pdf

Fountas, I., & Pinnell, G. S. (2022). *Fountas & Pinnell leveled books website.* https://www.fandpleveledbooks.com/

Foy, J. G. (2009). An effectiveness study of the Bookworm Project: A university-school partnership providing a low-cost, research-based early reading intervention for at-risk kindergartners. *Early Childhood Services*, *3*(4), 281-300.

Fuchs, D., Compton, D. L., Fuchs, L. S., Bryant, J., & Davis, G. N. (2008). Making “secondary intervention” work in a three-tier responsiveness-to-intervention model: Findings from the first-grade longitudinal reading study of the national research center on learning disabilities. *Reading & Writing*, *21*(4), 413-436. https://doi.org/10.1007/s11145-007-9083-9

Fuchs, L. S., Fuchs, D., Hamlet, C. L., Powell, S. R., Capizzi, A. M., & Seethaler, P. M. (2006). The effects of computer-assisted instruction on number combination skill in at-risk first graders. *Journal of Learning Disabilities*, *39*(5), 467-475. https://doi.org/10.1177/00222194060390050701

Fuchs, D., Hendricks, E., Walsh, M. E., Fuchs, L. S., Gilbert, J. K., Zhang Tracy, W., Patton III, S., Davis-Perkins, N., Kim, W., Elleman, A. M., & Peng, P. (2018). Evaluating a multidimensional reading comprehension program and reconsidering the lowly reputation of tests of near-transfer. *Learning Disabilities Research & Practice*, *33*(1), 11-23. https://doi.org/10.1111/ldrp.12162

Gersten, R., Haymond, K., Newman-Gonchar, R., Dimino, J., & Jayanthi, M. (2020). Meta-analysis of the impact of reading interventions for students in the primary grades. *Journal of Research on Educational Effectiveness*, *13*(2), 401-427. https://doi.org/10.1080/19345747.2019.1689591

Gilbert, J. K., Compton, D. L., Fuchs, D., Fuchs, L. S., Bouton, B., Barquero, L. A., & Cho, E. (2013). Efficacy of a first-grade responsiveness-to-intervention prevention model for struggling readers. *Reading Research Quarterly*, *48*(2), 135-154. https://doi.org/10.1002/rrq.45

Gillon, G. T. (2000). The efficacy of phonological awareness intervention for children with spoken language impairment. *Language, Speech & Hearing Services in Schools*, *31*(2), 126-141. https://doi.org/10.1044/0161-1461.3102.126

Goldberg, M., & Goldenberg, C. (2022). Lessons learned? Reading wars, Reading First, and a way forward. *The Reading Teacher 75*(5), 621-630. https://doi.org/10.1002/trtr.2079

Graham, S., Harris, K. R., & Chorzempa, B. F. (2002). Contribution of spelling instruction to the spelling, writing, and reading of poor spellers. *Journal of Educational Psychology*, *94*(4), 669-686. https://doi.org/10.1037/0022-0663.94.4.669

Gunn, B., Biglan, A., Smolkowski, K., & Ary, D. (2000). The efficacy of supplemental instruction in decoding skills for Hispanic and non-Hispanic students in early elementary school. *The Journal of Special Education*, *34*(2), 90-103. https://doi.org/10.1177/002246690003400204

Gunn, B., Smolkowski, K., Biglan, A., Black, C., & Blair, J. (2005). Fostering the development of reading skill through supplemental instruction: Results for Hispanic and non-Hispanic students. *The Journal of Special Education*, *39*(2), 66-85. https://doi.org/10.1177/00224669050390020301

Hagan-Burke, S., Kwok, O., Zou, Y., Johnson, C., Simmons, D., & Coyne, M. D. (2011). An examination of problem behaviors and reading outcomes in kindergarten students. *The Journal of Special Education*, *45*(3), 131-148. https://doi.org/10.1177/0022466909359425

Harris, T. L., Creekmore, M., & Greenman, M. H. (1972). *Keys to reading*. The Economy Company.

Hedberg, E. C. (2014). *ROBUMETA: Stata module to perform robust variance estimation in meta-regression with dependent effect size estimates*. Statistical Software Components.

Hiebert, E. H., & Fisher, C. W. (2007). Critical word factor in texts for beginning readers. *The Journal of Educational Research*, *101*(1), 3-11. https://doi.org/10.3200/JOER.101.1.3-11

Hiebert, E. H., & Fisher, C. W. (2016). *Young children's literacy experiences in home and school (Reading Research Report 16-01)*. TextProject, Inc. https://textproject.org/hiebert-fisher-2016/

Hiebert, E. H., Goodwin, A. P., & Cervetti, G. N. (2018). Core vocabulary: Its morphological content and presence in exemplar texts. *Reading Research Quarterly*, *53*(1), 29-49. https://doi.org/10.1002/rrq.183

Hiebert, E. H., & Mesmer, H. A. E. (2013). Upping the ante of text complexity in the Common Core State Standards: Examining its potential impact on young readers. *Educational Researcher*, *42*(1), 44–51. https://doi.org/10.3102/0013189X12459802

Hiebert, E. H., & Tortorelli, L. S. (2022). The role of word-, sentence-, and text-level variables in predicting guided reading levels of kindergarten and first-grade texts. *The Elementary School Journal*, *122*(4), 557-590. https://doi.org/10.1086/719658

Hoffman, J. V., Roser, N. L., Salas, R., Patterson, E., & Pennington, J. (2001). Text leveling and "little books" in first-grade reading. *Journal of Literacy Research, 33*(3), 507-528. https://doi.org/10.1080/10862960109548121

Hollands, F. M., Kieffer, M. J., Shand, R., Pan, Y., Cheng, H., & Levin, H. M. (2016). Cost-effectiveness analysis of early reading programs: A demonstration with recommendations for future research. *Journal of Research on Educational Effectiveness*, *9*(1), 30-53. https://doi.org/10.1080/19345747.2015.1055639

Hurry, J., & Sylva, K. (2007). Long-term outcomes of early reading intervention. *Journal of Research in Reading*, *30*(3), 227-248. https://doi.org/10.1111/j.1467-9817.2007.00338.x

Jacob, R., Armstrong, C., Bowden, A. B., & Pan, Y. (2016). Leveraging volunteers: An experimental evaluation of a tutoring program for struggling readers. *Journal of Research* *on Educational Effectiveness*, *9*(1), 67-92. https://doi.org/10.1080/19345747.2016.1138560

Jenkins, J. R., Peyton, J. A., Sanders, E. A., & Vadasy, P. F. (2004). Effects of reading decodable texts in supplemental first-grade tutoring. *Scientific Studies of Reading*, *8*(1), 53-85. https://doi.org/10.1207/s1532799xssr0801_4

Juel, C., Paratore, J. R., Simmons, D., & Vaughn, S. (2008). *My sidewalks on Reading Street*. Scott Foresman.

Juel, C., & Roper-Schneider, D. (1985). The influence of basal readers on first grade reading. *Reading Research Quarterly*, *20*(2), 134-152. https://doi.org/10.2307/747751

Kerins, M. R., Trotter, D., & Schoenbrodt, L. (2010). Effects of a tier 2 intervention on literacy measures: Lessons learned. *Child Language Teaching and Therapy*, *26*(3), 287-302. https://doi.org/10.1177/0265659009349985

Kraft, M. A. (2020). Interpreting effect sizes of education interventions. *Educational Researcher*, *49*(4), 241-253.

Kyle, F., Kujala, J., Richardson, U., Lyytinen, H., & Goswami, U. (2013). Assessing the effectiveness of two theoretically motivated computer-assisted reading interventions in the United Kingdom: GG Rime and GG Phoneme. *Reading Research Quarterly*, *48*(1), 61-76. https://doi.org/10.1002/rrq.038

Lane, K. L. (1999). Young students at risk for antisocial behavior: The utility of academic and social skills interventions. *Journal of Emotional and Behavioral Disorders*, *7*(4), 211-223. https://doi.org/10.1177/106342669900700403

Lane, H. B., Pullen, P. C., Hudson, R. F., & Konold, T. R. (2009). Identifying essential instructional components of literacy tutoring for struggling beginning readers. *Literacy Research and Instruction*, *48*(4), 277-297. https://doi.org/10.1080/19388070902875173

Lee, Y. S., Morrow-Howell, N., Jonson-Reid, M., & McCrary, S. (2012). The effect of the Experience Corps® program on student reading outcomes. *Education and Urban Society*, *44*(1), 97-118. https://doi.org/10.1177/0013124510381262

Lennon, J. E., & Slesinski, C. (1999). Early intervention in reading: Results of a screening and intervention program for kindergarten students. *School Psychology Review*, *28*(3), 353-364. https://doi.org/10.1080/02796015.1999.12085970

Lesgold, A., Resnick, L. B., & Hammond, K. (1985). Learning to read: A longitudinal study of word skill development in two curricula. In G. E. Mackinnon, & T. G. Waller (Eds.), *Reading research: Advances in theory and practice (vol. 4)* (pp. 107-138). Academic Press.

Little, M. E., Rawlinson, D., Simmons, D. C., Kim, M., Kwok, O., Hagan-Burke, S., Simmons, L. E., Fogarty, M., Oslund, E., & Coyne, M. D. (2012). A comparison of responsive interventions on kindergarteners' early reading achievement. *Learning Disabilities Research and Practice*, *27*(4), 189-202. https://doi.org/10.1111/j.1540-5826.2012.00366.x

Luftig, R. L. (2003). When a little bit means a lot: The effects of a short-term reading program on economically disadvantaged elementary schoolers. *Reading Research and Instruction*, *42*(4), 1-13. https://doi.org/10.1080/19388070309558393

Marston, D., Deno, S. L., Kim, D., Diment, K., & Rogers, D. (1995). Comparison of reading intervention approaches for students with mild disabilities. *Exceptional Children*, *62*(1), 20-37. https://doi.org/10.1177/001440299506200103

Mathes, P. G., & Babyak, A. E. (2001). The effects of peer-assisted literacy strategies for first-grade readers with and without additional mini-skills lessons. *Learning Disabilities Research and Practice*, *16*(1), 28-44. https://doi.org/10.1111/0938-8982.00004

Mathes, P. G., Denton, C. A., Fletcher, J. M., Anthony, J. L., Francis, D. J., & Schatschneider, C. (2005). The effects of theoretically different instruction and student characteristics on the skills of struggling readers. *Reading Research Quarterly*, *40*(2), 148-182. https://doi.org/10.1598/RRQ.40.2.2

Mathes, P. G., Torgesen, J. K., Clancy-Menchetti, J., Santi, K., Nicholas, K., Robinson, C., & Grek, M. (2003). A comparison of teacher-directed versus peer-assisted instruction to struggling first-grade readers. *The Elementary School Journal*, *103*(5), 459-479. https://doi.org/10.1086/499735

May, H., Sirinides, P. M., Gray, A. & Goldsworthy, H. (2016). *Reading Recovery: An evaluation of the four-year i3 scale-up.* https://repository.upenn.edu/cpre_researchreports/81

McCarthy, P., Newby, R. F., & Recht, D. R. (1995). Results of an early intervention program for first grade children at risk for reading disability. *Reading Research and Instruction*, *34*(4), 273-294. https://doi.org/10.1080/19388079509558187

McMaster, K. L., Fuchs, D., Fuchs, L. S., & Compton, D. L. (2005). Responding to nonresponders: An experimental field trial of identification and intervention methods. *Exceptional Children*, *71*(4), 445-463. https://doi.org/10.1177/001440290507100404

Meier, J. D., & Invernizzi, M. (2001). Book Buddies in the Bronx: Testing a model for America Reads. *Journal of Education for Students Placed at Risk*, *6*(4), 319-333. https://doi.org/10.1207/S15327671ESPR0604_2

Menon, S., & Hiebert, E. H. (2005). A comparison of first graders' reading with little books or literature-based basal anthologies. *Reading Research Quarterly*, *40*(1), 12-38. https://doi.org/10.1598/RRQ.40.1.2

Mesmer, H. A. E. (2005). Text decodability and the first-grade reader. *Reading & Writing Quarterly*, *21*(1), 61-86. https://doi.org/10.1080/10573560590523667

Mesmer, H. A., Cunningham, J. W., & Hiebert, E. H. (2012). Toward a theoretical model of text complexity for the early grades: Learning from the past, anticipating the future. *Reading Research Quarterly*, *47*(3), 235-258.

Miller, S. D. (2003). Partners-in-Reading: Using classroom assistants to provide tutorial assistance to struggling first-grade readers. *Journal of Education for Students Placed at Risk*, *8*(3), 333-349. https://doi.org/10.1207/S15327671ESPR0803_3

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. *Journal of Learning Disabilities*, *45*(2), 99-127. https://doi.org/10.1177/0022219409355472

Morris, D., Tyner, B., & Perney, J. (2000). Early Steps: Replicating the effects of a first-grade reading intervention program. *Journal of Educational Psychology*, *92*(4), 681-693. https://doi.org/10.1037/0022-0663.92.4.681

Murray, M. S., Munger, K. A., & Hiebert, E. H. (2014). An analysis of two reading intervention programs. *The Elementary School Journal*, *114*(4), 479-500. https://doi.org/10.1086/675635

National Academies of Sciences, Engineering, and Medicine. (2019). *Reproducibility and replicability in science*. The National Academies Press. https://www.nationalacademies.org/news/2019/05/new-report-examines-reproducibility-

and-replicability-in-science-recommends-ways-to-improve-transparency-and-rigor-in-research
Nelson, J. R., Benner, G. J., & Gonzalez, J. (2005). An investigation of the effects of a prereading intervention on the early literacy skills of children at risk of emotional disturbance and reading problems. *Journal of Emotional and Behavioral Disorders*, *13*(1), 3-12. https://doi.org/10.1177/10634266050130010101

Nicolson, R. I., Fawcett, A. J., Moss, H., Nicolson, M. K., & Reason, R. (1999). Early reading intervention can be effective and cost-effective. *British Journal of Educational Psychology*, *69*(1), 47-62. https://doi.org/10.1348/000709999157563

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

O'Connor, R. (2000). Increasing the intensity of intervention in kindergarten and first grade. *Learning Disabilities Research and Practice*, *15*(1), 43-54. https://doi.org/10.1207/SLDRP1501_5

O'Connor, R. E., Swanson, H. L., & Geraghty, C. (2010). Improvement in reading rate under independent and difficult text levels: Influences on word and comprehension skills. *Journal of Educational Psychology*, *102*(1), 1-19. https://doi.org/10.1037/a0017488

Osborn, J., Freeman, A., Burley, M., Wilson, R., Jones, E., & Rychener, S. (2007). Effect of tutoring on reading achievement for students with cognitive disabilities, specific learning disabilities, and students receiving Title I services. *Education and Training in Developmental Disabilities*, *42*(4), 467-474.

O'Shaughnessy, T. E., & Lee Swanson, H. (2000). A comparison of two reading interventions for children with reading disabilities. *Journal of Learning Disabilities*, *33*(3), 257-277. https://doi.org/10.1177/002221940003300304

Papadopoulos, T. C., Das, J. P., Parrila, R. K., & Kirby, J. R. (2003). Children at risk for developing reading difficulties: A remediation study. *School Psychology International*, *24*(3), 340-366. https://doi.org/10.1177/01430343030243006

Pearson, P. D. (2004). The reading wars. *Educational Policy*, *18*(1), 216-252. https://doi.org/10.1177/0895904803260041

Pericola Case, L., Speece, D. L., Silverman, R., Ritchey, K. D., Schatschneider, C., Cooper, D. H., Montanaro, E., & Jacobs, D. (2010). Validation of a supplemental reading intervention for first-grade children. *Journal of Learning Disabilities*, *43*(5), 402-417. https://doi.org/10.1177/0022219409355475

Price-Mohr, R., & Price, C. (2020). A comparison of children aged 4–5 years learning to read through instructional texts containing either a high or a low proportion of phonically-decodable words. *Early Childhood Education Journal*, *48*(1), 39-47. https://doi.org/10.1007/s10643-019-00970-4

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., Lane, H. B., & Monaghan, M. C. (2004). Effects of a volunteer tutoring model on the early literacy development of struggling first grade students. *Reading Research and Instruction*, *43*(4), 21-40. https://doi.org/10.1080/19388070409558415

Ransford-Kalson, C. R., Flynt, E. S., Ross, C. L., Franceschini, L., Zoblotsky, T., Huang, Y., & Gallagher, B. (2010). *Implementation of effective intervention: An empirical study to evaluate the efficacy of Fountas & Pinnell's Leveled Literacy Intervention system (LLI).* Center for Research in Educational Policy. https://files.eric.ed.gov/fulltext/ED544374.pdf

Ransford-Kaldon, C. R., Ross, C., Lee, C., Sutton Flynt, E., Franceschini, L., & Zoblotsky, T. (2013). *Efficacy of the Leveled Literacy Intervention system for K–2 urban students: An empirical evaluation of LLI in Denver Public Schools*. Center for Research in Education Policy.

Rashotte, C. A., Macphee, K., & Torgesen, J. K. (2001). The effectiveness of a group reading instruction program with poor readers in multiple grades. *Learning Disability Quarterly*, *24*(2), 119-134. https://doi.org/10.2307/1511068

Reading Recovery. (2022). *Training for teachers.* Reading Recovery Community. https://readingrecovery.org/reading-recovery/training/for-teachers/

Reutzel, D. R., Petscher, Y., & Spichtig, A. N. (2012). Exploring the value added of a guided, silent reading intervention: Effects on struggling third-grade readers' achievement. *The Journal of Educational Research*, *105*(6), 404-415. https://doi.org/10.1080/00220671.2011.629693

Rimm-Kaufman, S., Kagan, J., & Byers, H. (1998). The effectiveness of adult volunteer tutoring on reading among "at risk" first grade children. *Reading Research and Instruction*, *38*(2), 143-152. https://doi.org/10.1080/19388079909558284

Ryder, J. F., Tunmer, W. E., & Greaney, K. T. (2008). Explicit instruction in phonemic awareness and phonemically based decoding skills as an intervention strategy for struggling readers in whole language classrooms. *Reading & Writing*, *21*(4), 349-369. https://doi.org/10.1007/s11145-007-9080-z

Santa, C. M., & Høien, T. (1999). An assessment of early steps: A program for early intervention of reading problems. *Reading Research Quarterly*, *34*(1), 54-79. https://doi.org/10.1598/RRQ.34.1.4

Savage, R., & Carless, S. (2005). *Learning support assistants can deliver effective reading interventions for 'at-risk' children*. Informa UK Limited. https://doi.org/10.1080/0013188042000337550

Savage, R., Carless, S., & Stuart, M. (2003). The effects of rime- and phoneme-based teaching delivered by learning support assistants. *Journal of Research in Reading*, *26*(3), 211-233. https://doi.org/10.1111/1467-9817.00199

Scanlon, D. M., Vellutino, F. R., Small, S. G., Fanuele, D. P., & Sweeney, J. M. (2005). Severe reading difficulties-can they be prevented? A comparison of prevention and intervention approaches. *Exceptionality: The Official Journal of the Division for Research of the Council for Exceptional Children*, *13*(4), 209-227. https://doi.org/10.1207/s15327035ex1304_3

Schwartz, R. M. (2005). Literacy learning of at-risk first-grade students in the Reading Recovery early intervention. *Journal of Educational Psychology*, *97*(2), 257-267. https://doi.org/10.1037/0022-0663.97.2.257

Schwartz, S. (2019). The most popular reading programs aren't backed by science. *Education Week*, *39*(15), 19. https://www.edweek.org/teaching-learning/the-most-popular-reading-programs-arent-backed-by-science/2019/12

Simmons, D. C., Coyne, M. D., Hagan-Burke, S., Kwok, O., Simmons, L. C., Johnson, C., Zou, Y., Taylor, A. B., Mcalenney, A. L., Ruby, M., & Crevecoeur, Y. C. (2011). *Effects of* *supplemental reading interventions in authentic contexts: A comparison of kindergarteners' response*. SAGE Publications. https://doi.org/10.1177/001440291107700204

Smith, J. L. M., Nelson, N. J., Fien, H., Smolkowski, K., Kosty, D., & Baker, S. K. (2016). Examining the efficacy of a multitiered intervention for at-risk readers in grade 1. *The Elementary School Journal*, *116*(4), 549-573. https://doi.org/10.1086/686249

Stein, M., Johnson, B., & Gutlohn, L. (1999). Analyzing beginning reading programs: The relationship between decoding instruction and text. *Remedial and Special Education*, *20*(5), 275-287. https://doi.org/10.1177/074193259902000503

The Reading League. (2022). *Science of reading: Defining guide*. https://www.thereadingleague.org/what-is-the-science-of-reading/

Torgesen, J. K., Wagner, R. K., Rashotte, C. A., Herron, J., & Lindamood, P. (2010). Computer-assisted instruction to prevent early reading difficulties in students at risk for dyslexia: Outcomes from two instructional approaches. *Annals of Dyslexia*, *60*(1), 40-56. https://doi.org/10.1007/s11881-009-0032-y

Torgesen, J. K., Wagner, R. K., Rashotte, C. A., Rose, E., Lindamood, P., Conway, T., & Garvan, C. (1999). Preventing reading failure in young children with phonological processing disabilities: Group and individual responses to instruction. *Journal of Educational Psychology*, *91*(4), 579-593. https://doi.org/10.1037/0022-0663.91.4.579

Vadasy, P. F., Jenkins, J. R., Antil, L. R., Wayne, S. K., & O'Connor, R. E. (1997). Community-based early reading interventions fort a-risk first graders. *Learning Disabilities Research & Practice*, *12*(1), 29-39.

Vadasy, P. F., Jenkins, J. R., & Pool, K. (2000). Effects of tutoring in phonological and early reading skills on students at risk for reading disabilities. *Journal of Learning Disabilities*, *33*(6), 579-590. https://doi.org/10.1177/002221940003300606

Vadasy, P. F., & Sanders, E. A. (2008a). Repeated reading intervention: Outcomes and interactions with readers' skills and classroom instruction. *Journal of Educational Psychology*, *100*(2), 272-290. https://doi.org/10.1037/0022-0663.100.2.272

Vadasy, P. F., & Sanders, E. A. (2008b). Code-oriented instruction for kindergarten students at risk for reading difficulties: A replication and comparison of instructional groupings. *Reading & Writing*, *21*(9), 929-963. https://doi.org/10.1007/s11145-008-9119-9

Vadasy, P. F., & Sanders, E. A. (2009). Supplemental fluency intervention and determinants of reading outcomes. *Scientific Studies of Reading*, *13*(5), 383-425. https://doi.org/10.1080/10888430903162894

Vadasy, P. F., & Sanders, E. A. (2010). Efficacy of supplemental phonics-based instruction for low-skilled kindergarteners in the context of language minority status and classroom phonics instruction. *Journal of Educational Psychology*, *102*(4), 786-803. https://doi.org/10.1037/a0019639

Vadasy, P. F., & Sanders, E. A. (2011). Efficacy of supplemental phonics-based instruction for low-skilled first graders: How language minority status and pretest characteristics moderate treatment response. *Scientific Studies of Reading*, *15*(6), 471-497. https://doi.org/10.1080/10888438.2010.501091

Vadasy, P. F., Sanders, E. A., & Peyton, J. A. (2005). Relative effectiveness of reading practice or word-level instruction in supplemental tutoring. *Journal of Learning Disabilities*, *38*(4), 364-380. https://doi.org/10.1177/00222194050380041401

Vadasy, P. F., Sanders, E. A., & Peyton, J. A. (2006a). Code-oriented instruction for kindergarten students at risk for reading difficulties. *Journal of Educational Psychology, 98*(3), 508-528. https://doi.org/10.1037/0022-0663.98.3.508

Vadasy, P. F., Sanders, E. A., & Peyton, J. A. (2006b). Paraeducator-supplemented instruction in structural analysis with text reading practice for second and third graders at risk for reading problems. *Remedial and Special Education*, *27*(6), 365-378. https://doi.org/10.1177/07419325060270060601

Vadasy, P. F., Sanders, E. A., Peyton, J. A., & Jenkins, J. R. (2002). Timing and intensity of tutoring: A closer look at the conditions for effective early literacy tutoring. *Learning Disabilities Research and Practice*, *17*(4), 227-241. https://doi.org/10.1111/1540-5826.00048

Vadasy, P. F., Sanders, E. A., & Tudor, S. (2007). Effectiveness of paraeducator-supplemented individual instruction. *Journal of Learning Disabilities*, *40*(6), 508-525. https://doi.org/10.1177/00222194070400060301

Vandervelden, M. C., & Siegel, L. S. (1997). Teaching phonological processing skills in early literacy: A developmental approach. *Learning Disability Quarterly*, *20*(2), 63-81. https://doi.org/10.2307/1511215

Vaughn, S., Mathes, P., Linan-Thompson, S., Cirino, P., Carlson, C., Pollard-Durodola, S., Cardenas-Hagan, E., & Francis, D. (2006). Effectiveness of an English intervention for first-grade English language learners at risk for reading problems. *The Elementary School Journal*, *107*(2), 153-180. https://doi.org/10.1086/510653

Vauras, M., Kinnunen, R., & Rauhanummi, T. (1999). The role of metacognition in the context of integrated strategy intervention. *European Journal of Psychology of Education*, *14*(4), 555-569. https://doi.org/10.1007/BF03172979

Vellutino, F. R., & Scanlon, D. M. (2002). The interactive strategies approach to reading intervention. *Contemporary Educational Psychology*, *27*(4), 573-635. https://doi.org/10.1016/S0361-476X(02)00002-4

Vellutino, F. R., Scanlon, D. M., Small, S., & Fanuele, D. P. (2006). Response to intervention as a vehicle for distinguishing between children with and without reading disabilities. *Journal of Learning Disabilities*, *39*(2), 157-169. https://doi.org/10.1177/00222194060390020401

Vellutino, F. R., Scanlon, D. M., Zhang, H., & Schatschneider, C. (2008). Using response to kindergarten and first grade intervention to identify children at-risk for long-term reading difficulties. *Reading & Writing*, *21*(4), 437-480. https://doi.org/10.1007/s11145-007-9098-2

Vernon-Feagans, L., Kainz, K., Amendum, S., Ginsberg, M., Wood, T., & Bock, A. (2012). Targeted reading intervention: A coaching model to help classroom teachers with struggling readers. *Learning Disability Quarterly*, *35*(2), 102-114. https://doi.org/10.1177/0731948711434048

Wang, C., & Algozzine, B. (2008). Effects of targeted intervention on early literacy skills of at-risk students. *Journal of Research in Childhood Education*, *22*(4), 425-439. https://doi.org/10.1080/02568540809594637

Wanzek, J., Stevens, E. A., Williams, K. J., Scammacca, N., Vaughn, S., & Sargent, K. (2018). Current evidence on the effects of intensive early reading interventions. *Journal of Learning Disabilities*, *51*(6), 612-624. https://doi.org/10.1177/0022219418775110

Wanzek, J., & Vaughn, S. (2008). Response to varying amounts of time in reading intervention for students with low response to intervention. *Journal of Learning Disabilities*, *41*(2), 126-142. https://doi.org/10.1177/0022219407313426

Wanzek, J., Vaughn, S., Scammacca, N., Gatlin, B., Walker, M. A., & Capin, P. (2016). Meta-analyses of the effects of tier 2 type reading interventions in grades K-3. *Educational Psychology Review*, *28*(3), 551-576. https://doi.org/10.1007/s10648-015-9321-7

Wise, B. W., Ring, J., & Olson, R. K. (1999). Training phonological awareness with and without explicit attention to articulation. *Journal of Experimental Child Psychology*, *72*(4), 271-304. https://doi.org/10.1006/jecp.1999.2490

Wright, J., & Jacobs, B. (2003). Teaching phonological awareness and metacognitive strategies to children with reading difficulties: A comparison of two instructional methods. *Educational Psychology*, *23*(1), 17-47. https://doi.org/10.1080/01443410303217

Zach. (2021). *Hedges' g calculator.* Statology. https://www.statology.org/hedges-g-calculator/

Zeno, S., Ivens, S. H., Millard, R. T., Duvvuri, R., & Rothkopf, E. Z. (1995). *The educator's word frequency guide*. Touchstone Applied Science Associates.

Zvoch, K., & Stevens, J. J. (2013). Summer school effects in a randomized field trial. *Early Childhood Research Quarterly*, *28*(1), 24-32. https://doi.org/10.1016/j.ecresq.2012.05.002

ALIA PUGH (corresponding author) is a doctoral student in the Department of Educational Psychology at the University of Connecticut, Storrs, USA; email alia.pugh@uconn.edu. She researches the role of text characteristics in beginning reading skills.

DEVIN M. KEARNS is an associate professor in the Department of Educational Psychology at the University of Connecticut, Storrs, USA; email devin.kearns@uconn.edu. He researches reading disability and focuses on designing and testing reading interventions and studying how educational practice, cognitive science, and neuroscience can be linked.

ELFRIEDA H. HIEBERT is the CEO and president of TextProject, Santa Cruz, California, USA; email hiebert@textproject.org. Her research addresses how fluency, vocabulary, and knowledge can be fostered through appropriate texts.