Does Use of Text-to-Speech and Related Read-Aloud Tools Improve Reading Comprehension for Students with Reading Disabilities? A Meta-Analysis

Sarah G. Wood, B.S., **Jerad H. Moxley, M.S.**, **Elizabeth L. Tighe, M.S.**, and **Richard K. Wagner, Ph.D.**
Florida State University

Abstract

Text-to-speech and related read aloud tools are being widely implemented in an attempt to assist students' reading comprehension skills. Read aloud software, including text-to-speech, is used to translate written text into spoken text, enabling one to listen to written text while reading along. It is not clear how effective text-to-speech is at improving reading comprehension. This study addresses this gap in the research by conducting a meta-analysis on the effects of text-to-speech technology and related read aloud tools on reading comprehension for students with reading difficulties. Random effects models yielded an average weighted effect size of ($\bar{d} = .35$, with a 95% confidence interval (CI) of .14 to .56, $p < .01$). Moderator effects of study design were found to explain some of the variance. Taken together, this suggests that text-to-speech technologies may assist students with reading comprehension. However, more studies are needed to further explore the moderating variables of text-to-speech and read aloud tools' effectiveness for improving reading comprehension. Implications and recommendations for future research are discussed.

Keywords

reading comprehension; text-to-speech; reading disabilities; technology; meta-analysis

> Reading comprehension, which is defined as the ability to construct meaning from interacting with a text, is critical for students to succeed in today's educational settings

Correspondence should be addressed to Sarah G Wood, B.S., Department of Psychology, Florida State University, 1107 W. Call Street, Tallahassee, FL 32306-4301, USA. wood@psy.fsu.edu.
Jerad H. Moxley, M.S., Department of Psychology, 1107 W. Call Street, Tallahassee, Florida, 32306-4301, USA., moxley@psy.fsu.edu
Elizabeth L. Tighe, Ph.D., Department of Psychology, Georgia State University, P.O. Box 5010, Atlanta, Georgia, 30302-5010, etighe@gsu.edu
Richard K. Wagner, Ph.D., Department of Psychology, Florida State University, 1107 W. Call Street, Tallahassee, Florida 32306-4301, USA., rkwagner@psy.fsu.edu

[5]Heterogeneity is composed of a both random error and true heterogeneity among the studies. When heterogeneity is assumed to be only due to random error, as done in fixed effect models, and under this assumption grade level accounts for a significant amount of variance. Unlike fixed effect models, random effects models do not assume the same common effect (here the effect of read aloud tools on reading comprehension) and the heterogeneity present is from the heterogeneity among studies and random error. Under the assumptions of random effects model the effect of grade level is no longer seen. The calculation of Q is based on a fixed effects model. If we had assumed the fixed effects model was correct, grade would have accounted for significant heterogeneity QM(1)=1883.06, $p<.01$ with a parameter estimate of $B=-.06$ 95% CI[−.064-,−.058]. The REML model calculates QM based on the true effects accounting for heterogeneity (Viechbauer, 2010). As shown in table 2 the effect of grade is nearly zero under these less biased assumptions.

(Snow, 2002). For students with reading disabilities, reading comprehension is often difficult (Kim, Linan Thompson, & Misquitta, 2012). Most current theories argue that one of the primary causes of reading disabilities is a struggle to decode written text. This has a direct negative effect on reading comprehension by decreasing word reading accuracy and speed. Inefficient decoding may also tax cognitive resources, leaving fewer resources available for comprehension (Smythe, 2005).

Presenting reading material orally in addition to a traditional paper presentation format removes the need to decode reading material, and therefore, has the potential to help students with reading disabilities better comprehend written texts. There are several different technologies for presenting oral materials (e.g., text-to-speech, reading pens, audiobooks). Previously text was available orally through books-on-tape and human readers. More recently, text-to-speech technology has been used widely in educational settings from elementary school through college. Unfortunately, its implementation has outpaced the lagging research base on the effects of using text-to-speech to support comprehension. The research literature is characterized by contradictory results, with some studies reporting improved reading whereas others have not (Dalton & Strangman, 2006; Stetter & Hughes, 2010; Strangman & Hall, 2003). Due to these mixed findings, we wanted to conduct a current review of the literature. The goal of this meta-analysis is to synthesize the research literature on the effects of text-to-speech and related tools for oral presentation of material on reading comprehension for students with reading disabilities.

Individuals with reading disabilities have unexpected significant deficits in reading and its component skills (e.g., decoding, fluency), despite potential educational opportunities. This reading deficit is believed to have a neurobiological basis and is also characterized by a failure to respond to appropriate instruction and intervention (Fletcher, 2009; Lyon, Shaywitz, & Shaywitz, 2003; Wagner, Schatschneider, & Phythian-Sence, 2009).

Common problems for individuals with reading disabilities include inaccurate and slow word reading and reading of connected text, making comprehension challenging (LaBerge & Samuels, 1974; Perfetti, 1985). Mastery of lower-level decoding skills is essential for being able to use higher-level language skills to understand text (Cain, Oakhill, & Bryant, 2004).

Intervention Versus Compensation

Intervention studies seek to improve students' reading skills independent of the technology. In contrast, compensation studies seek to provide students with an assistive tool to help them with their reading. Intervention-oriented studies use text-to-speech tools to improve unassisted reading skills, whereas compensation-oriented studies address the use of text-to-speech tools to compensate for word-level skill deficits and gain access to written material. It is often unclear how to tell when intervention has failed and when students should begin to use compensation (Edyburn, 2007). Although educators would like to improve students basics skills as much as possible, not all students are able to reach proficient reading skills. The goal of compensation is designed to help students access texts. If comprehension problems of students with reading disabilities stem at least in part from decoding deficits, then reducing or eliminating the decoding requirement should improve reading

comprehension. Oral presentation of material including text-to-speech helps eliminate the decoding requirement by reading the words aloud to the student, thus enabling comprehension (Olson, 2000). Text-to-speech and read aloud text presentation tools are used in both interventions and compensation settings. We included both intervention and compensation studies in this review but also added a moderator variable to represent this distinction to determine whether comparable effects emerged from both kinds of studies.

Additionally, some students with reading disabilities are relatively accurate at decoding but are slow readers, and may be unable to keep up with reading assignments. This becomes increasingly true as students move into more advanced course work (i.e., college reading) (Jones, Schwilk, & Bateman, 2012). Text-to-speech may enable students to read material that is beyond their basic word reading level while matching their interests and listening comprehension abilities.

Use of Text-to-Speech and other Read-Aloud Tools in Education

Text-to-speech technology development began in the 1980s and has been rapidly increasing. With technological advances, practitioners and education researchers have begun to use text-to-speech and related tools to help students with reading disabilities. In text-to-speech, the text is paired either with a synthesized computer or human recording of that same text (Biancarosa & Griffiths, 2012). As computer technology has improved over the last 20 years, the prevalence and quality of electronic versions of books and computer software with text-to-speech capabilities have also increased. Text-to-speech software development is occurring worldwide in diverse locations, including the United States, Egypt, and Europe (Smythe, 2005). Some examples of such programs include: DecTalk, ClassMate Reader, Texthelp Read & Write Gold, and Kurzweil 3000 (Berkeley & Lindstrom, 2011). The number of free and easily accessible text-to-speech software programs is increasing (Berkeley & Lindstrom, 2011). There are many statewide implementations, including examples from Kentucky and Missouri (Goddard, Kaplan, Kuehnle, & Beglau, 2007; Hasselbring & Bausch, 2005). These programs commonly contain voice options, custom pronunciation, creation of synthetic audio files, and other assistive tools such as text highlighting (Peters & Bell, 2007). Text-to-speech technology features can greatly impact the users' experience and thus may impact its effectiveness. Some of these features include reading rate, voice type, document tagging (which impacts reading order), and dynamic highlighting. Reading rate can greatly impact the users' experience (Lionetti & Cole, 2004). When users choose text-to-speech software, a consistent comment is the desire to have software that follows natural speech and prosody. These subtle differences of reading rate, voice, document formatting, and dynamic highlighting between text-to-speech systems may impact overall user experience and thus may contribute to the mixed findings in the literature.

In addition to different features of text-to-speech tools, individual differences in users may impact the users' text-to-speech experience. Individuals with clinical diagnoses of reading disabilities are heterogeneous. Some may be more skilled at decoding than others and thus may benefit differently from text-to-speech/read aloud tools. Also, many have co-morbid diagnoses, such as attention deficit/hyperactivity disorder (ADHD), which may impact their

reading performance. In addition, personality and social factors interacting with their disability may either facilitate or decrease text-to-speech use (Parette & Scherer, 2004; Parr, 2012). Finally, having supportive and knowledgeable technical help with the software, particularly at school, increases students' ability to use and integrate this technology effectively into their daily lives (King Sears, Swanson, & Mainzer, 2011; Newton & Dell, 2009). These differences may lead to different compensation patterns, which may impact interactions with read aloud/text-to-speech accommodations.

As previously mentioned, the results of several reviews of the literature have not yielded consistent findings about whether text-to-speech improves comprehension (e.g., Hall, Hughes, & Filbert, 2000; MacArthur, Ferretti, Okolo, & Cavalier, 2001; Moran, Ferdig, Pearson, Wardrop, & Blomeyer, 2008; Stetter & Hughes, 2010; Strangman & Dalton, 2005). More recently, two meta-analyses investigated the compensatory effects of read aloud accommodations on assessments (Buzick & Stone, 2014b; Li, 2014b). These studies investigated read aloud accommodations for students with and without disabilities (focused more broadly than just reading disabilities) on English and math assessments. Comprehensive comparative comments on the two studies can be found in (Buzick & Stone, 2104a; Li, 2014a), however we will provide a brief description of the two studies to illustrate how the current study extends the previous findings.

Li (2014) was focused on two general questions of the effects of read aloud accommodations for students with and without disabilities on reading assesments, and which factors would influence those effects of read-aloud accommodations. Those factors were explored through moderators (e.g., disability status and content area) and effect sizes were compared utilizing hierarchical linear modeling. Therefore, they focused on using these moderators to explain the variability between student groups and topics. Having broad inclusionary criteria such as including both quasi-experimental and experimental studies as well as published and unpublished studies created a large enough pool, which enabled moderator analysis. The results revealed smaller effect sizes for read-aloud accommodations for math assessments compared to reading assessments regardless of student disability status. Additionally, students with disabilities did not benefit differentially on math assessments compared to students without disabilities.

Buzick and Stone (2014) carried out a meta-analysis that addressed the effects of read aloud accommodations for students with and without disabilities on standardized assessments, whether read aloud accommodations improve test scores more for students with disabilities compared to students without disabilities, and what factors contribute to differences between the effects of students with and without disabilities for read aloud accommodations. Their inclusionary criteria were stricter than those used by Li (2014a), resulting in a smaller number of studies included in the meta-analysis. The results supported those of Li (2014a) in that smaller effects were found for math assessments compared to reading assessments regardless of student disability status. Additionally, students with disabilities did not benefit differentially on math assessments compared to students without disabilities.

The two previous meta-analyses included students regardless of disability type and their focus was on large-scale standardized tests that were not specific to reading comprehension.

The focus of the current meta-analysis was the effects of text-to-speech and related tools on reading comprehension for individuals with reading disabilities. Additionally, we considered small n studies, sometimes known as single-case studies, which are commonly used in the field of special education. Recently new methods are available to calculate a standardized effect for single-case studies, which can be combined with between subject studies (Shadish & Lecy, 2015). These methods are now recommended for use as described in a recent paper commissioned by the National Center for Education Research (NCER) and the National Center for Special Education Research (Shadish, Hedges, Horner, & Odom, 2015).

We hope to extend the findings from the previous meta-analyses by: (a) calculating effect sizes for just students with reading disabilities using read aloud tools on reading comprehension assessments and; (b) including additional moderator variables (e.g., delivery method, intervention vs. compensation, amount of reading material presented using the read aloud). Three questions were addressed in this meta-analysis:

1.   What is the average weighted effect size of the use of text-to-speech and related tools on reading comprehension for students with reading disabilities?

2.   Are there identifiable moderators of the effects of text-to-speech and related tools on reading comprehension?

3.   What is the overall quantity and quality of the current research base on students with reading disabilities using text-to-speech software to assist with reading comprehension and are there important gaps in the literature?

Method

Inclusionary Criteria

Inclusionary criteria were developed to reflect the diversity of technology and settings in which participants used oral presentations of text. The below were used in this meta-analysis:

1.   Reading comprehension must be measured at the sentence-, paragraph- or passage-level. Both researcher-designed and norm-referenced assessments were included.

2.   Only studies in which the effect sizes could be calculated for students with dyslexia, reading disabilities or learning disabilities (subtype reading) were included. Due to the diversity and lack of consensus in the definition of reading disabilities (Quinn & Wagner, 2013), an inclusive definition was used. Students could have more than one disability but had to have a reading disability as well.

3.   All studies must include a condition with oral presentation of the reading material. Presentation type may be any of the following: human recorded audio, human readers, or a variety of technology, including synthesized text-to-speech, reading pens or other eBooks with text-to-speech or oral capabilities.

4.   Studies must be reported in English.

Additionally, in order to gain a complete quantitative synthesis with minimal publication bias, this meta-analysis included all types of literature: peer-reviewed studies, unpublished reports, theses, dissertations, conference proceedings, and poster presentations. Because of variability in the grade at which students with reading disabilities are identified, we restricted our meta-analysis to studies of third grade and above, the grade by which most students with reading disabilities have been identified. The search was not limited by quantitative study design (experimental, quasi-experimental or observational (utilization of pre-existing data), or year; however, qualitative studies were excluded.

Search Procedures

Studies were gathered through a comprehensive and systematic search, that utilized multiple methods: searching databases, checking previously published studies' bibliographies, and contacting experts in the text-to-speech research field. The following databases were used: Education Resources Information Center (ERIC), Dissertation and Theses (PQDT), PsycINFO, Linguistic and Language Behavior Abstracts, and ComDisDom provided by Proquest; Education Full Text and Academic Search Complete provided by EBSCO; Web of Science provided by Thomson Reuters; Expanded Academic ASAP provided by GALE; and Google Scholar provided by Google.

Three groups of search terms describing the outcome variable (reading comprehension) and the population of interest (students with reading disabilities/dyslexia) and the oral modality of reading material (read aloud/text-to-speech) were used in various combinations, depending on each database's thesaurus. The topic of "reading comprehension" was searched with the following terms: "*reading comprehension*," or "*comprehension tests*," or "*reading comprehension tests*." Next various terms referring to dyslexia were connected to the reading comprehension topic using the Boolean AND operator including: "*response to intervention*," or "*learning dis**," "*reading disabilities*," "*dyslex**," or "*special education students*," or "*disabilities*." These search statements were then combined using the Boolean AND operator with various terms referring to the pairing of audio and visual representations of the text simultaneously. These terms reflect the diversity of text delivery including text-to-speech technology, human readers, and other read aloud accommodations: "*text-to speech synthesizer*," or "*text to speech*," or "*text-to-speech*," or "*speech synthes**," or "*synthes* speech*" or "*speech synthesizers*," or "*oral reading*," or "*read aloud*," or "*book on tape*," or "*audio books*," or "*assistive technology*," or "*examinations*," or "*academic accommodations*," or "*testing accommodations*." Collectively all of the various search combinations yielded 2,933 records, which were reviewed, and finally duplicates were deleted. Initial screening of the search results involved reading titles and abstracts to assess their fit with this review of text-to-speech technology for individuals with reading disabilities. "Reading disabilities" was broadly defined so as not to exclude potentially useful studies. This review of titles and abstracts excluded 2,874 records, leaving 59 records remaining.

Additionally, "backwards mapping" was utilized by identifying 14 other relevant studies from the references list. Three of these studies did not have the appropriate population or disability categories (Crawford & Tindal, 2004; Koretz & Hamilton, 2000; Randall &

Engelhard, 2010). The 11 remaining studies were retained for further review. Also, experts in the field were contacted for additional studies. This resulted in finding six additional studies, of which three were already included (Boyle et al., 2003; Hodapp, Judas, Rachow, Munn, & Dimmitt, 2007; Roberts, Takahashi, Park, & Stodden, 2013). The last two studies of the six identified were not included because one was not about reading comprehension (Orr & Parks, 2007) and we were not able to extract data on the treatment groups for the other study (Swan, Kuhn, Groff, & Roca, unpublished).

Collectively, all of the search methods yielded 81 articles, which were reviewed in full text. Studies were excluded from the meta-analysis based on the following six exclusionary criteria: qualitative studies, synthesis studies (literature reviews and meta-analyses), multiple interventions without reading comprehension reported separately from other interventions, missing a silent reading condition, not conducted in the students native language, and inaccessible full text-versions of studies. More details are available upon request from the first author about methods, including which specific studies were excluded for which reasons. The remaining 43 records were then reviewed for possible inclusion within the meta-analysis.

Coding Procedure

Before coding began, studies for which the effect sizes could not be calculated were removed (see Note 1). In an attempt to get more information about the studies, we contacted the studies' authors. Next, the following characteristics of the studies were coded: *What was read aloud* (divided between partial versus entire passages, and whether comprehension questions were read aloud); *type of oral modality* (categorized into one of the three terms: synthesized text-to-speech, human readers, recorded human voice); *reading material difficulty* (categorized into at the participants' grade level or reading level); *intervention* (defined as using oral presentation to improve students' basic deficits so they can become better readers independent of technology); *compensation* (defined as utilizing technology as a long-term reading solution, like glasses for fixing vision problems); *ability to repeat oral reading material; decoding* (defined as the students' decoding skill being needed for any part of the passage or comprehension assessment); *participant grade* (defined as student's grade level represented in the sample (see Note 3) and *type of publication* (categorized into either published or unpublished). For between-subject studies, the following three additional characteristics were coded: *randomization* (defined as an equal chance of participants receiving random treatment or control condition); *differential sample attrition*; and *group equivalence* (defined as intervention and control group performing similarly on measured pre-test variables). To ensure coding accuracy a second rater coded 20% of the studies. Overall reliabilities were .95 for total inter-rater agreement and .94 for overall Kappa. The

---

[1]Studies were deleted for the following reasons: (a) studies with less than three participants (n=2) (b) studies with poor matching, defined as having a group score at least two standard deviations better in pretest (n=1); (c) studies that had non-independent samples across studies (e.g., Laitusis et al., 2010; n = 3).; (d) not enough information to estimate correlations for the effect sizes (n=8) and; (e) single-case study designs of alternating treatments and variations on that design as there is not a standardized effect size available to combine them with between subject studies (n=5).
[3]Studies' reports of grade level were inconsistent and we had to code it according to a few assumptions. Initially, if mean grade was reported, that was used, but if instead a range of grades were reported, the midpoint was used. If an age mean was reported, we used the grade that is typically equivalent to that age in the United States. If an age range was reported, we used the grade that is typically equivalent to the midpoint of the age. If the students were reported to be college students, grade was coded as 13.

range of the inter-rater agreement across the codes was .90 to 1.0 (except for decoding at .6). Kappa was 1.0 for all codes except: decoding (.28), intervention (.65), and what was read aloud (.73) respectively. Coding discrepancies were resolved through discussion except for the decoding moderator, which was coded by a third rater. The third rather was in agreement with the first rater with Kappa = 1.0, and not in agreement with the second rater with Kappa =.40. We went with the codes provided by the first and third raters as the second rater appeared to have trouble rating this variable.

Results

Effect size Calculations

Effect sizes were calculated for the 22 remaining studies (see Note 2). We calculated effect sizes for all studies, except single-case studies, using the equations in Cooper, Hedges, and Valentine (2009). We used the standardized mean difference (Hedges g, which is Cohen's d with a correction, noted in this paper as $\bar{d}$) as our effect size measure as it corrects for small sample size bias. For the within-subject studies, the correlations were estimated using other information provided in the publication. Effect sizes that are comparable to those of group studies are available for some case study designs, such as multiple baseline and AB^k designs (Shadish, 2014a; Shadish, 2014b), and these were included where possible.

Data Analyses

The analyses were conducted using the metafor package (Viechtbauer, 2010) as implemented in R. The package reports restricted maximum likelihood (REML: Harville, 1977) estimates of the pooled effect size and also reports several measures of heterogeneity including Q (Higgins & Thompson, 2002) and QE (for the moderator analysis (see Note 4), which we reported. We also estimated the effects of potential publication bias using the trim and fill method (Duval & Tweedie, 2004) as well as funnel plots (Cooper, Hedges, & Valentine, 2009).

Results

The results show that the use of text-to-speech tools has a significant impact on reading comprehension scores with $\bar{d}$ =.35, 95% CI [.14, .56], p <.01 (see Table 1 and Figure 1, respectively). The sampling variance was significant Q (22) = 2051.19 p <.001. Visual confirmation of this heterogeneity using a funnel plot is presented in Figure 1a. A trim and fill analysis suggested that four studies were missing on the left side of the mean effect size. Accounting for these studies, the trim and fill analysis suggested that the effect size would

---

[2](a) Raw data that were provided by researchers were used to calculate the effect size and the variance (n=1); (b) correlations estimated from data presented in the publication in either graphical or written form (n=13); (c) instead of using the difference score the post-test score was used to calculate the effect size as the correlation was not available (n=2); (d) effect size was calculated directly from the information provided (n=7); (e) variances were estimated by the published effect size (n=5); (f) average of two measures of reading comprehension (n=3). Citations of these articles are available upon request.
[4]Meta-regressions with each of moderators variable in the between-subject studies: average grade level, intervention, difficulty, delivery method, decoding needed, partly or fully oral passages, peer reviewed randomized and baseline (only coded for studies which were not randomized). Similarly, within-subject moderator analysis was run for moderators: average grade level, intervention, difficulty, delivery method, decoding needed, partly or fully oral passages, peer reviewed.

be $\bar{d}$ =.24, 95% CI [.02, .45], p=.03. Additionally, a series of sensitivity analyses were conducted.

Our effect size is consistent with the previous meta-analyses (Buzick & Stone, 2014b; Li, 2014b) although with a somewhat smaller average weighted effect size. A possible contributing factor to our smaller average effect size was that we had three within-subject studies with negative effect sizes in our pool that were not in either of the previous meta-analyses (Higgins & Raskind, 1997; Hodapp et al., 2007; Thurlow, Moen, Lekwa, & Scullin, 2010). Without these studies the results are more positive and remain significant ($\bar{d}$ = .49 [. 34, .58], p <.001).

In contrast to previous meta-analyses, our study contained post-secondary and adult subjects, but if we only analyzed studies for K-12 we still obtained similar results ($\bar{d}$ = .36 [. 13, .58] p < .01). Also, the previous meta-analyses included studies measuring reading for students with disabilities published in or after the 2000s. In this meta-analysis, when only the studies published after 2000 were analyzed the average weighted effect size became $\bar{d}$ = . 40 [.17, .64], p <.001. This pattern remained true for both intervention and compensation studies ($\bar{d}$ = .54 [.20, .89], p <.001, . $\bar{d}$ = 35 [.12, .59], p <.001, respectively). Overall, the effect of read aloud tools on reading comprehension found here was significant and positive. However, consistent heterogeneity was found and explored through moderator analyses.

For the moderator analyses, only a single, significant moderator emerged, namely whether the study design was a between- or within-subject study. The average weighted effect size for between-subject studies alone was $\bar{d}$ =.61, 95% CI [.39, .83], p <.001. For within subject studies, the average weighted effect size dropped to $\bar{d}$ =.15, 95% CI [−.13, .43], p >05. Previous meta-analyses asking what is the impact of presentation accommodations (segmented text and read aloud) on high stakes test for students with disabilities also found a lower effect size for within-subject study effect sizes as compared to between-subject studies. Li (2014b) also found this pattern but it did not remain significant after all of the predictors were added into the model.

Discussion

The impact of text-to-speech and related readaloud tools for text presentation on reading comprehension for individuals with reading disabilities was explored in this meta-analysis. Three questions were explored and will be discussed in turn.

Average effect size of the use of text-to-speech on measures of reading comprehension

Our meta-analysis provides a more focused approach than previously taken as we looked exclusively at students with reading disabilities using read aloud tools on reading comprehension assessments. This is motivated by a previous meta-analysis finding that indicated that presentation accommodations (segmented reading and read aloud) had a significant positive effect for students with learning disabilities but did not have an effect for students requiring special education services (Vanchu-Orosco, 2012). Both meta-analyses published in 2014 included many different types of disabilities without distinguishing among them; whereas in the current study we focus exclusively on students with reading

disabilities. Another difference is the previous meta-analysis (Buzick and Stone 2014b), focused mostly on state assessments and other large standardized assessments. Many large-scale assessments test general English skills and not solely reading comprehension. Here, we specifically chose to examine reading comprehension tests. Through this narrow focus on students with reading disabilities and reading comprehension assessments, our model hopes expand the previous studies to provide new information on moderators impacting the effectiveness of read aloud tools on reading comprehension.

The effects of text-to-speech and related read aloud tools indicate that oral presentation of text for students with disabilities helps their reading comprehension test scores. Our finding of an average weighted effect size of .35 is consistent with the recent meta-analysis results on read aloud accommodations for students with disabilities (Buzick & Stone, 2014b; Li, 2014b). Taken together, this meta-analysis is useful as a starting point to begin quantifying the wide range of published and unpublished read aloud literature.

Amount of variability present and identifiable moderators in effect sizes

There was significant heterogeneity present. The moderator of study design between-subject studies versus within-subject studies was found to be significant. Possible reasons for this may include regression to the mean, attrition, and order effects due to lack of counterbalancing treatments (Li, 2014b). Interestingly, none of the other moderators were significant in the random effects model. Potential carry-over or training effects of reading with assistive technology may contribute to these differences. Non-significant moderators included: what was read aloud, type of oral modality, reading material difficulty, intervention, compensation, ability to repeat oral reading material, decoding needed, and grade level. We were surprised that grade level did not moderate the effect of the text-to-speech/read aloud presentation. This is in contrast to the previous meta-analyses, which found grade level as a significant moderator (Buzick & Stone, 2014b). However, depending on how the variation in effect sizes between studies is conceptualized, it impacts the amount of variance accounted for by the grade level moderator (see note 6).

In addition, we were surprised that whether the study used text-to-speech/read aloud tools in an intervention setting did not emerge as a significant moderator. Using the text-to-speech/read aloud tools as intervention is theoretically and practically different than using text-to-speech/read aloud tools in a compensation setting. There were very few studies that were truly interventions despite many studies claiming to be interventions. More research comparing intervention versus compensation approaches needs to be conducted.

The quantity and quality of research on students using text-to-speech and related aloud tools for reading comprehension

Initial search results suggest that this is an active and growing area of research. We found 81 articles for full text review while applying exclusionary criteria. We found a small number of studies from the 1980s and 1990s. The literature is growing and the rate of new studies being conducted is increasing with time. Interestingly, the amount of studies using the computerized text-to-speech read aloud tool is increasing. For example, the studies from the

2010s all used text-to-speech technology. This may reflect the trends of wider access to improved text-to-speech technology.

However, study quality remains a concern with many studies missing key components such as proper controls or inadequate statistical reporting. Additionally, many studies used convenience samples (e.g. Boyle et al. (2003)). It is important to have randomized treatment-controlled studies to make causal inferences about the effect of text-to-speech/read aloud tools in an educational context, thus increasing ecological validity by capturing the diversity of the current classroom. However, there are trade offs between internal and external validity that are difficult to balance. It is ideal to have large sample sizes, but they should also be as homogeneous as possible, allowing the best possible assessment of the read aloud/text-to-speech presentation. However, small sample studies could be useful for identifying students' disabilities, and then allowing intensive study of these individual differences. The downside to them is external validity because you are not able to have a large enough sample size to extend to various types of learning disabilities. However, the studies' high internal validity is beneficial to this field. Taken together, new studies with clear and complete reporting of both between-subject and within-subject study designs will be beneficial.

Limitations and Future Directions

There are a few limitations that are worth noting. First, our study included a relatively small sample size (n = 22 studies); however, this is still larger than previous findings. Second, reading comprehension measures and text-to-speech interventions were diverse and were applied at different intensities. Analyzing the importance of intensity of interventions seems particularly interesting, but due to inconsistent reporting, we were unable to develop a meaningful measure of intensity. Third, placebo effects and motivation have yet to be explored or explicitly controlled for. Currently, there is a lack of randomized control trials, these studies will help provide evidence as to whether and how assistive technology impacts students' reading comprehension test scores.

In future studies, measurements of participants' reading skills such as decoding, working memory, vocabulary, and listening and reading comprehension, should be assessed. Text-to-speech technology features including speech speed, text highlighting, and intensity need to be explicitly addressed. With the advent of system-wide user tracking (e.g., length of use, user preferences) of text-to-speech programs, such as Read & Write Gold and Kurzeweil 3000, users' actions can be recorded. This data can be combined with student characteristics like reading level and disability status as well as their reading test outcomes to see how students should be utilizing the software. This data could then be compared to the students' reading comprehension measures with text-to-speech/read aloud tools in compensatory settings such as a state wide reading comprehension test. This enables large quantities of data to be collected but the quality of the data this monitoring provides is questionable. In summary, future research should investigate the following: features of text-to-speech tools, participant reading characteristics and disabilities, study design considerations, and larger-scale usage studies.

Conclusion

Text-to-speech/read aloud presentation positively impacts reading comprehension for individuals with reading disabilities, with average weighted effect sizes of $\bar{d} = = .35$ ($p < .001$). There is more variability than would be expected due to random error alone. Significant variance was accounted for by design type (e.g., within-subject studies versus between-subject studies). The overall quantity and quality of studies investigating the effectiveness of text-to-speech and related read aloud tools on reading comprehension is increasing. However, more studies are needed to explore diverse types of text presentation, for example different software packages and displays, and reading disability profiles. These future research studies can then be used to explore how students are interacting with these technologies and its impact on reading comprehension. Taken together, our results indicate that text-to-speech/read aloud presentation may help students, although the mechanisms remain unknown. Better study designs are required to control for these potential effects.

Acknowledgments

References

Note: References marked with an asterisk(*) indicates studies included in the meta analysis.

Berkeley S, Lindstrom JH. Technology for the struggling reader: Free and easily accessible resources. Teaching Exceptional Children. 2011; 43(4):48–55.

Berkeley S, Scruggs TE, Mastropieri MA. Reading Comprehension Instruction for Students With Learning Disabilities, 1995—2006: A Meta-Analysis. Remedial and Special Education. 2010; 36(6):423–436. DOI: 10.1177/0741932509355988

Biancarosa G, Griffiths GG. Technology Tools to Support Reading in the Digital Age. The Future of Children. 2012; 22(2):139–160. DOI: 10.1353/foc.2012.0014 [PubMed: 23057135]

*. Bielinski, J., Thurlow, M., Ysseldyke, J., Freidebach, J., Freidebach, M. Read-aloud accommodation: Effects on multiple-choice reading and math items. (Technical report 31). Minneapolis, MN: Univ. of Minnesota, National Center on Educational Outcomes; 2001. Retrieved from http://education.umn.edu/NCEO/OnlinePubs/Technical31.htm

*. Boyle EA, Rosenberg MS, Connelly VJ, Washburn SG, Brinckerhoff LC, Banerjee M. Effects of audio texts on the acquisition of secondary-level content by students with mild disabilities. Learning Disability Quarterly. 2003; 26(3):203.doi: 10.2307/1593652

Buzick HM, Stone EA. A Comment on Li: The same side of a different coin. Educational Measurement: Issues and Practice. 2014a; 33(3):34–35. DOI: 10.1111/emip.12042

Buzick H, Stone EA. A meta-analysis of research on the read aloud accommodation. Educational Measurement: Issues and Practice. 2014b; 33(3):17–30. DOI: 10.1111/emip.12040

Cain K, Oakhill J, Bryant P. Children's reading comprehension ability: concurrent prediction by working memory, verbal ability, and component skills. Journal of Educational Psychology. 2004; 96(1):31–42. DOI: 10.1037/0022-0663.96.1.31

Cooper, H., Hedges, LV., Valentine, JC. The handbook of research synthesis and meta-analysis. Russell Sage Foundation; 2009.

Crawford L, Tindal G. Effects of a read-aloud modification on a standardized reading test. Exceptionality. 2004; 12(2):89–106. DOI: 10.1207/s15327035ex1202_3

Cumming, G. Understanding the new statistics. New York: Routledge; 2012.

Dalton, B., Strangman, N. Improving Struggling Readers' Comprehension Through Scaffolded Hypertexts and Other Computer-Based Literacy Programs. In: McKenna, MC., editor. International handbook of literacy and technology volume two. Vol. 2. 2006. p. 75-92.

Dolan R, Hall TE, Banerjee M, Chun E, Strangman N. Applying principles of universal design to test delivery: The effect of computer-based read-aloud on test performance of high school students with learning disabilities. Journal of Technology, Learning, and Assessment. 2005; 3(7):4–32.

Duval S, Tweedie R. Trim and Fill: A Simple Funnel-Plot-Based Method of Testing and Adjusting for Publication Bias in Meta-Analysis. Biometrics. 2004; 56(2):455–463. http://doi.org/10.1111/j.0006-341X.2000.00455.x.

Edyburn DL. Technology- enhanced reading performance: Defining a research agenda. Reading Research Quarterly. 2007; 42(1):146–152. DOI: 10.1598/RRQ.42.1.7

*. Elkind J, Black MS, Murray C. Computer-based compensation of adult reading disabilities. Annals of Dyslexia. 1996; 46(1):159–186. DOI: 10.1007/BF02648175 [PubMed: 24234271]

*. Fasting RB, Lyster SH. The effects of computer technology in assisting the development of literacy in young struggling readers and spellers. European Journal of Special Needs Education. 2005; 20(1):21–40. DOI: 10.1080/0885625042000319061

Fletcher JM. Dyslexia: The evolution of a scientific concept. Journal of the International Neuropsychological Society. 2009; 15:501–508. DOI: 10.1017/S1355617709090900 [PubMed: 19573267]

*. Fletcher JM, Francis DJ, O'Malley K, Copeland K, Mehta P, Caldwell CJ, et al. Effects of a Bundled Accommodations Package on High-Stakes Testing for Middle School Students with Reading Disabilities. Exceptional Children. 2009; 75(4):447–463. http://doi.org/10.1177/001440290907500404.

*. Fletcher JM, Francis DJ, Boudousquie A, Copeland K. Effects of accommodations on high-stakes testing for students with reading disabilities. Exceptional Children. 2006; 72(2)doi: 10.1093/acref/9780195173697.001.0001/acref-9780195173697

*. Floyd KK, Judge SL. The efficacy of assistive technology on reading comprehension for postsecondary students with learning disabilities. Assistive Technology Outcomes and Benefits. 2012; 8(1):48–64.

Flynn LJ, Zheng X, Swanson HL. Instructing Struggling Older Readers: A Selective Meta- Analysis of Intervention Research. Learning Disabilities Research Practice. 2012; 27(1):21–32. DOI: 10.1111/j.1540-5826.2011.00347.x

Goddard, W., Kaplan, L., Kuehnle, J., Beglau, M. Voice recognition and speech-to-text pilot implementation in primary general education technology-rich emints classrooms. 2007. Retrieved from http://www.emints.org/wp-content/uploads/2012/02/TtS-VRpilot-qualitative.pdf

Hall TE, Hughes CA, Filbert M. Computer assisted instruction in reading for students with learning disabilities: A research synthesis. Education and Treatment of Children. 2000; 23(2):173–193.

Harville DA. Maximum likelihood approaches to variance component estimation and to related problems. Journal of the American Statistical Association. 1977; 72(358):320–338. DOI: 10.1080/01621459.1977.10480998

Hasselbring TS, Bausch ME. Assistive technologies for reading. Learning in the Digital Age. 2005; 63(4):72–75.

Higgins EL, Raskind MH. The compensatory effectiveness of the Quicktionary Reading Pen II on the reading comprehension of students with learning disabilities. Journal of Special Education Technology. 2005; 20:31–40.

*. Higgins EL, Raskind MH. The compensatory effectiveness of optical character recognition: speech synthesis on reading comprehension of postsecondary students with learning disabilities. Learning Disabilities a Multidisciplinary Journal. 1997; 8(2):1–15.

Higgins JPT, Thompson SG. Quantifying heterogeneity in a meta- analysis. Statistics in Medicine. 2002; 21(11):1539–1558. DOI: 10.1002/sim.1186 [PubMed: 12111919]

*. Hodapp, J., Judas, C., Rachow, C., Munn, C., Dimmitt, S. Iowa text reader project year 3: Longitudinal Results. Presented at the 26th Annual Closing the Gap Conference; October 20; Minneapolis, MN. 2007.

Horner RH, Kratochwill TR. Synthesizing Single-Case Research to Identify Evidence-Based Practices: Some Brief Reflections. Journal of Behavioral Education. 2012; 21(3):266–272. http://doi.org/10.1007/s10864-012-9152-2.

Jones, MG., Schwilk, CL., Bateman, DF. Reading by listening: Access to books in audio format for college students with print disabilities. In: Aitken, J.Fairley, J., Carlson, J., editors. Communication Technology for Students in Special Education and Gifted Programs. Hershey, PA: Information Science Reference; 2012. p. 249-272.

Kim W, Linan Thompson S, Misquitta R. Critical factors in reading comprehension instruction for students with learning disabilities: A research synthesis. Learning Disabilities Research & Practice. 2012; 27(2):66–78. DOI: 10.1111/j.1540-5826.2012.00352.x

King Sears ME, Swanson C, Mainzer L. TECHnology and Literacy for Adolescents With Disabilities. Journal of Adolescent & Adult Literacy. 2011; 54(8):569–578. DOI: 10.1598/JAAL.54.8.2

Koretz D, Hamilton L. Assessment of students with disabilities in Kentucky: Inclusion, student performance, and validity. Educational Evaluation and Policy Analysis. 2000; 22(3):255–272. DOI: 10.3102/01623737022003255

*. Kosciolek, S., Ysseldyke, JE. Technical report 28. Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes; 2000. Effects of a reading accommodation on the validity of a reading test. Retrieved [today's date], from http://education.umn.edu/NCEO/OnlinePubs/Technical28.htm

LaBerge D, Samuels SJ. Toward a theory of automatic information processing in reading. Cognitive Psychology. 1974; 6(2):293–323. DOI: 10.1016/0010-0285(74)90015-2

*. Laitusis CC. Examining the impact of audio presentation on tests of reading comprehension. Applied Measurement in Education. 2010; 23(2):153–167. DOI: 10.1080/08957341003673815

*. Lange AA, McPhillips M, Mulhern G, Wylie J. Assistive software tools for secondary-level students with literacy difficulties. Journal of Special Education Technology. 2006; 21(3):13–22.

Lenz AS. Using Single-Case Research Designs to Demonstrate Evidence for Counseling Practices. Journal of Counseling & Development. 2015; 93(4):387–393. http://doi.org/10.1002/jcad.12036.

Li H. Comment on Buzick and Stone: Two Tales of One City. Educational Measurement: Issues and Practice. 2014a; 33(3):31–33. http://doi.org/10.1111/emip.12043.

Li H. The effects of read- aloud accommodations for students with and without disabilities: A meta-analysis. Educational Measurement: Issues and Practice. 2014b; n/a–n/a. doi: 10.1111/emip.12027

Lionetti TM, Cole CL. A comparison of the effects of two rates of listening while reading on oral reading fluency and reading comprehension. Education Treatment of Children. 2004; 27(2):114–129.

*. Lundberg I, Olofsson Å. Can computer speech support reading comprehension? Computers in Human Behavior. 1993; 9:283–293. DOI: 10.1016/0747-5632(93)90012-H

Lyon GR, Shaywitz SE, Shaywitz BA. A definition of dyslexia. Annals of Dyslexia. 2003; 53(1):1–14. DOI: 10.1007/s11881-003-0001-9

MacArthur CA, Ferretti RP, Okolo CM, Cavalier AR. Technology applications for students with literacy problems: A critical review. The Elementary School Journal. 2001; 101(3):273–301.

*. McCove, T. Doctoral dissertation. 2012. The use of the kindle to access textbook resources by secondary students in automotive technology. Retrieved from ProQuest Dissertations and Theses database (UMI No 3555373)

*. Meloy LL, Deville C, Frisbie DA. The effect of a read aloud accommodation on test scores of students with and without a learning disability in reading. Remedial and Special Education. 2002; 23(4):248–255. DOI: 10.1177/07419325020230040801

*. Meyer NK, Bouck EC. The impact of text-to-speech on expository reading for adolescents with LD. Journal of Special Education Technology. 2014; 29(1)

Moran J, Ferdig R, Pearson PD, Wardrop J, Blomeyer R Jr. Technology and reading performance in the middle-school grades: A meta-analysis with recommendations for policy and practice. Journal of Literacy Research. 2008; 40(1):6–58. DOI: 10.1080/10862960802070483

Newton DA, Dell AG. Issues in assistive technology implementation: Resolving at/it conflicts. Journal of Special Education Technology. 2009; 24(1):51–56.

Olson RK. Individual Differences in Gains from Computer-Assisted Remedial Reading. Journal of Experimental Child Psychology. 2000; 77(3):197–235. DOI: 10.1006/jecp.1999.2559 [PubMed: 11023657]

Orr, A., Parks, L. Assisted reading software – teachers tell it like it is. Tech&Learning. 2007 Aug 1. Retrieved November 1, 2014, from http://www.techlearning.com/showArticle.php?articleID=196604599

Parette P, Scherer M. Assistive technology use and stigma. Education Training in Developmental Disabilities. 2004; 39(3):217–226.

Parr M. The future of text-to-speech technology: How long before it's just one more thing we do when teaching reading? Procedia - Social and Behavioral Sciences. 2012; 69:1420–1429. DOI: 10.1016/j.sbspro.2012.12.081

Perfetti, CA. Reading ability. New York, NY: Oxford University Press; 1985.

Peters T, Bell L. Choosing and using text-to-speech software. Computers in Libraries. 2007; 27(2):26–29.

Quinn JM, Wagner RK. Gender differences in reading impairment and in the identification of impaired readers: Results from a large-scale study of at-risk readers. Journal of Learning Disabilities. 2013; (Advance online publication). doi: 10.1177/0022219413508323

Randall J, Engelhard G. Performance of students with and without disabilities under modified conditions: Using resource guides and read-aloud test modifications on a high-stakes reading test. The Journal of Special Education. 2010; 44(2):79–93. DOI: 10.1177/0022466908331045

*. Roberts, K., Takahashi, K., Park, HJ., Stodden, R. Does text-to-speech use improve reading skills of high school students?. Paper presented at the Assistive Technology Industry Association; Orlando, FL. 2013 Jan. Abstract retrieved from: http://www.cds.hawaii.edu/steppingstones/downloads/presentations/pdf/TTS_ATIA2013.pdf

*. Schmitt AJ, Hale AD, McCallum E, Mauck B. Accommodating remedial readers in the general education setting: Is listening-while-reading sufficient to improve factual and inferential comprehension? Psychology in the Schools. 2011; 48(1):37–45. DOI: 10.1002/pits.20540

Shadish WR. Analysis and meta-analysis of single-case designs: An introduction. Journal of School Psychology. 2014a; 52(2):109–122. http://doi.org/10.1016/j.jsp.2013.11.009. [PubMed: 24606971]

Shadish WR. Statistical Analyses of Single-Case Designs The Shape of Things to Come. Current Directions in Psychological Science. 2014b; 23(2):139–146. http://doi.org/10.1177/0963721414524773.

Shadish WR, Hedges LV, Pustejovsky JE. Analysis and meta-analysis of single-case designs with a standardized mean difference statistic: A primer and applications. Journal of School Psychology. 2014; 52(2):123–147. http://doi.org/10.1016/j.jsp.2013.11.005. [PubMed: 24606972]

Shadish WR, Lecy JD. The meta-analytic big bang. Research Synthesis Methods. 2015; 6(3):246–264. http://doi.org/10.1002/jrsm.1132. [PubMed: 26212600]

Shadish, WR., Hedges, LV., Horner, RH., Odom, SL. The Role of Between-Case Effect Size in Conducting, Interpreting, and Summarizing Single-Case Research. Washington, DC: 2015. p. 1-109.(No. NCER 2015-002)

*. Shany MT, Biemiller A. Assisted reading practice: Effects on performance for poor readers in grades 3 and 4. Reading Research Quarterly. 1995; 30(3):382–395.

Smythe, I. Provision and use of information technology with dyslexic students in university in Europe. EU funded project Welsh Dyslexia Project. 2005. Retrieved from http://eprints.soton.ac.uk/264151/1/The_Book.pdf

Snow, CE. Reading for understanding: Toward a research and development program in reading comprehension. Arlington, VA: RAND; 2002. p. 1-174.

*. Sorrell CA, Bell SM, McCallum RS. Reading rate and comprehension as a function of computerized versus traditional presentation mode: A preliminary study. Journal of Special Education Techology. 2007; 22(1):1–12.

Stetter ME, Hughes MT. Computer-assisted instruction to enhance the reading comprehension of struggling readers: A review of the literature. Journal of Special Education Techology. 2010; 25(4):1–16.

Strangman, N., Dalton, B. Using technology to support struggling readers: A review of the literature. In: Edyburn, D., Higgins, K., editors. Handbook of Special Education Technology Research and Practice. Vol. 2. Milwaukee, WI: 2005. p. 545-570.

Strangman, N., Hall, T. Text Transformations. Wakefield, MA: National Center on Accessing the General Curriculum; 2003. Retrieved from http://aim.cast.org/learn/historyarchive/backgroundpapers/text_transformations

Swan A, Kuhn MR, Groff C, Roca J. Evaluating the effectiveness of *Learning Through Listening®* program on the reading abilities of students with learning disabilities. 2005 Unpublished.

Swanson HL. Reading research for students with LD: A meta-analysis of intervention outcomes. Journal of Learning Disabilities. 1999; 32(6):504–532. DOI: 10.1177/002221949903200605 [PubMed: 15510440]

*. Thurlow, ML., Moen, RE., Lekwa, AJ., Scullin, SB. Examination of a reading pen as a partial auditory accommodation for reading assessment. Minneapolis, MN: University of Minnesota, Partnership for Accessible Reading Assessment; 2010.

Vanchu-Orosco, M. Doctoral dissertation. Electronic Theses and Dissertations; Denver, Colorado: 2012. A Meta-analysis of Testing Accommodations for Students with Disabilities: Implications for High-stakes Testing.

Viechtbauer W. Conducting meta-analyses in R with the metafor package. Journal of Statistical Software. 2010; 36(3)

Wagner, RK., Schatschneider, C., Phythian-Sence, C. Beyond decoding. New York, NY: Guilford Press; 2009.

Wendel E, Cawthon SW, Ge JJ, Beretvas SN. Alignment of Single-Case Design (SCD) Research With Individuals Who Are Deaf or Hard of Hearing With the What Works Clearinghouse Standards for SCD Research. Journal of Deaf Studies and Deaf Education. 2015; 20(2):103–114. http://doi.org/10.1093/deafed/enu049. [PubMed: 25672901]
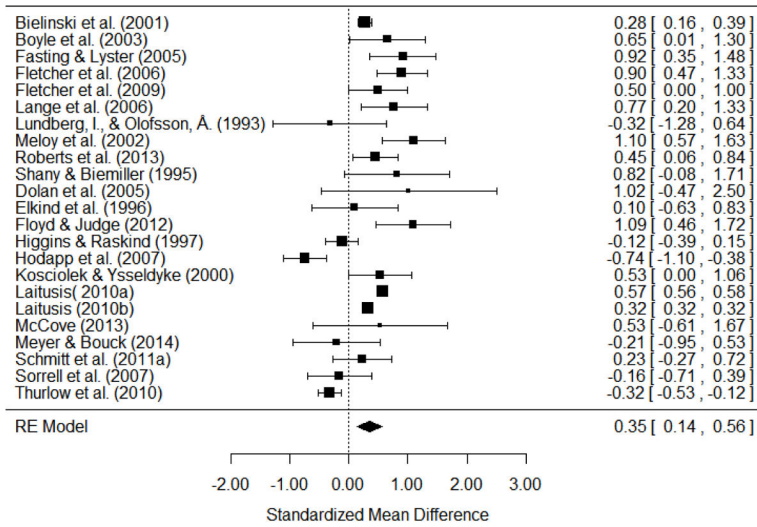
Figure 1.
A) Forest plot of effect sizes.

A)

B)

Figure 2.
Funnel plot showing (A) the heterogeneity of all studies included in the meta-analysis and (B) the residual heterogeneity after controlling for if the study was between subjects or within subjects.

Summary of Included Articles

Table 1

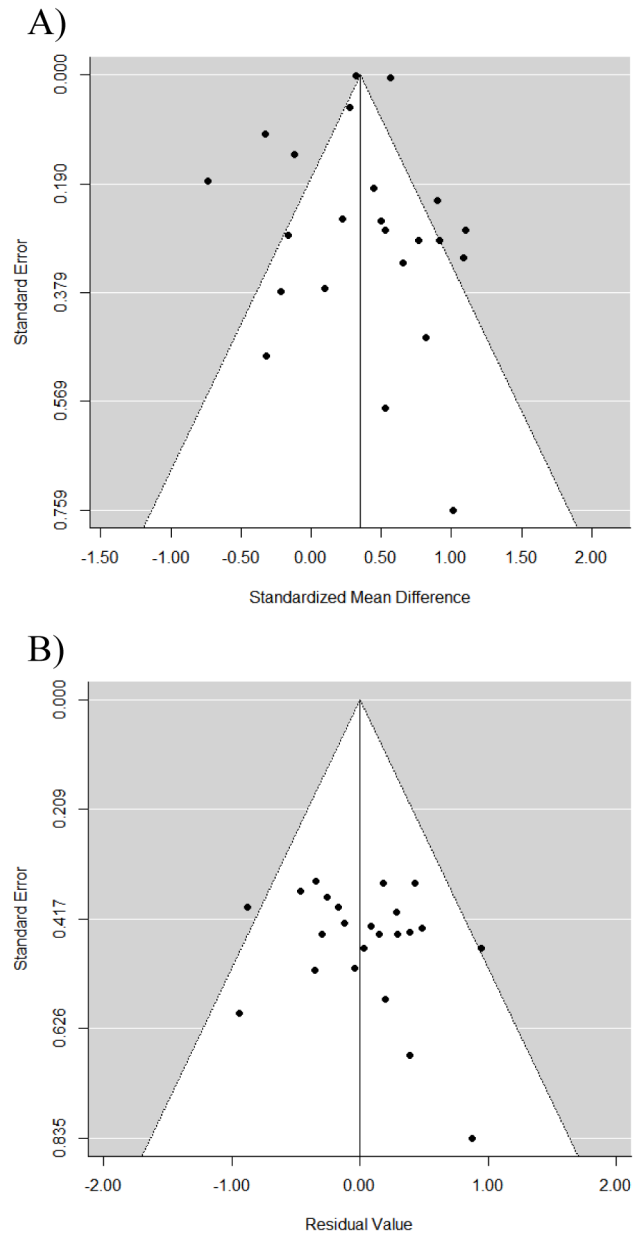| Between-Subject Study Design	d	V	n	Avg. grade	Delivery method	Goal of study	What was read aloud	Reading material difficulty	Decoding needed	Peer reviewed	Measure
Bielinski et al. (2001)	0.28	0.003	1261	3	HR	C	CP&Q/T	GL	No	No	MAP
Boyle et al. (2003)	0.65	0.108	38	10.5	HR&RHV	C	CP&Q/T	GL	No	Yes	RD
Fasting & Lyster (2005)	0.92	0.083	52	6	TTS	I&C	CP	GL	Yes	Yes	LS60
Fletcher et al. (2006)	0.90	0.048	91	3	HR	C	PP&Q/T	GL	Yes	Yes	TAKS
Fletcher et al. (2009)	0.50	0.065	62	7	HR	C	PP&Q/T	GL	Yes	Yes	TAKS
Lange et al. (2006)	0.77	0.083	54	9.5	TTS	C	CP&Q/T	GL	No	Yes	NARA*
Lundberg & Olofsson (1993)	-0.32	0.241	15	5.6	TTS	I	PP	CR	Yes	Yes	RD
Meloy et al. (2002)	1.10	0.073	62	7	HR	C	CP&Q/T, RA	CL	No	Yes	ITBS
Roberts et al. (2013)	0.45	0.039	164	9	TTS	I	CP	CL	Yes	No	GMRT
Shany & Biemiller (1995)	0.81	0.209	18	3.5	HR	I	CP	CR	Yes	Yes	SAT*
Within-Subject Study Design											
Dolan et al. (2005)	1.02	0.575	9	11.5	TTS	C	CP&Q/T	GL	No	Yes	NAEP
Elkind et al. (1996)	0.10	0.139	50	17	TTS	C	CP&Q/T	GL	No	Yes	NDRT
Floyd & Judge (2012)	1.09	0.102	6	17	TTS	C	CP&Q/T	GL	No	Yes	SAT*
Higgins & Raskind (1997)	-0.12	0.019	37	13	TTS	C	CP&Q/T, RA	GL	No	Yes	FRI
Hodapp et al. (2007)	-0.74	0.034	27	7.5	TTS	C	CP&Q/T	GL	No	No	JRS
Kosciolek & Ysseldyke (2000)	0.53	0.073	14	4	RHV	C	CP&Q/T	GL	No	No	CAT/5
Laitusis(2010a)	0.57	0.000	527	4	RHV	C	CP&Q/T, RA	GL	No	Yes	GMRT
Laitusis (2010b)	0.32	0.000	376	8	RHV	C	CP&Q/T, RA	GL	No	Yes	GMRT
McCove (2013)	0.53	0.337	3	10	TTS	I&C	CP	GL	Yes	No	STAR
Meyer & Bouck (2014)	-0.21	0.143	3	7.6	TTS	C	CP&Q/T	GL	No	Yes	SWP
Schmitt et al. (2011a)	0.23	0.063	25	7	TTS	C	CP	GL	Yes	Yes	TRL
Sorrell et al. (2007)	-0.16	0.078	4	4	TTS	C	CP	CR	Yes	Yes	AR
Thurlow et al. (2010)	-0.32	0.011	44	6.7	RP	C	PP&Q/T	GL	Yes	No	GSRT

Note: * = researcher modified, $\bar{d}$ = effect size, v = variance, n = number of subjects, CP = complete passages, PP = partial passages, Q/T = questions/test(s), RA = explicitly stated that they could request oral reading material on questions, passages or comprehension tests, GL = grade level, CR = student's reading level, HR = human reader, C = compensation, I = intervention, RHV = recorded human voice, TTS = text-to-speech, RP = reading pen, MAP = Missouri Assessment Program, TAKS = Texas Assessment of Knowledge and Skills, FRI = Formal Reading Inventory, RD = Researcher Developed, NARA* = Modified Neale Analysis of Reading Ability II, ITBS = Iowa Test of Basic Skills the reading comprehension part, GMRT = Gates-MacGinitie Reading Tests, NAEP = NCES U.S. History and Civics tests, NDRT = Nelson-Denny Reading Test, LS60 = Test of Silent Sentence Reading, CTBS = Canadian Test of Basic Skills reading comprehension part, SAT* = SAT

Critical Readings, TRL = Timed Readings in Literature, AR = Accelerated Reader, STAR = STAR Reading Assessment Test, SWP = Six Way Paragraphs Middle Level, CAT/5 = California Achievement Test 5th Edition comprehension part, GSRT = The Gray Silent Reading Test, JRS = Jamestown Reading Series.

Table 2

Moderator Analysis

Moderator	K	N	B	Qm	Qe
Between Subjects	23	2942	.48*	6.17	2049.60**
Grade Level	23	2942	<−.01	.01	168.13**
Intervention	23	2942	.20	.47	2049.02**
Difficulty	23	2942	.30	67	2049.14
Delivery Method (2 modes below)	23	2942		4.45	2009.41
a. *Recorded Voice*			.29		
b. *Human Reader*			.44		
Decoding Needed	23	2942	−.02	.01	2041.46**
Partly Oral Passages	23	2942	−.16	.32	2032.51**
Ability to Repeat Oral Reading	23	2942	.11	.16	2040.92**
Peer Reviewed	23	2942	.41	3.49	2028.56
Randomized	23	2942	.28	1.34	2046.64
Baseline	3	1337	.59	1.86	1.46

Note. *K* = number of studies included, B = change in effect size of 1 unit increase in moderator. Qm = heterogeneity explained by the moderator, Qe = remaining unexplained heterogeneity.