A Systematic Review on Quality Indicators of Randomized Control Trial Reading Fluency Intervention Studies

Nithya Naveenkumar, George K. Georgiou, Ana Paula Alves Vieira, Sandra Romero, and Rauno Parrila

IN PRESS

READING & WRITING QUARTERLY

Author note

Nithya Naveenkumar, Department of Educational Psychology, University of Alberta, Canada; George K. Georgiou, Department of Educational Psychology, University of Alberta, Canada; Ana Paula Alves Vieira, Department of Educational Psychology, University of Alberta, Canada; Sandra Romero, Department of Educational Psychology, University of Alberta, Canada; and Rauno Parrila, Department of Educational Studies, Macquarie University, Australia.

The study was supported by a Social Sciences and Humanities Research Council of Canada grant (#RES0029061) to the second author.

Correspondence concerning this article should be sent to Dr. George K. Georgiou, Department of Educational Psychology, 6-102 Education North, University of Alberta, Edmonton-AB, Canada, T6G 2G5; Email: georgiou@ualberta.ca: Phone: +1(780) 492-8247

Abstract

Although several studies have examined the effectiveness of reading fluency interventions, the methodological rigor of these studies remains unclear. The purpose of this study was to examine the quality of randomized controlled trial (RCT) reading fluency intervention studies using the Council for Exceptional Children (CEC) standards. Twenty-six studies met the inclusion criteria for the review. All included studies met the four quality indicators for context and setting, participants, intervention agent, and data analysis. In addition, 22 out of the 26 studies met the quality indicators for internal validity and implementation fidelity, and 18 met the quality indicator for outcome measure. Finally, 16 of the studies met all eight quality indicators, and the majority of them also reported significant effects on reading fluency outcomes. Findings and future directions are discussed considering the quality indicators for methodologically rigorous reading fluency intervention studies.

Keywords: CEC standards, evidence-based practices, fluency programs, intervention, reading fluency, research quality indicators.

A Systematic Review on Quality Indicators of Randomized Control Trial Reading Fluency Intervention Studies

Reading fluency, defined as the ability to read accurately with appropriate speed and expression (Kuhn & Stahl, 2003; NICHD, 2000), is a fundamental skill in the process of becoming proficient in reading. Its development is supported by two factors: automaticity (i.e., fast and accurate word recognition) and the prosodic features of language (e.g., voice inflection, pausing) (Kuhn & Stahl, 2003; Kuhn et al., 2010; Rasinski et al., 2011). Although reading fluency training has been overlooked in literacy curricula and remedial school programs (Allington, 1983, 2006; see also Rasinski et al., 2011, for a historical account), research on reading fluency intervention is abundant. Despite the proliferation of fluency interventions in the last two decades, recent meta-analyses have reported somewhat limited effects from the intervention studies (see e.g., Lee & Yoon, 2017; Stevens et al., 2017). Scammacca et al. (2015) argued that the increased use of more rigorous methodological designs may partly explain the smaller effect sizes. Unfortunately, none of the existing meta-analyses on reading fluency interventions have examined the methodological quality of their studies. This raises concerns about whether current treatment methods qualify as evidence-based practices.

Quality Indicators for Evidence-Based Practices in Reading Fluency Intervention

Three literature reviews examining the methodological quality of reading fluency intervention studies have been published (Begeny et al., 2018; Chard et al., 2009; and O'Keeffe et al., 2012). These systematic reviews used the quality indicators (QIs) suggested by Gersten et al.'s (2005) for group comparison studies and Horner et al.'s (2005) for single-case studies.[1]

[1]Gersten et al. (2005) listed ten essential and eight desirable QIs. Their essential QIs were further grouped into four categories: QIs for describing participants, QIs for implementation of the intervention and description of comparison conditions, QIs for outcome measures, and QIs for data analysis. Horner et al. (2005) identified seven QIs

In the most recent review, Begeny et al. (2018) assessed the methodological quality of fluency intervention studies that targeted small-group fluency instruction. They included 11 single-case design (SCD) studies and two group-design studies (GDS). All the included studies targeted at-risk elementary students. Results showed that only three (27.3%) of the SCD studies met all seven QIs, six (54.5%) studies met six QIs, and the remaining two (18.2%) studies met five QIs. Also, only one of the two group design studies met the standards, but with a few limitations. Begeny et al. concluded that few high-quality studies were found in SCD studies and that the dearth of GDS prevented them from determining whether methodologically rigorous reading fluency intervention studies exist.

Chard et al. (2009) and O'Keeffe et al. (2012) conducted systematic reviews of intervention studies that used the Repeated Reading approach. Both reviews concluded that Repeated Reading, a widely used reading fluency intervention method, should be further scrutinized as none of the studies examined by these reviews met all quality standards.

Moreover, both statistical and non-statistical systematic reviews on reading fluency interventions have reported a scarcity of GDS. Thus, the reported effects of fluency interventions are largely based on single-subject design studies, limiting the external validity of their findings (this is not to say that single-subject design studies are not important, but that their external validity is limited). Therefore, as Stevens et al. (2017) pointed out, inconclusive outcomes from reading fluency interventions might be due to the scarcity of studies that fully meet the standards for group comparison studies.

(description of participants and settings, dependent variable, independent variable, baseline, experimental control/internal validity, external validity, and social validity) with 21 subcomponents.

The scarcity of GDS is more severe when it comes to Randomized Control Trials (RCTs). RCT design is thought to be the "gold standard" in developing evidence-based practices (Davies, 1999; Odom et al., 2005; Torgerson & Torgerson, 2001) as the most rigorous and robust research method. According to the group design standards developed by What Works Clearinghouse (2020), intervention studies must randomly assign participants to treatment conditions to be eligible for the highest rating of methodological quality. The main reason for this is the increased risk of selection bias when researchers assign participants to groups using a non-random procedure (Cook et al., 2015). To date, there have not been any systematic reviews that solely examined the quality of reading fluency intervention studies with an RCT design. Begeny et al. (2009) did not include RCT design studies, Chard et al. (2009) included only three, and O'Keeffe et al. (2012) included eight, two of which overlapped with the previous reviews. Since more RCT fluency studies have been published in the last few years, a methodological analysis of these studies would be useful.[2]

Finally, Begeny et al. (2018), Chard et al. (2009), and O'Keeffe et al. (2012) used the group design standards established by Gersten et al. (2005); however, some researchers have argued that some of their QIs are insufficiently operationalized (Cook et al., 2009; Montague & Dietz, 2009). The type of framework used to analyze the studies could partially explain the lack of reading fluency intervention studies that met the quality standards reported in previous systematic reviews. Thus, the purpose of our study was to employ the more systematic and better operationalized set of QIs from the Council of Exceptional Children (CEC, 2014) to evaluate whether reading fluency intervention studies using an RCT design meet the criteria for

[2] Notice that none of the RCT studies included in our review were examined by Chard et al. (2009) or O'Keeffe et al. (2012) because their RCT studies were published before 2000, which was used here as the beginning year of our search (see Methods section).

methodologically rigorous studies. In addition, we were interested in identifying which reading fluency instructional practices for children in elementary school years are supported by RCT design studies published in the last 20 years that meet all CEC (2014) QIs and reported significant improvement in reading fluency measures.

Method

Article Selection Procedures

The search procedure is detailed in Figure 1. We first searched electronic databases (ERIC, PsycINFO, PubMed, and Google Scholar) to identify experimental group-comparison studies of reading fluency intervention published in peer-reviewed journals in English from January 2000 to December 2019. A combination of terms related to reading fluency (*reading fluency* OR *reading rate* OR *reading speed*) crossed with terms related to reading intervention (*reading programs* OR *reading fluency intervention* OR *fluency program* OR *reading intervention* OR *school-based intervention*) was used to identify the initial pool of studies. We further hand searched specialized journals (*Journal of Learning Disabilities, Journal of Literacy Research, Journal of School Psychology, Learning Disabilities Quarterly, Learning Disabilities Research & Practice, Literacy Research and Instruction, Reading Research Quarterly, Reading & Writing Quarterly, Remedial and Special Education, School Psychology Quarterly, Scientific Studies of Reading, and Exceptional Children*). The results from this search yielded 328 studies. An ancestral search of studies mentioned in the reference lists of the previously found studies helped us identify 15 additional studies.

To achieve our research objectives, we applied the following inclusion criteria: (a) the article was published in English in a peer-reviewed journal, [3] (b) the study was published

[3] Book chapters and dissertations were excluded from our review as we could not check whether they went through a rigorous peer review process. For the same reason, we did not include any unpublished studies.

between January 2000 and December 2019,[4] (c) the study was conducted in a school setting (Kindergarten to Grade 12), (d) the study employed an RCT research design, and (e) the examiners measured and reported the effects of the intervention on students' reading fluency performance.

After removing 23 duplicate studies, we scrutinized the 320 abstracts and further eliminated 288 studies because they had non-random assignment of participants, had duplicate samples, did not report inferential statistics, or were published as book chapters or dissertations. The remaining full-text articles (n = 32) were assessed for eligibility and six studies were further excluded. The final sample (n = 26) was then independently read in full by the first, third, and fourth authors who coded the articles against the inclusion criteria. The percentage of agreement was 100% for the 26 studies that were subsequently included in the review.

Quality Coding Procedures

To assess the methodological quality of empirical studies, the Council of Exceptional Children (CEC, 2014) established operationally–defined QIs for experimental design studies and standards for determining evidence-based practices. Although the samples in the studies included in this review are not limited to children with or at-risk for reading disabilities, the use of the CEC (2014) standards allowed for a more thorough examination of the studies than other approaches.[5]

[4] We selected the year 2000 as our starting point for two reasons: first, it gives us a reasonable time window (19 years) to identify enough studies for the purpose of this methodological analysis, and second, it seems to be the starting point of most recently published meta-analyses and systematic reviews.

[5] We considered using the methodological design steps proposed by What Works Clearinghouse (WWC, 2020) to rate GDS. However, 10 of the 26 studies included in this review did not provide sufficient information regarding the reliability of the outcome measure or attrition and differential attrition levels to adequately receive one of the three rankings from WWC (i.e., meets standards without reservations, meets standards with reservations, and does not meets standards). To avoid the exclusion or the potentially inappropriate rating of these studies in our review, we decided to only use the more detailed list of 23 indicators for GDS proposed by the CEC (2014).

To begin the quality coding process, the authors evaluated the studies using the CEC (2014) QIs and coded each study. The CEC standards include eight categories of QIs: 1) Context and Setting, 2) Participants, 3) Intervention Agent, 4) Description of Practice, 5) Implementation Fidelity, 6) Internal Validity, 7) Outcome Measure, and 8) Data Analysis. Each of these categories consists of one or more indicator components (see also Table 1).

The third author created a coding sheet with all eight quality indicators and included two columns to mark whether each QI was adequately met = 1 or insufficiently met = 0. Each of the QI categories were assessed as "met" if all the indicator components were either stated explicitly or could be inferred from the existing studies.[6] For example, the Participants category has two indicators (participants' characteristics and disability status). If both indicators are found in the study, then a score of 1 is given for the Participants category. If one or both indicators are missing, then the study receives a score of 0 for that category. Since there are eight QI categories, the studies could receive a score from 0 to 8. Notice that while studies are ranked according to their compliance to the eight general categories, the standards proposed by the CEC (2014) encompass a total of 23 individual indicators for group comparison studies with an RCT design. The QI categories and their specific indicator components are described below.

QI1.Context and Setting

The first CEC (2014) QI asks whether the study provides adequate information on essential features of the context and/or setting (e.g., geographical setting, type of school, language of instruction). For the interest of our review, we looked at whether the studies specified the school setting. However, if the contextual information was relevant to the

[6] In the absence of explicit information, a QI can still be met if the missing information does not threaten the validity of the study (Cook et al., 2015). For example, some instructional procedures and materials are commonly understood and do not require an explicit description.

interpretation of the results, the researchers needed to report other essential features such as type of classroom (e.g., special education classes) or instructional approach.

QI2. Participants

The CEC (2014) QI for participants requires the studies to report on two components. For the first component, QI 2.1, studies should provide relevant characteristics of their participants. At a minimum, we assessed studies on reporting two major descriptors relevant to our review: a) age or grade level to determine school-age status, and b) reading level (e.g., reading disabled, average, ELLs). However, other relevant descriptors include gender, ethnicity, socioeconomic status (SES), and language. The second component (QI 2.2) requires an explanation of participants' disability or risk status and, if any, then the identification process of determining disability status (e.g., by the school using state guidelines, teacher nomination, standardized intelligence test, curriculum-based measurement probes, or rating scales). Concerning the latter, if the US students with special education needs are identified through the Individuals with Disabilities Education Improvement Act (IDEA) and are receiving special education services, this is considered adequate to meet this QI component (Cook et al., 2015). Since we included RCT reading fluency intervention studies with all kinds of school-aged children, the studies that did not target children with learning or developmental disabilities/difficulties were not required to meet QI 2.2.

QI3. Intervention Agent

Studies met the third QI when adequate information was provided on two components: (QI 3.1) researchers should describe the role of the intervention agent (e.g., research assistant or school teacher) and their background information (e.g., ethnicity or race) when relevant; and (QI 3.2) the authors were required to report on training or qualifications of the interventionists.

QI4. Description of Practice

As stated by CEC (2014), the studies fulfill this QI if they describe and provide essential information on these two research components: (QI 4.1) a clear explanation of the intervention procedure and components, including intervention dosage, or a study should cite one or more accessible sources to offer sufficient information on intervention processes; and (QI 4.2) intervention materials should be described, or include references on how to access information on them. Cook et al. (2015) observed that a description of every intervention procedure detail is not required to satisfy this QI. However, Cook et al. clarify that studies must provide the necessary information for others to implement the critical features of the intervention.

QI5. Implementation Fidelity

According to the CEC (2014), the QI for implementation fidelity asks whether three components are assessed and adequate evidence is provided to meet these criteria: (QI 5.1) in terms of adherence through direct or reliable measures, such as observation checklist; (QI 5.2) dosage or exposure using direct or reliable measures, namely observations or self-report of the duration, frequency, and curriculum coverage of implementation; and (QI 5.3) frequent assessment of adherence or dosage throughout the intervention and by any unit of analysis (e.g., by interventionist or participant).

QI6. Internal Validity

CEC (2014) QIs include nine components of internal validity; however, we disregarded three indicators (QI 6.5, QI 6.6, and QI 6.7) as they only pertain to single-subject design studies. According to the six indicators for internal validity of GDS, the studies should: (QI 6.1) control and systematically manipulate the independent variable, (QI 6.2) describe control or comparison conditions, (QI 6.3) establish limited or no access of control or comparison group to the

intervention condition, (QI 6.4) include random assignment of study participants, (QI 6.8) report attrition and have rates of attrition below 30%, and (QI 6.9) report differential attrition with a minimal difference between the groups (< 10%). For QI 6.8 and QI 6.9, the articles should report or provide sufficient information on attrition (e.g., difference between the randomized subjects and the number of subjects, degrees of freedom from statistical analysis) and differential attrition (i.e., loss of subjects between groups) to infer the rate required to meet this QI (Cook et al., 2015).

QI7. Outcome Measures/Dependent Variables

The CEC (2014) standards incorporate six components that are concerned with outcome measures and dependent variables. However, one component (QI 7.4) was excluded from the coding criteria;[7] therefore, the included studies were evaluated on the following five components: (QI 7.1) outcomes are socially important, (QI 7.2) a specific description of dependent variable measures is included, (QI 7.3) statement of intervention effects on all outcome measures is included, and evidence of reliability (QI 7.5; e.g., inter-observer reliability, test-retest reliability, parallel-form reliability) and validity (QI 7.6; e.g., content, construct, criterion or social) is provided. If a study had multiple measures for the outcome variable and reported evidence of reliability and validity for only some of the measures, it can be considered to meet this QI component (Cook et al., 2015). For our review, we targeted reading fluency outcomes (e.g., words read correctly per minute) and, therefore, the studies were required to report the evidence for reliability and validity for at least one measure of reading fluency.

QI8. Data Analysis

[7] We excluded QI 7.4 (requires information on frequency and timing of outcome measures for single-case and alternating treatment designs) as it was not relevant for RCTs.

According to CEC (2014), to ensure the proper use of data analysis, the reviewers should assess the studies on three components; however, QI 8.2 was excluded as it is not related to GDS. The remaining two indicators for data analysis require that (QI 8.1) the study uses appropriate techniques to compare the group differences (e.g., *t* test, ANOVAs/ MANOVAs, ANCOVAs/ MANCOVAs, hierarchical linear modelling, structural equation modeling), and (QI 8.3) the effect size statistics for all outcome variables are reported or required data to calculate effect sizes are provided (CEC, 2014).

Intercoder Agreement

To ensure the coding rules were consistently followed, the first and third author independently coded two studies and then estimated the inter-rater reliability (IRR). The IRR across the two studies was 96%. A consensus meeting was then conducted between the coders and any discrepancies were discussed and resolved. Subsequently, the first, third, and fourth authors read and coded all 26 articles independently. The percentage of IRR was calculated by dividing the sum of agreements by the total number of QI and multiplied by 100. Overall, the IRR percentage of coding QIs for all the components was 98% (*SD* = 3.29; ranged from 87% to 100%). The disagreements in the coding were mainly due to scarce information provided in a few studies regarding attrition levels, internal validity, and outcome measures. All the disagreements in coding the QI components were discussed and resolved until we reached full consensus.

Results

The results of the methodological analysis are reported in Table 2. Overall, the reviewed studies accounted for a mean of 91% (*SD* = 13.24; range 62.5 to 100) of the QIs met. Of the 26 studies, 16 met 100% of all the QIs (Algozzine et al., 2009; Ardoin et al., 2016; Begeny, 2011;

Begeny et al., 2012; Hammerschmidt-Snidarich et al., 2018; O'Connor et al., 2007, 2010, 2013; Patton et al., 2010; Reed et al., 2019; Therrien et al., 2012; Vadasy & Sanders, 2008a; 2008b; Vess et al., 2018; Wexler et al., 2010; Young et al., 2018). All 26 studies achieved 100% in four QIs (context and setting, participants, intervention agent, and data analysis). Description of practice was met by 96% of the studies, implementation fidelity was met by 85%, and internal validity by 81%. The QI for outcome measures/dependent variables was the least frequently addressed QI. Eight studies did not meet the criteria to satisfy the QI for outcome measures.

QI1. Context and Setting

All 26 studies met this QI. The great majority of the studies (n = 24; 92%) were conducted in the United States. Of the two studies that took place outside of the United States, one was conducted in Finland (Heikkilä et al., 2013) and the other in Uganda (Friedland et al., 2017). Twenty-three intervention studies were conducted in primary schools. The rest were conducted in a high school (Wexler et al., 2010) or middle schools (Allinder et al., 2001; Vess et al., 2018). The most frequently reported school variables were the geographic location, race/ethnicity, and the percentage of students receiving free and reduced-price lunches. Studies working with children with learning disabilities or other exceptionalities were required to report additional information. Such studies adequately addressed QI 1 by reporting the percentage of school population receiving special education services (Allinder et al., 2001; Begeny et al., 2011, 2012; Esteves & Whitten, 2011; Therrien et al., 2012; Vess et al., 2018) and English as second language (ESL) services (Begeny et al., 2012; Vess et al., 2018).

QI2. Participants

Overall, all included studies (100%) met this QI. All 26 studies reported the primary demographic data of 2,955 participants. Although three of the studies provided limited

information on participants' demographics (Ardoin et al., 2013; Esteves & Whitten, 2011; Keehn, 2003), the researchers provided the essential information to meet the criteria for this QI. The samples across the studies were from Grade 1 to Grade 12.

Fifteen of the intervention studies (58%) involved poor readers, students with reading fluency deficits, reading disabilities, or students performing below grade-level, and reported that the students were identified using state guidelines, standardized assessments on reading fluency, records of individualized education programs in the area of reading, and/or teacher nominations (Algozzine et al., 2009; Begeny et al., 2011; Esteves & Whitten, 2011; Hammerschmidt-Snidarich et al., 2018; Heikkilä et al., 2013; Kuhn et al., 2006; Martens et al., 2007; O'Connor et al., 2007, 2010; Therrien et al., 2012; Vadasy & Sanders, 2008a, 2008b; Vess et al., 2018; Wexler et al., 2010; Young et al., 2018). Allinder et al. (2001) targeted students from remedial reading classes and clearly reported the portion of students with identified disabilities. In two studies, O'Connor et al. (2013) and Begeny et al. (2012) examined the reading fluency gains of English-language learners. Six studies involved only typically-developing students (Ardoin et al., 2013, 2016; Keehn, 2003; Begeny, 2011; Patton et al., 2010; Reed et al., 2019). Begeny et al. (2010) and Friedland et al. (2017) stated that their samples were unselected.

QI3. Intervention Agent

In total, all the included studies (100%) reported adequate details regarding the intervention agents, including their training or qualifications. In eight out of 26 studies the intervention was delivered by undergraduate and/or postgraduate students (Begeny, 2011; Hammerschmidt-Snidarich et al., 2018; Martens et al., 2007; O'Connor et al., 2010, 2013; Vess et al., 2018; Wexler et al., 2010; Young et al., 2018), and in three studies (Ardoin et al., 2013; Begeny et al., 2010, 2012), the primary authors also acted as intervention agents with

undergraduate and/or postgraduate students. In Ardoin et al. (2016) and O'Connor et al. (2007), the experimenters served as the principal intervention agents. In seven studies, the classroom teachers acted as interventionists (Allinder et al., 2001; Begeny et al., 2011; Esteves & Whitten, 2011; Friedland et al., 2017; Heikkilä et al., 2013; Keehn, 2003; Kuhn et al., 2006). One of the studies included peers as student coaches to deliver the intervention (Algozzine et al., 2009). In Reed et al. (2019), the interventions were implemented by students working in pairs whose fidelity to the procedures needed to be monitored by their teachers. In three studies (Patton et al., 2010; Therrien et al., 2012; Vadasy & Sanders, 2008a), the intervention was delivered by paraprofessionals, and in one study (Vadasy & Sanders, 2008b), the tutors who implemented the intervention were matched to paraeducator competency requirements under the *No Child Left Behind Act* from 2001.

QI4. Description of Practice

As shown in Table 2, most of the studies (25, 96%) met this QI. Overall, 100% of the studies provided essential features of the intervention procedure. However, Friedland et al. (2017) did not meet QI 4.2 due to an incomplete description of the reading materials (e.g., text grade level) used during the intervention sessions, nor did they cite external sources where this information could be obtained from.

Seventeen out of 26 studies (65%) implemented Repeated Readings as their main intervention strategy (Algozzine et al., 2009; Ardoin et al., 2013, 2016; Hammerschmidt-Snidarich et al., 2018; Heikkilä et al., 2013; Keehn, 2003; Kuhn et al., 2006; Martens et al., 2007; O'Connor et al., 2007; Patton et al., 2010; Reed et al., 2019; Therrien et al., 2012; Vadasy & Sanders, 2008a, 2008b; Vess et al., 2018; Wexler et al., 2010; Young et al., 2018). Six of these studies (Ardoin et al., 2016; Hammerschmidt-Snidarich et al., 2018; Kuhn et al., 2006; O'Connor

et al., 2007; Reed et al., 2019; Wexler et al., 2010) compared Repeated Reading with continuous oral reading of different passages (also known as Wide Reading). O'Connor et al. (2010, 2013) also provided continuous reading practice; however, they compared the effect of 10 vs. 20 minutes of practice (O'Connor et al., 2013) and text level difficulty (O'Connor et al., 2010).

Five studies (Begeny, 2011; Begeny et al., 2010, 2011, 2012; Vess et al., 2018) administered the Helping Early Literacy with Practice Strategies (HELPS) program that combines several evidence-based fluency-building instructional strategies (i.e., repeated reading, modeling, phrase-drill error correction, verbal cueing, goal setting, performance feedback, and reward system). Furthermore, the authors provided references and sources for additional information upon the HELPS program.

Two studies (Patton et al., 2010; Therrien et al., 2012) implemented the Reread-Adapt and Answer-Comprehend (RAAC) intervention. According to these studies, the program involves oral reading, comprehension monitoring, Repeated Reading, corrective feedback, performance feedback on fluency markers, and adjusting passages according to reading level.

Two studies administered the Great Leaps Reading Program (GLR) (Begeny et al., 2010; Patton et al., 2010). GLR also trains readers on other reading skills such as phonics and sight-word recognition; however, its focus is on reading fluency.

Esteves and Whitten (2011) and Friedland et al. (2017) assisted struggling readers through audiobooks designed to provide a model of fluent reading while doing independent reading practice.

Finally, Allinder et al. (2001) studied the effects of prompting students to focus on a specific reading strategy on fluency performance. The strategies included reading with inflection,

not adding words, pausing at periods and commas, self-monitoring for accuracy, reading at an appropriate rate, watching for word endings, and finger tracking.

QI5. Implementation Fidelity

Twenty-two out of 26 studies (85%) reported sufficient information to meet the criteria. In general, most studies employed direct observation of intervention sessions (Algozzine et al., 2009; Ardoin et al., 2013; Ardoin et al., 2016; Begeny, 2011; Begeny et al., 2010, 2012; Hammerschmidt-Snidarich et al., 2018; O'Connor et al., 2007, 2010, 2013; Patton et al., 2010; Therrien et al., 2012; Vadasy & Sanders, 2008a, 2008b; Vess et al., 2018; Wexler et al., 2010; Young et al., 2018). In addition, some of the aforementioned studies used a second strategy, such as self-reports (Algozzine et al., 2009; Ardoin et al., 2016; Begeny et al., 2011; O'Connor et al., 2007, 2010, 2013; Vess et al., 2018; Wexler et al., 2010; Young et al., 2018) and audio recording assessments (Begeny, 2011; Begeny et al., 2010, 2012; Martens et al., 2007). One study (Kuhn et al., 2006) implemented a revised version of the *CIERA School Change Classroom Observation Scheme* to code the applied procedures. In Reed et al. (2019), the fidelity to the procedures were monitored by teachers in two ways. First, they checked all the students' reading logs where they had to record the date of the session, passage names, amount of time spent reading a passage, and the number of errors made. Second, teachers submitted procedural checklists biweekly.

Four studies did not meet this QI (Allinder et al., 2001; Friedland et al., 2017; Heikkilä et al., 2013; Keehn, 2003). Although Allinder et al. and Keehn observed the intervention sessions regularly, they did not report the implementation fidelity. Heikkilä et al. and Friedland et al. used electronic devices (e.g., computers, tablets) as intervention tools and did not provide any information on either adherence or dosage.

QI6. Internal Validity

Of the 26 articles, 22 (85%) studies provided adequate information to meet the requirements of internal validity (Algozzine et al., 2009; Allinder et al., 2001; Ardoin et al., 2013, 2016; Begeny, 2011; Begeny et al., 2010, 2011, 2012; Esteves & Whitten, 2011; Friedland et al., 2017; Hammerschmidt-Snidarich et al., 2018; Reed et al., 2019; O'Connor et al., 2007, 2010, 2013; Patton et al., 2010; Therrien et al., 2012; Vadasy & Sanders, 2008a; 2008b; Vess et al., 2018; Wexler et al., 2010; Young et al., 2018).

Sixteen studies compared one or more types of fluency intervention with a "Business-as-Usual" (BU) control group (Algozzine et al., 2009; Allinder et al., 2001; Ardoin et al., 2016; Begeny, 2011; Begeny et al., 2010, 2011, 2012; Esteves & Whitten, 2011; Friedland et al., 2017; Kuhn et al., 2006; O'Connor et al., 2007, 2010; Vadasy & Sanders, 2008a, 2008b; Wexler et al., 2010; Young et al., 2018). Thirteen of these studies provided a more detailed description of the literacy curriculum implemented in the regular classroom (Algozzine et al., 2009; Allinder et al., 2001; Begeny, 2011; Begeny et al., 2010, 2011, 2012; Esteves & Whitten, 2011; Friedland et al., 2017; Kuhn et al., 2006; O'Connor et al., 2010; Vadasy & Sanders, 2008b; Wexler et al., 2010; Young et al., 2018). Although providing information on the type of literacy activities or literacy program followed by instructors in the control group was not a requirement to meet QI 6, it is highly recommended that studies provide such a description.

On the other hand, four studies compared different types of interventions without a control group (Ardoin et al., 2013; Hammerschmidt-Snidarich et al., 2018; Reed et al., 2019; Vess et al., 2018). Additionally, five studies compared variations of the same kind of intervention (Heikkilä et al., 2013; Keehn, 2003; O'Connor et al., 2013; Patton et al., 2010; Therrien et al., 2012). More specifically, Keehn (2003) compared treatment A (Readers Theater repertory and specific oral fluency instruction) with treatment B in which the same program was

delivered but with no specific fluency instruction. O'Connor et al. (2013) compared 10 minutes intervention of reading aloud one-to-one against 20 minutes. Patton et al. (2010) delivered the GLR program in one group and GLR program plus comprehension training with the second group. Therrien et al. (2012) compared the RAAC program with a modified RAAC version without the repeated reading component. Finally, in the study by Heikkilä et al., training was delivered using a computer program; however, children in the control condition practiced solving arithmetic problems and did not engage in reading activities.

Four studies did not meet this QI (Heikkilä et al., 2013; Keehn, 2003; Kuhn et al., 2006; Martens et al., 2007). In the studies by Keehn, Kuhn et al., and Martens et al., general (QI 6.8) and differential (QI 6.9) attrition was not reported or possible to infer. In Heikkilä et al., high general and differential attrition levels were reported. However, QI 6.9 is met since the study demonstrated pretest equivalence of groups (Cook et al., 2015).

QI7. Outcome Measures/Dependent Variables

Eighteen (69%) studies met all five components of this QI (Algozzine et al., 2009; Ardoin et al., 2016; Begeny, 2011; Begeny et al., 2012; Hammerschmidt-Snidarich et al., 2018; Kuhn et al., 2006; Martens et al., 2007; O'Connor et al., 2007, 2010, 2013; Patton et al., 2010; Reed et al., 2019; Therrien et al., 2012; Vadasy & Sanders, 2008a, 2008b; Vess et al., 2018; Wexler et al., 2010; Young et al., 2018). All 18 studies reported evidence for reliability. Twelve studies reported evidence of validity. For the remaining six studies, we assessed the content validity of their outcome measure.[8] The outcome measure validity was assessed as met since all studies used well-known standardized reading fluency tests (i.e., AIMSweb, DIBELS, GORT-5, and TOWRE-2).

[8] According to Cook et al. (2015), the QI for validity (QI 7.6) is met if reviewers consider that the outcome measure accurately represents the construct (i.e., content validity).

Sixteen studies measured reading fluency through standardized assessments such as AIMSweb Reading Curriculum Based Measurement (R-CBM; Shinn & Shinn, 2002), Dynamic Indicators of Basic Early Literacy (DIBELS; Good & Kaminski, 2002), Formative Assessment System for Teachers (FASTBridge; Christ et al., 2014); Gray Oral Reading Tests – Fourth and Fifth Edition (GORT-4; Wiederholt & Bryant, 2001; GORT-5; Wiederholt & Bryan, 2012), Test of Word Reading Efficiency-First and Second Edition (TOWRE; Torgesen et al., 1999; TOWRE-2; Torgesen et al., 2012), and Woodcock-Johnson III (Woodcock et al., 2007).

Five studies used reading curriculum-based measurement (CBM-R). Wexler et al. (2010) assessed oral reading fluency with three standardized assessment passages from the AIMSWeb system. Ardoin et al. (2016) used three probes from *Formative Assessment Instrumentation and Procedures for Reading* (FAIP-R) to calculate words read correctly per minute (WRCM). The authors reported that CBM-R oral reading fluency measure was comparable to other CBM-Rs (Ardoin et al., 2013; Christ et al., 2013). Similarly, Hammerschmidt-Snidarich and colleagues (2018) evaluated WRCM of three grade-level CBM-R probes from *FastBridge Learning.* Reed et al. (2019) also used the *FastBridge* to measure reading rate and accuracy. All five studies provided sufficient psychometric information with reliability ranging from 0.94 to 0.98 (Baker et al., 2008) and 0.89 to 0.96 (Roberts et al., 2005). The researchers demonstrated adequate concurrent (r = 0.67 – 0.82; Baker et al., 2008) and predictive validity (r = 0.62 – 0.92; Harn et al., 2008) for CBM-R.

One study provided information on the reliability and validity of a measurement tool for the prosodic component of reading fluency.[9] More specifically, Young et al. (2018) measured reading fluency using the Multidimensional Fluency Scale (MFS; Zutell & Rasinski, 1991) and

[9] Most studies included in this review measured WRCM to assess reading fluency intervention gains.

reported interrater reliability ($\kappa = .77$, $p < .001$). The evidence for validity was provided in a previously published study.

Eight (30%) studies did not provide sufficient information to meet the standard of outcome measure. More specifically, four studies (Allinder et al., 2001; Begeny et al., 2010, 2011; Esteves & Whitten, 2011) did not meet the requirement for reliability scores; one study (Ardoin et al., 2013) did not report validity of reading fluency measures; and three studies (Friedland et al., 2017; Heikkilä et al., 2013; Keehn, 2003) failed to report on both reliability and validity.

QI 8. Data Analysis

All 26 (100%) studies met this QI. The researchers in all studies either reported effect sizes for both statistically significant and non-significant reading fluency outcomes or provided adequate data to estimate effect sizes for all outcome variables to meet the standards of data analysis. Thus, all the studies in our analysis fulfilled these criteria.

Discussion

The primary purpose of this review was to identify and evaluate the methodological rigor of reading fluency interventions that employed an RCT design. To our knowledge, this review is the first attempt to assess RCT-design reading fluency intervention studies using the CEC (2014) standards. Overall, most of the studies in our review had a high level of methodological quality. We found 16 studies that met all eight QIs and five studies that met seven. The remaining five studies could be considered of moderate quality by meeting six or five QIs (see Table 2). In what follows, we will first discuss the findings on quality indicators and then examine what practices could be considered evidence-based in light of these findings.

Indicators of Quality Standards in Reading Fluency Intervention Studies

The authors of all 26 studies provided sufficient information on circumstantial and environmental features (QI1), participant demographics, and, if relevant, the determination of their disability or risk status (QI2). All studies provided details regarding the individuals who delivered the intervention as well as their educational qualification or particular training to implement the intervention procedures (QI3) and appropriate methods of data analysis for group comparison and information to observe or calculate effect sizes (QI8).

Across the four QIs that were not met, three QIs (i.e., internal validity, implementation fidelity, and outcome measures) were inadequately addressed by several studies. Keehn (2003), Kuhn et al. (2006), and Martens et al. (2007) did not meet the QI for internal validity. These studies included insufficient data to estimate the attrition and differential attrition rates at the participant level. Not providing adequate information about the loss of subjects within and between groups weakens the attribution of the results to the independent variable and therefore threatens internal validity. It can also be a threat to the external validity of the study by generating a final sample that is no longer representative; thus, the results cannot be generalized to the population sampled (Barry, 2005). Reporting high overall and differential attrition also threatens the validity of the results. In our review, we opted to use a liberal criterion (as suggested by Cook et al., 2015) for overall attrition (≤30%) and differential attrition (≤10%). As a result, one more study (Heikkilä et al., 2013) also failed to meet QI 6. However, Heikkilä et al. reported no significant differences between groups across demographic, cognitive, and reading variables prior to the introduction of the intervention. Thus, the baseline equivalence requirement was met, and the potential impact of differential attrition over internal validity can be disregarded (Cook et al., 2015; WWC, 2020).

Regarding the QI for outcome measures, five studies did not report evidence of reliability assessment (Allinder et al., 2001; Esteves & Whitten, 2011; Friedland et al., 2017; Heikkilä et al., 2013; Keehn, 2003). Two studies (Begeny et al., 2010, 2011) cited previous reports to inform readers about reliability; however, we could not access the reports as they were not publicly available. Even though Begeny et al. and Esteves and Whitten used a well-known reading fluency test (i.e., DIBELS), all studies are required to report a specific coefficient or agreement level to meet this QI (Cook et al., 2015). In addition, validity (QI 7.6) was missed by a few studies (Ardoin et al., 2013; Friedland et al., 2017; Heikkilä et al., 2013; Keehn, 2003). A similar problem was reported by Begeny et al. (2018) and Chard et al. (2009). According to Gersten et al. (2005), including tests with unknown validity as measures of the dependent variable automatically reduces the power of a study.

In experimental intervention studies, it is also of paramount importance to collect data on all aspects of intervention and report treatment fidelity. Even though we would consider this to be standard practice in intervention studies, we found four studies that did not meet the criterion of implementation fidelity (Allinder et al., 2001; Friedland et al., 2017; Heikkilä et al., 2013; Keehn, 2003). Lack of description of treatment fidelity procedures in fluency intervention studies has also been reported in previous reviews (e.g., Begeny et al., 2018; Chard et al., 2009). Assessing and providing adequate evidence on implementation fidelity is necessary to evaluate the effectiveness of any given intervention. Researchers should continue to evaluate and report both the treatment fidelity and how it was established to increase both the confidence in the intervention and the other researchers' chances to replicate the study.

Previous reviews on reading fluency interventions did not find group design studies that complied with 100% of QIs (e.g., Begeny et al., 2018; Chard et al., 2009). In general, these

reviews concluded that reading fluency intervention studies provided limited descriptions of the selection of participants, treatment fidelity, outcome measures, and effect sizes. The fact that we found 16 recently published studies that met all CEC (2014) QIs, and five studies meeting all but one, might be an indicator that RCT design reading fluency interventions are improving methodologically. Another possible explanation is that by using CEC (2014) standards, we were able to pinpoint with greater ease the elements that each standard encompasses. Therefore, we highly recommend that researchers planning intervention studies consider the CEC (2014) standards.

Evidence-Based Practices Supported by Studies Meeting All QIs

Although the great majority of the studies met most of the 23 components across the eight categories of methodological quality, CEC (2014) requires that evidence-based practices must be supported by studies meeting 100% of the indicators. Also, instructional or intervention practices must be supported by at least two RCT design studies with positive treatment effects. Therefore, in this section, we discuss the evidence-based practices supported by two or more studies that met all QIs and that reported significant treatment effects (see Table 3).[10]

First, the studies appear to have some common characteristics worth noting. Most of them involved: a) participants with reading difficulties in elementary school years, b) one-on-one intervention delivery; c) simultaneous implementation of several fluency-based strategies, and d) words correct per minute as the preferred outcome measure to assess reading fluency (only

[10] Three studies met all CEC (2014) QIs but reported non-significant treatment effects on reading fluency: Patton et al. (2010), Vadasy and Sanders (2008a), and Wexler et al. (2010). Patton et al. attributed the results to the fact that most children in their sample were too young (i.e., Grade 1) to benefit from a fluency training program. In the study by Vadasy and Sanders, the fluency program had significant effects on other reading skills such as passage comprehension and vocabulary; however, no improvement was observed in WCPM. The authors concluded that deficient word reading skills might have prevented students from developing fluency. Finally, Wexler et al. argued that lack of treatment effects was due to the participants' severe reading difficulties (i.e., below the 5th percentile).

Young et al., 2018, used the Multidimensional Fluency Scale to measure the prosodic features of language).

Regarding intervention practices, one strategy was common across all studies: oral reading practice. Previous reports have highlighted the importance of extensive oral reading practice to improve reading fluency (NICHD, 2000; Rasinski & Hoffman, 2003). Moreover, oral reading practice was delivered through Repeated Reading practice by 11 out of the 13 studies with three to four repetitions (Algozzine et al., 2009; Ardoin et al., 2016; Begeny, 2011; Begeny et al., 2012; Hammerschmidt-Snidarich et al., 2018; O'Connor et al., 2007; Reed et al., 2019; Therrien et al., 2012; Vadasy & Sanders, 2008b; Vess et al., 2018; Young et al., 2018).

Interestingly, six of the studies that used Repeated Reading compared the treatment with the continuous oral reading practice of varied texts without repetition (Ardoin et al., 2016; Hammerschmidt-Snidarich et al., 2018; O'Connor et al., 2007; Reed et al., 2019; Therrien et al., 2012; Young et al., 2018). All of the studies reported significant growth from pre-test to post-test in both treatment conditions; thus, children's gains in the Repeated Reading intervention were not superior to those in the condition without repeated reading practice. These results seem to agree with Kuhn et al. (2006) who concluded that the increase in reading practice and text exposure, and not repetitive practice over the same texts, leads to familiarity with orthographic patterns and greater automaticity in word recognition, which, in turn, improves reading fluency. This conclusion seems to be in stark contrast with several meta-analyses that have identified Repeated Reading as an effective evidence-based practice (e.g., Chard et al., 2002; Lee & Yoon, 2015; Stevens et al., 2017; Therrien, 2004). However, it is important to consider that these meta-analyses included studies with different methodological designs, not just RCTs. Furthermore, Stevens et al. (2017) acknowledged that very few studies in their meta-analysis met all quality

design standards. Therefore, future reading fluency meta-analyses with Repeated Reading intervention studies should consider assessing the methodological quality of their studies to help settle this debate.

Other frequently used instructional strategies included corrective feedback (Algozzine et al., 2009; Ardoin et al., 2016; Hammerschmidt-Snidarich et al., 2018; O'Connor et al., 2007, 2010, 2013; Therrien et al., 2012; Vadasy & Sanders, 2008b; Vess et al., 2018); providing a model of fluent reading (Algozzine et al., 2009; Begeny, 2011; Begeny et al., 2012; Vadasy & Sanders, 2008b; Vess et al., 2018; Young et al., 2018); reading rate goal setting (Algozzine et al., 2009; Begeny, 2011, Begeny et al., 2012; Hammerschmidt-Snidarich et al., 2018; Vadasy & Sanders, 2008b), use of passages according to the participants' grade level equivalent (Ardoin et al., 2016; Hammerschmidt-Snidarich et al., 2018; O'Connor et al., 2007, 2013; Therrien et al., 2012), and progressively adjusting the text difficulty level (Algozzine et al., 2009; Begeny, 2011; Begeny et al., 2012; Reed et al., 2019; Therrien et al., 2012).

Less frequent instructional strategies included word drill (Begeny, 2011; Begeny et al., 2012), choral reading (Vadasy & Sanders, 2008b; Young et al., 2018), verbal cueing to read with fluency (Begeny, 2011; Begeny et al., 2012; Vadasy & Sanders, 2008b; Therrien et al., 2012), and peer-mediated fluency practice (Algozzine et al., 2009, Reed et al., 2019; Vess et al., 2018).

Beyond specific instructional strategies, differences in reading fluency outcomes between treatment conditions appear to be influenced by the frequency and dosage of intervention implementation. For example, Algozzine et al. (2009) and Begeny et al. (2012) found significant improvement from intervention programs that lasted more than five months (most studies reported duration of less than three months). In turn, Young et al. (2018) reported significant differences between treatment and control group after just a few weeks of intervention. However,

the intervention program of Young et al. was delivered frequently, three sessions a week. Other short-term intervention studies found significant improvement from pre-test to post-test after providing intensive training (at least three times per week) (Ardoin et al., 2016; Hammerschmidt-Snidarich et al., 2018; O'Connor et al. al., 2013; Reed et al., 2019; Therrien et al., 2012; Vadasy & Sanders, 2008b; Vess et al., 2018; Young et al., 2018). The importance of frequency and dosage seems to be further supported by Patton et al.'s (2010) lack of significant gains in fluency across conditions. Although Patton et al.'s study met all QIs, their intervention program consisted of two ten-minute sessions per week and the intervention lasted less than three months.

Limitations and Future Directions

Some limitations of the present study should be noted. First, we included only RCTs in this review. More studies would have been identified if other study designs such as quasi-experimental and single-case experiments were included. Although we acknowledge this as a limitation, it is important for researchers and practitioners to understand the methodological quality of RCT fluency studies in the literature. Second, we excluded the studies that have not necessarily gone through a rigorous peer review process (e.g., book chapters, dissertations). By excluding such studies, we may have overlooked a few methodologically rigorous studies. Third, only one (Ardoin et al., 2013) of the 26 intervention studies included in our review focused on the prosodic component of reading fluency. Thus, even though we refer to the methodological quality of reading fluency interventions, in essence, we can only speak about the quality of studies that focused on the automaticity component of fluency. Finally, the selected studies in this review did not focus on any specific population. In our review, the sample characteristics varied considerably across the 26 studies. Having considered the significant role of reading

fluency in the reading performance of students in all ability levels regardless of age and grades, we did not target a specific learner population.

To conclude, identifying methodologically sound reading fluency intervention studies is critical for developing evidence-based practices. Our review provided some promising findings, but also revealed areas that require researchers' attention. Currently, the research base holds moderate, but clearly increasing, proportion of high-quality studies and there is a strong need for more methodologically rigorous studies on reading fluency. The good news is that such studies are both possible and feasible in this field (as exemplified by the 16 we found), and it is clear that the researchers can utilize the operationalized set of QIs (CEC, 2014) to design, implement, and report high-quality intervention studies.

References

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

*Algozzine, B., Marr, M. B., Kavel, R. L., & Dugan, K. K. (2009). Using peer coaches to build oral reading fluency. *Journal of Education for Students Placed at Risk, 14*(3), 256-270. https://doi.org/10.1080/10824660903375735

*Allinder, R. M., Dunse, L., Brunken, C. D., & Obermiller-Krolikowski, H. J. (2001). Improving fluency in at-risk readers and students with learning disabilities. *Remedial and Special Education, 22*(1), 48-54. https://doi.org/10.1177/074193250102200106

Allington, R. (1983). Fluency: The neglected reading goal. *The Reading Teacher*, *36*(6), 556-561. Retrieved May 19, 2021, from http://www.jstor.org/stable/20198272

Allington, R. L. (2006). *Fluency: Still waiting after all these years*. In S. S. Samuels & A. E. Farstrup (Eds.), *What research has to say about fluency instruction* (p. 94–105). International Reading Association.

*Ardoin, S. P., Binder, K. S., Foster, T. E., & Zawoyski, A. M. (2016). Repeated versus wide reading: A randomized control design study examining the impact of fluency interventions on underlying reading behavior. *Journal of School Psychology, 59*, 13-38. https://doi.org/10.1016/j.jsp.2016.09.002

*Ardoin, S. P., Morena, L. S., Binder, K. S., & Foster, T. E. (2013). Examining the impact of feedback and repeated readings on oral reading fluency: Let's not forget prosody. *School Psychology Quarterly, 28*(4), 391-404. https://doi.org/10.1037/spq0000027

Baker, S. K., Smolkowski, K., Katz, R., Fien, H., Seeley, J. R., Kame'Enui, E. J., & Beck, C. T. (2008). Reading fluency as a predictor of reading proficiency in low-performing, high-poverty schools. *School Psychology Review, 37*(1), 18-37.

Barry, A. E. (2005). How attrition impacts the internal and external validity of longitudinal research. *The Journal of School Health, 75*(7), 267-270. https://doi.org/10.1111/j.1746-1561.2005.00035.x

*Begeny, J. C. (2011). Effects of the helping early literacy with practice strategies (HELPS) reading fluency program when implemented at different frequencies. *School Psychology Review, 40*(1), 149-157. https://doi.org/10.1080/02796015.2011.12087734

*Begeny, J. C., Laugle, K. M., Krouse, H. E., Lynn, A. E., Tayrose, M. P., & Stage, S. A. (2010). A control-group comparison of two reading fluency programs: The Helping Early Literacy with Practice Strategies (HELPS) program and the great leaps K-2 reading program. *School Psychology Review, 39*(1), 137-155. https://doi.org/10.1080/02796015.2010.12087795

Begeny, J. C., Levy, R. A., & Field, S. A. (2018). Using small-group instruction to improve students' reading fluency: An evaluation of the existing research. *Journal of Applied School Psychology, 34*(1), 36-64. https://doi.org/10.1080/15377903.2017.1328628

*Begeny, J. C., Mitchell, R. C., Whitehouse, M. H., Samuels, F. H., & Stage, S. A. (2011). Effects of the HELPS reading fluency program when implemented by classroom teachers with low-performing second-grade students. *Learning Disabilities Research and Practice, 26*(3),122-132. https://doi.org/10.1111/j.1540-5826.2011.00332.x

*Begeny, J. C., Ross, S. G., Greene, D. J., Mitchell, R. C., & Whitehouse, M. H. (2012). Effects of the helping early literacy with practice strategies (HELPS) reading fluency program with Latino English language learners: A preliminary evaluation. *Journal of Behavioral Education, 21*, 134-149. https://doi.org/10.1007/s10864-011-9144-7

Berninger, V. W., Abbott, R. D., Trivedi, P., Olson, E., Gould, L., Hiramatsu, S., Holsinger, M., McShane, M., Murphy, H., Norton, J., Scuilli Boyd, A., & York Westhaggen, S. (2010). Applying the multiple dimensions of reading fluency to assessment and instruction. *Journal of Psychoeducational Assessment, 28*(1), 3–18. https://doi.org/10.1177/0734282909336083

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

Chard, D. J., Ketterlin-Geller, L. R., Baker, S. K., Doabler, C., & Apichatabutra, C. (2009). Repeated reading interventions for students with learning disabilities: Status of evidence. *Exceptional Children, 75*(3), 263-281. https://doi.org/10.1177/001440290907500301

Chard, D. J., Vaughn, S., & Tyler, B. (2002). A synthesis of research on effective interventions for building reading fluency with elementary students with learning disabilities. *Journal of Learning Disabilities, 35*(1), 386-406. https://doi.org/10.1080/15377903.2017.1328628

Christ, T. J., White, M. J., Ardoin, S. P., & Eckert, T. L. (2013). Curriculum based measurement of reading: Consistency and validity across best, fastest, and question reading conditions. *School Psychology Review, 42*(4), 415-436. https://doi.org/10.1080/02796015.2013.12087463

Christ, T. J., Arañas, Y. A., Kember, J. M., Kiss, A. J., McCarthy-Trentman, A., Monaghen, B. D., et al. (2014). *Formative assessment system for teachers technical manual: earlyReading, CBMReading, aReading, aMath, and earlyMath*. Minneapolis, MN: Formative Assessment System for Teachers.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Erlbaum.

Cook, B. G., Buysse, V., Klingner, J., Landrum, T. J., McWilliam, R. A., Tankersley, M., & Test, D. W. (2015). CEC's standards for classifying the evidence base of practices in special education, *Remedial and Special Education, 36*(4), 220-234. https://doi.org/10.1177/0741932514557271

Cook, B. G., Tankersley, M., & Landrum, T. J. (2009). Determining evidence-based practices in special education. *Council for Exceptional Children, 75*(3), 365-383. https://doi.org/10.1177/001440290907500306

Council for Exceptional Children. (2014). Council for Exceptional Children Standards for evidence-based practices in special education. https://www.cec.sped.org/~/media/Files/Standards/Evidence%20based%20Practices%20and%20Practice/EBP%20FINAL.pdf

Davies, P. (1999). What is evidence-based education? *British Journal of Educational Studies, 47*(2), 108-121. https://doi.org/10.1111/1467-8527.00106

*Esteves, K. J., & Whitten, E. (2011). Assisted reading with digital audiobooks for students with reading disabilities. *Reading Horizons, 51*(1), 21-40.

Fuchs, L. S., Fuchs, D., Hosp, M. K., & Jenkins, J. R. (2001). Oral reading fluency as an indicator of reading competence: A theoretical, empirical, and historical analysis. *Scientific Studies of Reading, 5*, 239–256. doi:10.1207/S1532799XSSR0503_3

Field, A. (2013). *Discovering statistics using IBM SPSS Statistics* (4th ed.). Sage.

*Friedland, A., Gilman, M., Johnson, M., & Demeke, A. (2017). Does reading-while-listening enhance students' reading fluency? Preliminary results from school experiments in rural Uganda. *Journal of Education and Practice, 8*(7), 82-95.

Good, R. H., & Kaminski, R. A. (2002). *Dynamic Indicators of Basic Early Literacy Skills* (6th ed.). Eugene, OR: Institute for the Development of Educational Achievement.

Gersten, R., Fuchs, L. S., Coyne, M., Greenwood, C., & Innocenti, M. S. (2005). Quality indicators for group experimental and quasi-experimental research in special education. *Council for Exceptional Children, 71*(2), 149-164. https://doi.org/10.1177/001440290507100202

*Hammerschmidt-Snidarich, S. M., Maki, K. E., & Adams, S. R. (2018). Evaluating the effects of repeated reading and continuous reading using a standardized dosage of words read. *Psychology in the Schools, 56*(5), 635-651. https://dx.doi.org/10.1002/pits.22241" https://dx.doi.org/10.1002/pits.22241

Harn, B. A., Stoolmille, M., & Chard, D. J. (2008). Measuring the dimensions of alphabetic principle on the reading development of first graders. *Journal of Learning Disabilities, 41*(2), 143-157. https://doi.org/10.1177/0022219407313585

Hedges, L.V. (2007). Effect sizes in cluster-randomized designs. *Journal of Educational and Behavioral Statistics*, *32*(4), 341-370. Retrieved from https://login.ezproxy.library.ualberta.ca/login?url=https://www-proquest-com.login.ezproxy.library.ualberta.ca/docview/216383910?accountid=14474

*Heikkilä, R., Aro, M., Närhi, V., Westerholm, J., & Ahonen, T. (2013). Does training in syllable recognition improve reading speed? A computer-based trial with poor readers from second and third grade. *Scientific Studies of Reading, 17*(6), 398-414. https://doi.org/10.1080/10888438.2012.753452

Horner, R. H., Carr, E. G., Halle, J., McGee, G., Odom, S., & Wolery, M. (2005). The use of single-subject research to identify evidence-based practice in special education. *Exceptional Children, 71*(2), 165-179. https://doi.org/10.1177/001440290507100203

*Keehn, S. (2003). The effect of instruction and practice through readers theatre on young readers' oral reading fluency. *Literacy Research and Instruction, 42*(4), 40-61. https://doi.org/10.1080/19388070309558395

Kim, Y-S., Wagner, R.K., & Foster, E. (2011) Relations among oral reading fluency, silent reading fluency, and reading comprehension: A latent variable study of first-grade readers. *Scientific Studies of Reading, 15*(4), 338-362, https://doi.org/10.1080/10888438.2010.493964

Kuhn, M. R., Schwanenflugel, P. J., Meisinger, E. B., & Levy, B. A., & Rasinski, T. V. (Eds.). (2010). Aligning theory and assessment of reading fluency: Automaticity, prosody, and definitions of fluency. *Reading Research Quarterly, 45*(2), 230–251. https://doi.org/10.1598/RRQ.45.2.4

*Kuhn, M. R., Schwaneflugel, P. J., Morris, R. D., Morrow, L. M., Woo, D. G., Meisinger, E. B., Sevick, R. A., Bradley, B.A., & Stahl, S. A. (2006). Teaching children to become fluent and automatic readers. *Journal of Literacy Research, 38*(4), 357-387. https://doi.org/10.1207/s15548430jlr3804_1

Kuhn, M. R., & Stahl, S. A. (2003). Fluency: A review of developmental and remedial practices. *Journal of Educational Psychology, 95*(1), 3-21. http://dx.doi.org/10.1037/0022-0663.95.1.3

Lee, J., & Yoon, S.Y. (2017). The effects of Repeated Reading on reading fluency for students with reading disabilities: A meta-analysis. *Journal of Learning Disabilities, 50*(2), 213-224. https://doi.org/10.1177/0022219415605194

*Martens, B. K., Eckert, T. L., Begeny, J. C., Lewandowski, L. J., DiGennaro, F. D., Montarello, S. A., Arbolino, L. A., Reed, D.D., & Fiese, B. H. (2007). Effects of a fluency-building program on the reading performance of low-achieving second and third grade students. *Journal of Behavioral Education, 16*, 39-54. https://doi.org/10.1007/s10864-006-9022-x

Montague, M., & Dietz, S. (2009). Evaluating the evidence base for cognitive strategy instruction and mathematical problem solving. *Exceptional Children, 75*(3), 285-302. https://doi.org/10.1177/001440290907500302

National Institute of Child Health and Human Development. (2000). *Report of the National Reading Panel. Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction: Reports of the subgroups* (NIH Publication No. 00-4754). Washington, DC: U.S. Government Printing Office.

Odom, S. L., Brantlinger, E., Gersten, R., Horner, R. H., Thompson, B., & Harris, K. R. (2005). Research in special education: Scientific methods and evidence-based practices. *Exceptional Children, 71*(2), 137-148. https://doi.org/10.1177/001440290507100201

*O'Connor, R. E., Gutierrez, G., Teague, K., Checca, C., Kim, J. S., & Ho, T.-H. (2013). Variations in practice reading aloud: Ten versus twenty minutes. *Scientific Studies of Reading, 17*(2), 134-162. https://doi.org/10.1080/10888438.2011.624566

*O'Connor, R. E., White, A., & Swanson, H. L. (2007). Repeated reading versus continuous reading: Influences on reading fluency and comprehension. *Exceptional Children, 74*(1), 31-46. https://doi.org/10.1177/001440290707400102

*O'Connor, R. E., White, A., 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

O'Keeffe, B., Slocum, T., Burlingame, C., Snyder, K., & Bundock, K. (2012). Comparing Results of Systematic Reviews: Parallel Reviews of Research on Repeated Reading. *Education and Treatment of Children, 35*(2), 333-366. Retrieved from http://www.jstor.org/stable/42900160

Papadopoulos, T. C., Georgiou, G. K., & Kendeou, P. (2009). Investigating the double-deficit hypothesis in Greek: Findings from a longitudinal study. *Journal of Learning Disabilities, 42*(6), 528–547. https://doi.org/10.1177/0022219409338745

*Patton, B., Crosby, S., Houchins, D., & Jolivette, K. (2010). The comparative effect of fluency instruction with and without a comprehension strategy for elementary school students. *International Journal of Special Education, 25*, 100-112.

Rasinski, T.V., & Hoffman, J.V. (2003). Oral reading in the school literacy curriculum. *Reading Research Quarterly, 38*(4), 510-522. https://doi.org/10.1598/RRQ.38.4.5

Rasinski, T.V., Reutzel, C.R., Chard, D., & Linan-Thompson, S. (2011). *Reading fluency*. In M.L. Kamil, P.D. Pearson, E.B. Moje, & P. Afflerbach (Eds.), *Handbook of Reading Research* (pp. 286–319). New York: Routledge.

*Reed, D. K., Zimmermann, L. M., Reeger, A. J., & Aloe, A. M. (2019). The effects of varied practice on the oral reading fluency of fourth-grade students. *Journal of School Psychology, 77*, 24-35. https://doi.org/10.1016/j.jsp.2019.10.003

Roberts, G., Good, R., & Corcoran, S. (2005). Story retell: A fluency-based predictor of reading comprehension. *School Psychology Quarterly, 20*(3), 304-317. https://doi.org/10.1521/scpq.2005.20.3.304

Scammacca, N.K., Roberts, G., Vaughn, S., & Stuebing, K.K. (2015). A meta-analysis of interventions for struggling readers in Grades 4-12: 1980-2011. *Journal of Learning Disabilities*, *48*(4), 369-390. https://doi.org/10.1177/0022219413504995

Shinn, M. R., & Shinn, M. M. (2002). *Aimsweb training workbook: Administration and scoring of reading maze for use in general outcome measurement*. Eden Prairie, MN: Edformation.

Stevens, E. A., Walker, M. A., & Vaughn, S. (2017). The effects of reading fluency interventions on the reading fluency and reading comprehension performance of elementary students with learning disabilities: A synthesis of the research from 2001 to 2014. *Journal of Learning Disabilities, 50*(5), 576-590. https://doi.org/10.1177/0022219416638028

Therrien, W. J. (2004). Fluency and Comprehension Gains as a Result of Repeated Reading: A Meta-Analysis. *Remedial and Special Education*, *25*(4), 252–261. https://doi.org/10.1177/07419325040250040801

*Therrien, W. J., Kirk, J. F., & Woods-Groves, S. (2012). Comparison of a reading fluency intervention with and without passage repetition on reading achievement. *Remedial and Special Education, 33*(5), 309-319. https://doi.org/10.1177/0741932511410360

Torgerson, C., & Torgerson, D. (2001). The need for randomised controlled trials in educational research. *British Journal of Educational Studies, 49*(3), 316-328. Retrieved from http://www.jstor.org/stable/3122243

Torgesen, J., Wagner, R. K., & Rashotte, C. (1999). *Test of Word Reading Efficiency.* Austin, TX: Pro-Ed.

Torgesen, J. K., Wagner, R. K., & Rashotte, C. A. (2012). *TOWRE-2Test of Word Reading Efficiency* (2nd ed.). Austin, TX: Pro-Ed.

*Vadasy, P. F., & Sanders, E. A. (2008a). Benefits of repeated reading intervention for low-achieving fourth- and fifth-grade students. *Remedial and Special Education, 29*(4), 235-249. https://doi.org/10.1177/0741932507312013

*Vadasy, P. F., & Sanders, E. A. (2008b). 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

*Vess, S. F., Begeny, J. C., Norwalk, K. E., & Ankey, R. N. (2018). Tier 2 reading fluency interventions with middle school students: A comparison of the HELPS-SG program and a teacher-directed evidence-based intervention. *Journal of Applied School Psychology, 34*(4), 316-337. https://doi.org/10.1080/15377903.2018.1443985

What Works Clearinghouse. (2020). *What Works Clearinghouse Standards Handbook*, Version 4.1. Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance. This report is available on the What Works Clearinghouse website at https://ies.ed.gov/ncee/wwc/handbooks.

*Wexler, J., Vaughn, S., Roberts, G., & Denton, C. A. (2010). The efficacy of repeated reading and wide reading practice for high school students with severe reading disabilities.

Learning Disabilities Research and Practice, 25(1), 2-10. https://doi.org/10.1111/j.1540-5826.2009.00296.x

Wiederholt, J. L., & Bryant, B. R. (2001). *Gray Oral Reading Test-4*. Austin, TX: Pro-Ed.

Wiederholt, J. L., & Bryant, B. R. (2012). GORT-5 *Gray Oral Reading Tests* (5th ed.). Austin, TX: Pro-Ed.

Woodcock, R. W., McGrew, K. S., & Mather, N. (2007). *Woodcock-Johnson Tests of Cognitive Abilities and Tests of Achievement* (3rd ed). Rolling Meadows, IL: Riverside.

*Young, C., Pearce, C., Gomez, J., Christensen, R., Pletcher, B., & Fleming, K. (2018). Read two impress and the neurological impress method: Effects on elementary students' reading fluency, comprehension, and attitude. *The Journal of Educational Research, 111*(6), 657-665. https://doi.org/10.1080/00220671.2017.1393650

Zutell, J., & Rasinski, T.V. (1991). Training teachers to attend to their students' reading fluency. *Theory into Practice*, 30, 211–217

Table 1

Quality Indicators for Assessing Methodological Rigor of RCT Intervention Studies

Quality indicator	Indicator element
1. Context and setting	1.1 Describes critical features of the context or setting (school or classroom).
2. Participants	2.1 Describes participants' demographics. 2.2 Describes disability or risk status and method for determining status.
3. Intervention agents	3.1 Describes role of the intervention agent, and background when relevant to review. 3.2 Describes agents' training or qualifications.
4.Description of practice	4.1 Describes detailed intervention procedures and agents' actions or cites accessible sources for that information. 4.2 Describes, when relevant, study materials described or cites accessible source.
5.Implementation fidelity	5.1 Assesses and reports implementation fidelity related to adherence with direct, reliable measures. 5.2 Assesses and reports implementation fidelity related to dosage or exposure with direct, reliable measures 5.3 Assesses and reports implementation fidelity (adherence/dosage) throughout intervention and by unit of analysis.
6. Internal validity	6.1 Researcher controls and systematically manipulates independent variable. 6.2 Describes baseline or control conditions 6.3 During baseline or control conditions, participants have no/extremely limited access to intervention. 6.4 Random assignment of groups 6.8 Attrition is low across groups 6.9 Attrition differential is low between groups.
7. Outcome measures/ dependent variables	7.1 Outcomes are socially important. 7.2 Defines and describes measurement of dependent variables. 7.3 Reports effects of intervention on all measures. 7.5 Provides evidence of adequate internal reliability. 7.6 Provides evidence of adequate validity.
8. Data analysis	8.1Techniques are appropriate for detecting change in performance. 8.3 Reports appropriate effect size statistic(s) or provides data to calculate the effect size.

Source. Council for Exceptional Children (2014).

Table 2

Methodological Rigor by Quality Indicator

Intervention studies	Quality Indicators								
	1.Context and setting	2.Participants	3.Intervention agent	4.Description of practice	5.Implementation fidelity	6.Internal validity	7.Outcome measures	8.Data analysis	QI met (%)
Algozzine et al. (2009)	1/1	2/2	2/2	2/2	3/3	6/6	5/5	2/2	8 (100)
Allinder et al. (2001)	1/1	2/2	2/2	2/2	1/3	6/6	4/5	2/2	6 (75)
Ardoin et al. (2013)	1/1	2/2	2/2	2/2	3/3	6/6	4/5	2/2	7 (88)
Ardoin et al. (2016)	1/1	2/2	2/2	2/2	3/3	6/6	5/5	2/2	8 (100)
Begeny et al. (2010)	1/1	2/2	2/2	2/2	3/3	6/6	4/5	2/2	7 (88)
Begeny (2011)	1/1	2/2	2/2	2/2	3/3	6/6	5/5	2/2	8 (100)
Begeny et al. (2011)	1/1	2/2	2/2	2/2	3/3	6/6	4/5	2/2	7 (88)
Begeny et al. (2012)	1/1	2/2	2/2	2/2	3/3	6/6	5/5	2/2	8 (100)
Esteves & Whitten, (2011)	1/1	2/2	2/2	2/2	3/3	6/6	3/5	2/2	7 (88)
Friedland et al. (2017)	1/1	2/2	2/2	1/2	0/3	6/6	3/5	2/2	5 (63)
Hammerschmidt-Snidarich et al. (2018)	1/1	2/2	2/2	2/2	3/3	6/6	5/5	2/2	8 (100)
Heikkilä et al. (2013)	1/1	2/2	2/2	2/2	0/3	5/6	3/5	2/2	5 (63)
Keehn (2003)	1/1	2/2	2/2	2/2	1/3	4/6	3/5	2/2	5 (63)
Kuhn et al. (2006)	1/1	2/2	2/2	2/2	3/3	4/6	5/5	2/2	7 (88)
Martens et al. (2007)	1/1	2/2	2/2	2/2	3/3	4/6	5/5	2/2	7 (88)
O'Connor et al. (2007)	1/1	2/2	2/2	2/2	3/3	6/6	5/5	2/2	8 (100)
O'Connor et al. (2013)	1/1	2/2	2/2	2/2	3/3	6/6	5/5	2/2	8 (100)
O'Connor et al. (2010)	1/1	2/2	2/2	2/2	3/3	6/6	5/5	2/2	8 (100)
Patton et al. (2010)	1/1	2/2	2/2	2/2	3/3	6/6	5/5	2/2	8 (100)
Reed et al. (2019)	1/1	2/2	2/2	2/2	3/3	6/6	5/5	2/2	8(100)
Therrien et al. (2012)	1/1	2/2	2/2	2/2	3/3	6/6	5/5	2/2	8 (100)
Vadasy & Sanders (2008a)	1/1	2/2	2/2	2/2	3/3	6/6	5/5	2/2	7 (88)
Vadasy & Sanders (2008b)	1/1	2/2	2/2	2/2	3/3	6/6	5/5	2/2	8 (100)
Vess et al. (2018)	1/1	2/2	2/2	2/2	3/3	6/6	5/5	2/2	8 (100)
Wexler et al. (2010)	1/1	2/2	2/2	2/2	3/3	6/6	5/5	2/2	8 (100)
Young et al. (2018)	1/1	2/2	2/2	2/2	3/3	6/6	5/5	2/2	8 (100)
Quality indicators met (%)	26 (100)	26 (100)	26 (100)	25 (96)	22 (85)	22 (85)	18 (65)	26 (100)	

Table 3

Evidence-based Practices Supported by RCT Studies Meeting All CEC (2014) Quality Indicators

Intervention studies	Instructional Practices											
	Oral Reading	Repeated Reading	Continuous Reading	Corrective Feedback	Fluency Goal Setting	Grade Equivalent Texts	Adjust Text Difficulty	Word Drill	Modeling	Peer Reading	Fluency Cueing	Choral Reading
Algozzine et al. (2009)	✓	✓		✓	✓		✓		✓	✓		
Ardoin et al. (2016)	✓	✓	✓	✓		✓						
Begeny (2011)	✓	✓			✓		✓	✓	✓		✓	
Begeny et al. (2012)	✓	✓			✓		✓	✓	✓		✓	
Hammerschmidt-Snidarich et al. (2018)	✓	✓	✓	✓	✓	✓						
O'Connor et al. (2007)	✓	✓	✓	✓		✓						
O'Connor et al. (2010)	✓		✓	✓								
O'Connor et al. (2013)	✓		✓	✓		✓						
Reed et al. (2019)	✓	✓	✓				✓			✓		
Therrien et al. (2012)	✓	✓	✓	✓		✓	✓				✓	
Vadasy & Sanders (2008b)	✓	✓		✓	✓				✓		✓	✓
Vess et al. (2018)	✓	✓		✓					✓	✓		
Young et al. (2018)	✓	✓	✓						✓			✓

Figure 1

Flow Diagram for the Search and Inclusion of Studies