HHS Public Access

Author manuscript

Except Child. Author manuscript; available in PMC 2021 October 07.

Published in final edited form as:

Except Child. 2021 July ; 87(4): 397–417. doi:10.1177/0014402921993406.

Current State of the Evidence: Examining the Effects of Orton-Gillingham Reading Interventions for Students With or at Risk for Word-Level Reading Disabilities

Elizabeth A. Stevens[1], **Christy Austin**[3], **Clint Moore**[2], **Nancy Scammacca**[2], **Alexis N. Boucher**[2], **Sharon Vaughn**[2]

[1]Georgia State University

[2]University of Texas at Austin

[3]University of Utah

Abstract

Over the past decade, parent advocacy groups led a grassroots movement resulting in most states adopting dyslexia-specific legislation, with many states mandating the use of the Orton-Gillingham approach to reading instruction. Orton-Gillingham is a direct, explicit, multisensory, structured, sequential, diagnostic, and prescriptive approach to reading for students with or at risk for word-level reading disabilities (WLRD). Evidence from a prior synthesis and What Works Clearinghouse reports yielded findings lacking support for the effectiveness of Orton-Gillingham interventions. We conducted a meta-analysis to examine the effects of Orton-Gillingham reading interventions on the reading outcomes of students with or at risk for WLRD. Findings suggested Orton-Gillingham reading interventions do not statistically significantly improve foundational skill outcomes (i.e., phonological awareness, phonics, fluency, spelling; effect size [ES] = 0.22; $p = .40$), although the mean ES was positive in favor of Orton-Gillingham-based approaches. Similarly, there were not significant differences for vocabulary and comprehension outcomes (ES = 0.14; $p = .59$) for students with or at risk for WLRD. More high-quality, rigorous research with larger samples of students with WLRD is needed to fully understand the effects of Orton-Gillingham interventions on the reading outcomes for this population.

Approximately 13% of public school students receive special education services under the Every Student Succeeds Act (ESSA; 2015–2016), with 34% identified with a specific learning disability (SLD; Depaoli et al., 2015). Approximately 85% of students identified with SLD have a primary disability in the area of reading (Depaoli et al., 2015). Reading achievement data from the National Assessment of Educational Progress demonstrate that students with disabilities persistently perform far below their nondisabled peers in reading, with only 32% performing at a basic level and 30% performing above a basic level (National Center for Education Statistics, 2017, 2019). The majority of students reading below grade level after the early elementary grades require remediation in word-level decoding and reading fluency (Scammacca et al., 2013; Vaughn et al., 2010).

Corresponding Author: Elizabeth A. Stevens, Department of Learning Sciences, College of Education and Human Development, Georgia State University, P.O. Box 3978, Atlanta, GA 30302-3978, USA. estevens11@gsu.edu.

The International Dyslexia Association (IDA; 2002) and National Institute of Child Health and Human Development (Eunice Kennedy Shriver National Institute of Child Health and Human Development, n.d.) define dyslexia as an SLD that is neurobiological in origin and characterized by difficulties with accurate or fluent word recognition, poor spelling, and poor decoding. These word-reading deficits result in secondary consequences, including reduced exposure to text, poor vocabulary and background knowledge development, and limited reading comprehension (Lyon et al., 2003). Over the past decade, considerable support for screening, assessing, and providing appropriate educational services for students with dyslexia has occurred at local and state levels (National Center on Improving Literacy [NCIL], 2021). Forty-seven states established legislation to protect the rights of individuals with dyslexia beyond the requirements of the Individuals With Disabilities Education Act (IDEA, 2004; U.S. Department of Education, 2019; NCIL, 2019). Students with dyslexia may receive specialized instruction as a student with SLD under ESSA (2015) or through Section 504 of the Rehabilitation Act (1973). These students demonstrate word-reading and spelling difficulties, so they may be identified with SLD in basic reading, reading fluency, or written expression (Odegard et al., 2020). Because dyslexia can be identified as a SLD, some schools may not utilize the dyslexia label when identifying a student. All students with word-level reading disabilities (WLRD) require instruction to address their difficulties in word recognition, spelling, and decoding.

Many states require teacher training and implementation of Orton-Gillingham (OG) methodology (see Table 1). The OG approach to reading instruction is a “direct, explicit, multisensory, structured, sequential, diagnostic, and prescriptive way to teach reading and spelling” (OG Academy, 2020 October 14) commonly used for students with and at risk for reading disabilities, such as dyslexia (Ring et al., 2017). The OG Academy further defines each descriptor of the OG approach, stating OG is *direct* and *explicit* by “employing lesson formats which ensure that students understand what is to be learned, why it is to be learned, and how it is to be learned”; *structured* and *sequential* by “presenting information in a logical order which facilitates student learning and progress, moving from simple, well-learned material to that which is more and more complex as mastery is achieved”; *diagnostic* in that “the instructor continuously monitors the verbal, nonverbal, and written responses of the student to identify and analyze both the student’s problems and progress” and *prescriptive* in that lessons “contain instructional elements that focus on a student’s difficulties and build upon a student’s progress from the previous lessons”; and finally, *multi-sensory* by “using all learning pathways: seeing, hearing, feeling, and awareness of motion” (OG Academy, 2020 October 14, “What Is the Orton-Gillingham Approach?” section).

The OG Institute for Multi-Sensory Education (2020a October 11, “What Orton-Gillingham Is All About” section) further explains multi-sensory instruction as involving the simultaneous use of “sight, hearing, touch, and movement to help students connect and learn the concepts” and identifies this as the “most effective strategy for children with difficulties in learning to read” (Institute for Multi-Sensory Education, 2020b October 12, “Components of Multi-Sensory Instruction” section). Examples of visual activities include seeing words and graphemes via charts, flashcards, lists, visual cues, and pictures; examples of auditory activities include hearing sounds and directions aloud, rhymes, songs, and

mnemonics; examples of kinesthetic and tactile activities include fine motor (e.g., finger tapping, use of hands to manipulate objects, writing graphemes in sand, finger tracing) and whole-body movements (e.g., arm tapping, moving in order to focus and learn; Institute for Multi-Sensory Education, 2020b October 12). Most early reading programs emphasize the visual (discrimination between letters, seeing a word) and auditory (naming sounds, reading words aloud) senses, and some include the kinesthetic or tactile sense (handwriting practice, spelling words). OG intervention is described as different from others in the *simultaneous* use of visual, auditory, and kinesthetic or tactile experiences. An example of all three senses being simultaneously employed could involve simultaneously seeing the letters *sh* on a sound card (visual), hearing the sound/sh/made by the letters *sh*, (auditory), and tracing the letters *sh* on a textured mat (kinesthetic or tactile). When the OG approach was first introduced in the early 1900s, it was unique for (a) its emphasis on direct, explicit, structured, and sequential instruction individually introducing each phonogram and the rules for blending phonograms into syllables and (b) utilizing visual, auditory, and kinesthetic teaching techniques reinforcing one another (Ring et al., 2017). More recently, non-OG programs have adopted many of the descriptors or characteristics of the OG approach (direct, explicit, structured, sequential, diagnostic, and prescriptive word-reading instruction), and therefore OG and non-OG programs have overlapping characteristics. However, OG remains widely used with students with WLRD, in part, due to dyslexia legislation (Uhry & Clark, 2005; WWC, 2010).

The professional standards of the Council for Exceptional Children (2015) and U.S. federal regulations of the Every Student Succeeds Act (2015–2016) reauthorized by the Elementary and Secondary Education Act mandate the use of evidence-based practices and interventions to the greatest extent possible. However, the efficacy of OG instruction remains unclear based on results of prior systematic reviews. For example, Ritchey and Goeke (2006) published a systematic review of OG interventions implemented with elementary, adolescent, and college students between 1980 and 2005. Findings demonstrated limited evidence to support the use of OG instruction. The authors noted the limited number of studies ($N = 12$) and the poor methodological rigor of those studies, calling for additional research investigating OG interventions; others in the field have also noted the lack of rigorous research examining OG interventions (Lim & Oei, 2015; Ring et al., 2017). Since the Ritchey and Goeke (2006) review, the What Works Clearinghouse (WWC) also reviewed branded OG programs (i.e., published, commercially available OG programs; WWC, 2010a, 2010b, 2010c, 2010d, 2010e, 2010f, 2010h, 2010i, 2012, 2013) and unbranded OG interventions (i.e., unpublished curricula based on the principles of a sequential, multisensory OG approach to teaching reading; WWC, 2010g), finding little evidence supporting the effectiveness of the OG methodology.

Rationale and Purpose

Despite the limited evidence supporting its efficacy, OG has become a popular, widely adopted and used approach to providing reading instruction to students with or at risk for WLRD (Lim & Oei, 2015; Ring et al., 2017). Laws requiring the use of evidence-based practices for addressing WLRD may also mandate the use of OG—seemingly assuming that OG approaches are associated with statistically significant effects for target students.

Considering that the WWC reviews occurred 10 years ago and the Ritchey and Goeke (2006) review occurred nearly 15 years ago, we aimed to update and extend Ritchey and Goeke's review to inform the field on the current state of the evidence regarding this popular and widely utilized instructional approach. We addressed the following research question: What are the effects of OG interventions for students identified with or at risk for WLRD in Grades K through 12? Due to the lack of methodological rigor noted for studies included in these prior reviews, we also examined whether the effects are moderated by study quality, as determined by research design, the nature of the instruction in the comparison condition, implementation fidelity, and year of publication.

Method

Operational Definitions

Due to the inconsistent application of the term "dyslexia" and identification of students with dyslexia across the literature, we included studies with participants formally diagnosed with dyslexia and those without a diagnosis but who exhibited WLRD (i.e., students at risk for dyslexia, students with a learning disability in reading, or struggling readers performing in the bottom quartile on a standardized reading measure). We refer to this population as "students with or at risk for WLRD."

We utilized WWC definitions of "branded OG programs" and "unbranded OG interventions" to guide this review. *Branded OG programs* are "curricula based on the principles of sequential, multisensory Orton-Gillingham approach to teaching reading" (WWC, 2010a). To include a comprehensive list of branded programs in this review, authors utilized each of the branded programs identified by WWC (i.e., Alphabetic Phonics, Barton Reading and Spelling System, Fundations, Herman Method, Wilson Reading System, Project Read, and Dyslexia Training Program; WWC, 2010a, 2010b, 2010c, 2010d, 2010e, 2010h, 2010i). We also included additional branded programs identified in Ritchey and Goeke's (2006) initial review (i.e., Project ASSIST, the Slingerland Approach, the Spalding Method, Starting Over) or identified in Sayeski (2019; i.e., Language!, Lindamood Bell, Recipe for Reading, S.P.I.R.E., Take Flight, and the Writing Road to Reading).

Unbranded OG interventions (WWC, 2010g) are interventions based on general OG principles or interventions that combine multiple branded products based on OG principles. We required authors to self-identify instruction as OG (i.e., the authors identified the intervention as OG instruction in the manuscript) to be included in this review as an unbranded intervention.

Search Procedures

To locate all relevant studies examining OG interventions, we searched published and unpublished studies through March 2019. We did not specify a start date to conduct a comprehensive review of the evidence base, including and extending studies from Ritchey and Goeke (2006). We conducted a computerized search of three electronic databases (i.e., Education Source, Educational Resources Information Clearinghouse, and PsycINFO) and ProQuest Dissertations using the following search terms: "Orton-

Gillingham," "Wilson Reading," "Wilson Language," "Alphabetic Phonics," "Herman Method," "Project ASSIST," "Slingerland Approach," "Spalding Method," "Starting Over," "Project Read," "Take Flight," "Barton Reading & Spelling System," "Barton Reading and Spelling System," "Fundations," "Dyslexia Training Program," "Recipe for Reading," or "S.P.I.R.E." See Figure 1 for a PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses; Liberati et al., 2009) diagram detailing the search process.

We conducted a 2-year hand search of the following journals: *Annals of Dyslexia, Exceptional Children, Journal of Learning Disabilities, The Journal of Special Education, Learning Disabilities Research & Practice*, and *Learning Disability Quarterly*. We selected these journals because Ritchey and Goeke's (2006) conducted a hand search of these journals, and they contain relevant empirical research in the field of intervention research and special education. We identified two additional articles in the hand search. Finally, we conducted an ancestral search using the reference lists from WWC reports of branded and unbranded programs (WWC, 2010a, 2010b, 2010c, 2010d, 2010e, 2010f, 2010g, 2010h, 2010i, 2012, 2013); we identified 16 additional studies in the WWC reports. After removing the duplicates, we screened 354 abstracts. The first two authors independently reviewed 10% of the abstracts to determine if the full text of the study should be excluded or further reviewed for inclusion in the systematic review. The authors sorted these abstracts with 98% reliability and proceeded with sorting the remaining abstracts. We reviewed the full text of 109 articles, and 24 studies met inclusion criteria.

Inclusion Criteria

We included studies that met the following criteria:

1. Published in a peer-reviewed journal or an unpublished dissertation printed in English through March 2019.
2. Employed an experimental, quasiexperimental, or single-case design providing a treatment and comparison to determine the experimental effect (i.e., multiple-treatment, single-group, pre-test-posttest, AB single-case, qualitative, and case study designs were excluded).
3. Included participants in kindergarten through 12th grade identified with dyslexia, reading disabilities, learning disabilities, at risk for reading failure, or reading difficulty or at risk for reading failure as determined by low performance on a standardized reading measure. Studies with additional participants (e.g., students without reading difficulty) were included if at least 50% of the sample included the targeted population or disaggregated data were provided for these students. We included English learners, students with behavioral disorders, and students with attention deficit hyperactivity disorder if they were also identified with reading difficulty as described previously. We excluded studies targeting students with autism, intellectual disabilities, and vision or hearing impairments.
4. Examined a branded or unbranded OG reading intervention (see Operational Definitions) provided one-on-one or in small groups (i.e., we excluded OG instruction provided in the whole-class, general education setting). We excluded

multicomponent interventions (e.g., interventions targeting OG and additional components of reading instruction, such as vocabulary).

5. Assessed at least one of the following dependent variables: word reading, oral reading fluency, phonological awareness, phonics, spelling, vocabulary, listening comprehension, or reading comprehension.

Coding Procedures

We coded studies that met inclusion criteria using a protocol (Vaughn et al., 2014) developed for education-related intervention research based on study features described in the WWC Design and Implementation Device (Valentine & Cooper, 2008) and used in previous meta-analyses (e.g. Stevens et al., 2018).

Data extraction and quality coding.—We extracted the following data from each study: (a) participant information (e.g., socioeconomic status, risk type, age, grade, and criteria used for the selection of participants); (b) research design; (c) a detailed description of all treatment and comparison groups; (d) the length, frequency, and duration of the intervention provided; (e) measures; and (f) results and effect sizes (ESs).

We coded each study for study quality based on three indicators: research design, comparison group, and implementation fidelity. We utilized the coding procedures applied in a previous meta-analysis examining study quality (Austin et al., 2019). For each indicator, we assigned a rating of exemplary, acceptable, or unacceptable. For research design, a study received an exemplary rating for utilizing a randomized design with a sufficiently large sample (≥20), an acceptable rating for use of a randomized design with an insufficient sample size (<20) or a nonrandomized design with a large sample, and an unacceptable rating for use of a nonrandomized design with a small sample size. For implementation fidelity, we rated a study exemplary if clear, replicable operational definitions of treatment procedures were provided, data demonstrated high procedural fidelity (≥75%), and interobserver reliability was equal to or exceeded .90. A study received an acceptable rating if adequate operational definitions of treatment procedures were provided, data demonstrated high procedural fidelity (≥75%), and interobserver reliability was at least .80. A study received an unacceptable rating if the description of treatment was such that replication would not be possible, data demonstrated poor implementation fidelity (<75%), data demonstrated poor intercoder agreement (<.80), or fidelity was not reported. For the comparison group indicator, studies received an exemplary rating if the majority of the students in the comparison group received an alternate treatment (i.e., supplemental, small-group reading intervention), an acceptable rating if the comparison group served as an active control (i.e., minimal intervention, business-as-usual intervention with minimal description), and an unacceptable rating if the comparison group received no intervention or insufficient information was provided to determine what the group received.

We used the gold-standard method (Gwet, 2001) to establish interrater reliability prior to coding. The first author, a researcher with experience using and publishing systematic reviews with the code sheet, provided an initial 4.5-hr training session to the remaining authors (i.e., PhD level researcher and two PhD graduate research assistants studying

reading intervention research). The researcher described the code sheet and modeled each step of the coding process for a sample intervention study, and then the research assistants practiced coding additional intervention studies of different design types. Upon completion of the training, the coders independently coded a study to establish reliability. Coders achieved interrater reliability scores of .96, .92, and .98 as determined by the number of items in agreement divided by the total number of items. After establishing initial reliability, each study was independently coded by two coders. The coders met to review each code sheet and to identify and resolve any discrepancies. When the coders were unable to resolve a specific code, the first author reviewed the study, and the author team made final decisions by consensus.

Meta-Analysis Procedures for the Group Design Studies

Standardized mean difference ESs were computed as Hedges's *g* for all studies that used an experimental or quasiexperimental group design. To compute *g*, we used the means, standard deviations, and group sizes for the treatment and comparison groups when study authors reported these data. When studies did not contain this information, we computed *g* from Cohen's *d* and group sample sizes or from group means, sample sizes, and the *p* value of tests of group differences. All ESs and standard errors were computed using Comprehensive Meta-Analysis (Version 3.3.070) software (Borenstein et al., 2014).

Data analysis.—ESs from measures of foundational reading skills (including phonological awareness, decoding, word identification, fluency, and spelling) were meta-analyzed separately from ESs for measures of vocabulary and comprehension (reading comprehension, listening comprehension, and vocabulary) in order to determine the effects of OG instruction versus comparison instruction on both types of reading outcomes. Fifteen studies reported results for one or more foundational skill measures, and 10 studies reported results for one or more measures of vocabulary and comprehension. Most studies in each meta-analysis reported results on multiple foundational skill and/or reading comprehension measures, and some included comparisons of two or more interventions with a comparison condition. As a result, we used robust variance estimation (RVE; Hedges et al., 2010) in conducting the meta-analyses. RVE accounts for the dependency within a study when the study contributes more than one ES to a meta-analysis by adjusting the standard errors within a meta-regression model.

Using the robumeta package for R (Fisher & Tipton, 2015), we calculated beta coefficients for the meta-regression model, mean ESs, and standard errors. Because the meta-analyses included fewer than 40 studies, we implemented a small-sample correction to avoid inflating Type I error (Tipton, 2015; Tipton & Pustejovsky, 2015). The mean within-study correlation between all pairs of ESs (ρ) must be specified to estimate study weights and calculate the variance between studies when using RVE. As shown by Hedges et al. (2010), the value of ρ has a minimal effect on meta-regression results when implementing RVE. As recommended by Hedges et al., we evaluated the impact of ρ values of .2, .5, and .8 on the model parameters. The differences were minimal. We reported results from the model where ρ = .8. Using robumeta, we first estimated intercept-only models to compute the weighted mean ESs and standard errors for foundational skill measures and vocabulary and comprehension

measures. Next, two moderator variables (study quality score and publication year) were included in the meta-regression models as covariates.

Results

We applied a more stringent inclusion criteria than that used by Ritchey and Goeke (2006; i.e., we excluded college participants and studies that examined OG instruction in general education, whole-class settings). The previous review included 12 studies examining OG instruction using primarily quasiexperimental designs. In the current corpus, we identified 24 studies. Of the 24 studies, six were also included in the original review; we excluded the remaining six studies because (a) they included college students, (b) they provided OG instruction in general education settings, or (c) we were unable to determine if participants were students with or at risk for WLRD. We included 16 of the 24 studies in the quantitative meta-analysis (see Table 2). We were unable to include the remaining eight studies due to insufficient sample size (i.e., <10 in each group; Giess, 2005; Hook, et al. 2001; Wade, 1993; Wille, 1993; Young, 2001) or insufficient information provided to calculate ESs (Kuveke, 1996; Oakland et al., 1998; Simpson et al., 1992).

The weighted mean ES for the 15 studies that included one or more measures of foundational skills was 0.22 (SE = 0.25; 95% confidence interval [CI] = [−0.33, 0.77]). The mean ES was not statistically significantly different from zero (p = .40), indicating that students who received OG interventions did not experience significantly larger effects on these measures than students who received comparison reading instruction. The I^2 estimate of the percentage of heterogeneity in ESs between studies that likely is not due to chance was 88.74%, which is considered large and sufficient for conducting moderator analyses to determine if one or more moderator variables can explain the heterogeneity (Higgins et al., 2003). The τ^2 estimate of the true variance in the population of effects for this analysis was .71, which also indicates the presence of considerable heterogeneity in the effects of the studies in the analysis. However, the meta-regression model that included quality score and publication year as covariates indicated that neither moderator significantly predicted study ES (for quality score, b = 0.43, SE = 1.03, p = .70; for publication year, b = −0.04, SE = 0.03, p = .25).

In the meta-analysis of vocabulary and comprehension measures, the weighted mean ES for the 10 included studies was 0.14 (SE = 0.23; 95% CI = [−0.39, 0.66]). As with the foundational skills measures, the effect of OG interventions across studies was not significantly different from zero (p = .59), meaning that students in OG interventions did not experience significantly greater benefit than students in the comparison condition. The I^2 estimate of heterogeneity not likely due to chance variation was 81.53%, which is considered large (Higgins et al., 2003), and the τ^2 estimate of the true variance in the population of effects was .38. As in the analysis of foundational skills measures, quality score was not a significant predictor of ES magnitude (b = 0.49, SE = 0.55, p = .47). However, publication year did predict the magnitude of ESs, with older studies having larger effects (b = −0.05, SE = 0.02, p = .02).

Publication Bias

We evaluated the study corpus for each meta-analysis for the likelihood of studies with null effects being absent from the analysis due to publication bias. Duval and Tweedie's (2000) trim-and-fill approach indicated that no studies likely were missing from either meta-analysis as a result of publication bias. Egger's regression test (Egger et al., 1997) also did not indicate that publication bias was present in the corpus used for each of the meta-analyses.

Study Quality

We examined study quality in terms of three indicators: research design, comparison condition instruction, and implementation fidelity (Table 3). Studies received a mean quality rating from 0 to 2, with scores interpreted as unacceptable (0), acceptable (1), and exemplary (2). The mean quality rating for research design was 0.95, with most studies receiving unacceptable or acceptable ratings. Few studies conducted randomized designs that included sufficiently large samples, and all but one of these studies were conducted after the previous review (i.e., Christodoulou et al., 2017; Reuter, 2006; Torgesen et al., 2007; Wanzek & Roberts, 2012). Authors employed a quasiexperimental design in 15 studies and a randomized design in nine studies. The comparison group instruction resulted in a mean rating of 1.0. Twelve studies provided exemplary instruction to the comparison group, meaning the majority of the students received an alternate treatment, such as business-as-usual supplemental intervention. The remaining studies received unacceptable ratings because either students in the comparison group received no instruction or not enough information was reported to determine the type of instruction. Finally, implementation fidelity resulted in a mean rating of 0.17, with most studies (n = 20) receiving an unacceptable rating due to a lack of implementation fidelity data reported.

Discussion

We aimed to systematically review existing evidence of the effects of OG interventions for students with or at risk for WLRD through 2019. We also examined whether study quality (i.e., determined by research design, comparison condition instruction, implementation fidelity, and publication year) moderated the effects of OG interventions.

Is There Scientific Evidence to Support OG Instruction for Students With WLRD?

The major finding in Ritchey and Goeke's (2006) review revealed the research was simply insufficient, in the number of studies conducted and study quality, to support OG instruction as an evidence-based practice. Nearly 15 years later, the results of this meta-analysis suggest OG interventions do not statistically significantly improve foundational skill outcomes or vocabulary and reading comprehension outcomes for students with or at risk for WLRD over and above comparison condition instruction. Despite the finding that effects were not statistically significant, we interpret a mean effect of 0.22 as indicating promise that OG may positively impact student outcomes. For students with significant WLRD, who often demonstrate limited response to early reading interventions (Nelson et al., 2003; Tran et al., 2011), 0.22 may be indicative of educationally meaningful reading progress. However, until a sufficient number of high-quality research studies exist, we

echo the cautionary recommendation provided in that initial review: Despite the continued widespread acceptance, use, and support for OG instruction, there is little evidence to date that these interventions significantly improve reading outcomes for students with or at risk for WLRD over and above comparison group instruction.

Methodological Rigor

On a scale of 0 to 2 (0 is unacceptable, 1 is acceptable, and 2 is exemplary), the mean quality rating across studies and quality indicators was 0.76, which falls below the acceptable level and suggests concerns about the study quality represented in this corpus. In the foundational skill and vocabulary and comprehension meta-analyses, study quality did not significantly predict study ES, indicating student outcomes did not differ for studies rated unacceptable, acceptable, and exemplary. A closer inspection of the quality ratings for individual studies may help to explain the lack of relationship found between study quality and ES. The five studies that received unacceptable design ratings (i.e., authors used a nonrandomized design with a small sample) were not included in the meta-analysis because sample size was less than 10 (Giess, 2005; Hook et al., 2001; Wade, 1993; Wille, 1993) or insufficient information was provided to calculate ESs (Kuveke, 1996). Three of these studies received the lowest overall quality ratings (i.e., 0.00; Kuveke, 1996; Wade, 1993; Wille, 1993). It may be that the limited number of studies ($n = 16$) and the lack of variability in quality ratings (i.e., only three studies receive mean rating above 1.00; three studies with mean rating of 0.00 were dropped from the meta-analysis) prohibited detecting a relationship between reading outcomes and study quality.

The current corpus revealed limited reporting of implementation fidelity ($M = 0.17$). This finding is particularly concerning given fidelity is a group design quality indicator (Gersten et al., 2005). With the exception of four studies that received acceptable (Fritts, 2016; Geiss, 2005; Wanzek & Roberts, 2012) or exemplary (Torgesen et al., 2007) ratings, the remaining studies did not provide implementation fidelity data or described it with insufficient detail such that replication would not be possible. Knowing whether the intervention was implemented as intended is essential to establishing a causal connection between the independent and dependent variables, raising concerns about the internal validity of the included studies, particularly given the importance of measuring multiple dimensions of implementation fidelity (i.e., procedural, dosage, quality; Gersten et al., 2005).

We also examined publication year as a moderator of intervention effectiveness. Of the 16 studies included in the meta-analysis, one was published in 1979, six were published in the 1990s, two were published between 2000 and 2010, and seven were published after 2010. Scammacca and colleagues (2013) reported a decline in ESs for reading interventions over time, with statistically significantly different mean effects for studies published in 1980 to 2004 and 2005 to 2011. We expected studies conducted more recently would result in smaller effects due to an increased use in standardized measures, more rigorous research designs, and improvement in business-as-usual instruction. This was not the case for foundational reading skill measures, as publication year did not significantly predict these outcomes for these students. Although we expected study quality to increase in more recent studies, this was not the case. Overall low study quality across time in this corpus

may have prevented detecting a relationship between year of publication and foundational skill outcomes. On the other hand, publication year significantly predicted ES for reading comprehension outcomes, with older studies reporting larger effects; this finding aligns with the findings from Scammacca et al. (2013). These findings need to be interpreted in light of the overall low quality of studies in this corpus. We echo Ritchey and Goeke's (2006) recommendation: We simply need more high-quality, rigorous research with larger samples of students with or at risk for WLRD to fully understand the effects of OG interventions on the reading outcomes for this population.

Limitations

Several limitations are worth noting. First, we expected to identify more studies that met our inclusion criteria, but these findings were based on only 24 studies. We replicated the 2-year hand search procedures used in Ritchey and Goeke (2006), which did not include international and American Speech-Language-Hearing Association journals; however, it is important to note these journals were included in the electronic database search. Second, the overall study quality of the corpus was low, limiting confidence in the findings and potentially limiting our ability to detect a relationship between study quality and the effects of OG interventions. With a more heterogenous representation of study quality across studies, it is possible that a relationship between study quality and intervention effects may well exist. Third, the ES for foundation skills 0.22 was not statistically significant in part due to the wide range in the magnitude of the ESs across studies. In addition, the small number of students per condition in most studies resulted in large standard errors, leading to a wide confidence interval for the mean ES. Fourth, because multiple measures were used in nearly all studies, RVE needed to be used in estimating the mean ES and its standard error; the RVE tends to result in larger standard errors when there is a smaller number of studies included (<40) in the analysis. Given the mean ES of 0.22 it is worth considering whether or not the findings would be similar across a corpus of studies with higher study quality, particularly because higher-quality studies are often associated with smaller ESs. Finally, we were limited in the moderator analyses we could conduct due to the small number of studies and the limited descriptions of interventions provided in the corpus. With more studies and more detailed descriptions of interventions, additional moderator analyses could have investigated how variables such as grade level or dosage moderated the effects of OG interventions.

Implications for Future Research

The findings from this meta-analysis raise concerns about legislation mandating OG. The findings from this synthesis suggest "promise" but not confidence or evidence-based effects given the research findings currently available. Future intervention studies that utilize high-quality research designs, have sufficiently large samples, and report multiple dimensions of treatment fidelity will determine whether OG interventions positively impact the reading outcomes for students with or at risk for WLRD. First, high-quality, rigorous research needs to examine the effects of OG compared with typical school instruction. Many studies in the corpus did not provide a sufficient description of business-as-usual instruction, which limited our ability to determine the extent to which phonics was taught explicitly in the comparison condition. It is important for researchers to report the nature

of instruction provided in the comparison condition, particularly with regard to explicit phonics instruction. These types of studies will determine whether OG interventions lead to improved reading outcomes for students with or at risk for WLRD compared with typical practice. Next, rigorous research might also compare the effects of OG interventions to non-OG programs that share many of the same characteristics of OG interventions (i.e., systematic, explicit, sequential phonics instruction). It appears that multisensory instruction may be the defining feature that sets OG interventions apart from other programs providing direct instruction in reading and spelling, but there is still a lack of clarity about how OG interventions differ from non-OG interventions that provide direct instruction in decoding and encoding. We did not include multiple treatment studies comparing OG with other reading intervention programs (Acalin, 1995; Foorman et al., 1997; Moore, 1998; Torgesen et al., 2001); however, these types of studies might help determine whether OG intervention is differentially better for students with and at risk for WLRD when compared with explicit phonics programs with less emphasis on multisensory instruction. Finally, it would be important to examine the effects of OG for students with or at risk for WLRD at various grade levels to determine for whom and under what conditions these programs are or are not effective.

Implications for Practitioners, Parents, and Policy Makers

Recently, practitioners, parents, and policy makers have adopted the term “science of reading” to describe a national movement that advocates for reading instruction that aligns with extensive scientific research conducted over several decades and disciplines. Unfortunately, despite this extensive research base, many teachers are uninformed about effective early reading intervention (Spear-Swearling, 2007). Consequently, individuals with WLRD and their families have been significantly challenged in regard to receiving evidence-based instruction that is profitable. These challenges have resulted in families sensing that schools and educators have given up on their children. As a result, they have reached out to groups they perceive as more responsive to their needs and have formed advocacy groups that are actively involved in advocating and securing dyslexia-specific legislation aimed at improving the outcomes for students with and at risk for WLRD. Often, it appears that these parent-led advocacy groups pushed legislation (see Table 1) to provide the practices they felt were most helpful for their children, hoping that these practices would result in positive outcomes. However, we are still at the beginning stages of documenting what evidence is effective for students with WLRD, such as dyslexia. The findings from this meta-analysis do not definitively prove that OG interventions are not impactful for students with dyslexia. In addition, we are not suggesting that other reading programs are more effective than OG. Instead, findings from this meta-analysis indicate that we do not yet know the answers to these questions. Current evidence suggests promise but not confidence that this approach significantly impacts reading outcomes for this population; furthermore, current evidence does not suggest confidence that this is the only approach to remediating word-reading difficulties for these students. It is our hope that this meta-analysis can serve as an impetus for future research and provide evidence-based guidance to practitioners, parents, and policy makers regarding instruction for this population of students.

Finally, many practitioners, parents, and policy makers value the multisensory component of OG instruction (International Dyslexia Association, 2020a, February 11). The majority of states have legislation mandating the use of multisensory reading interventions for students with WLRD. It is possible that many OG interventions are used with students with WLRD because they are marketed as providing that multisensory instruction required in state dyslexia legislation. In addition, it is possible that OG interventions continue to be used in practice, despite the limited evidence supporting their effectiveness, because there remains a prevailing myth that individuals with dyslexia require specialized, multisensory instruction that is inherently different than the instruction required by other students experiencing WLRD (Thorwarth, 2014).

We argue that there are two reasons to question promoting multisensory instruction as a necessary component of reading intervention for students with WLRD. There is little consensus in the field around how we define and operationalize multisensory reading instruction. There is no universal definition of this type of instruction beyond the simultaneous use of visual, auditory, and kinesthetic or tactile learning experiences during reading and spelling instruction. One concern with identifying the multisensory component as the crucial ingredient in OG instruction is that there is not a clear understanding of what multisensory instruction includes across OG programs, how it is applied, and the proportion of instruction it occupies. Effective literacy instruction, in general, involves all of a reader's senses—visual and auditory experiences seeing and reading words aloud and kinesthetic or tactile experiences spelling and writing words. In fact, substantial evidence supports the integration of phonics and spelling instruction to improve students' word reading (e.g., Graham & Santangelo, 2014), which would lead many to believe that most early reading programs offer multisensory instruction. Current research does not indicate that the simultaneous use of these senses positively impacts students reading outcomes, but additional research is needed to understand what this type of instruction looks like in OG interventions and whether this type of instruction has added benefit for students with and at risk for WLRD.

Conclusion

In summary, the findings from this meta-analysis do not provide definitive evidence that OG interventions significantly improve the reading outcomes of students with or at risk for WLRD, such as dyslexia. However, the mean ES of 0.22 indicates OG interventions may hold promise for positively impacting the reading outcomes of this population of students. Additional high-quality research is needed to identify whether OG interventions are or are not effective for students with and at risk for WLRD. Because OG interventions are firmly entrenched in policy and practice with limited evidence supporting their use, we hope that this meta-analysis propels researchers to conduct additional high-quality research to provide the evidence necessary to inform policies and practices for students with WLRD.

Acknowledgments

This research was supported in part by the 5P50 HD052117-12 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development and Grant H325H140001 from the Office of Special Education Programs, U.S Department of Education. The content is solely the responsibility of the authors and does not

necessarily represent the official views of the Eunice Kennedy Shriver National Institute of Child Health and Human Development, the National Institutes of Health, or the U.S. Department of Education. We thank Dr. Jack Fletcher for his feedback and guidance on this manuscript.

References

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

Acalin TA (1995). A comparison of Reading Recovery to Project READ (Publication No. 1361908) [Master's thesis, California State University–Fullerton]. ProQuest Dissertations and Theses Global.

Austin CR, Wanzek J, Scammacca NK, Vaughn S, Gesel SA, Donegan RE, & Engelmann ML (2019). The relationship between study quality and the effects of supplemental reading interventions: A meta-analysis. Exceptional Children, 85(3), 347–366. 10.1177/0014402918796164 [PubMed: 31588147]

*. Bisplinghoff SE (2015). The effectiveness of an Orton-Gillingham-Stillman-influenced approach to reading intervention for low achieving first-grade students (Publication No. 3718126) [Master's thesis, Florida Gulf Coast University]. ProQuest Dissertations and Theses Global.

Borenstein M, Hedges LV, Higgins JPT, & Rothstein HR (2014). Comprehensive meta-analysis (Version 3.3.070) [Computer software]. Biostat. https://www.meta-analysis.com/

*. Christodoulou JA, Cyr A, Murtagh J, Chang P, Lin J, Guarino AJ, Hook P, & Gabrieli JD (2017). Impact of intensive summer reading intervention for children with reading disabilities and difficulties in early elementary school. Journal of Learning Disabilities, 50(2), 115–127. 10.1177/0022219415617163 [PubMed: 26712799]

Council for Exceptional Children. (2015). What every special educator must know: Professional ethics and standards (7th ed.).

DePaoli JL, Fox JH, Ingram ES, Maushard M, Bridgeland JM, & Balfanz R (2015). Building a grad nation: Progress and challenge in ending the high school dropout epidemic. Civic Enterprises. http://gradnation.americaspromise.org/sites/default/files/d8/18006_CE_BGN_Full_vFNL.pdf

*. Dooley B (1994). Multisensorially integrating reading and composition: Effects on achievement of remedial readers in middle school (Publication No. 9428328) [Doctoral dissertation, Texas Woman's University]. ProQuest Dissertations and Theses Global.

Duval S, & Tweedie R (2000). Trim and fill: A simple funnel-plot–based method of testing and adjusting for publication bias in meta-analysis. Biometrics, 56(2), 455–463. 10.1111/j.0006-341X.2000.00455.x [PubMed: 10877304]

Egger M, Davey Smith G, Schneider M, & Minder C (1997). Bias in meta-analysis detected by a simple, graphical test. British Medical Journal, 315(7109), 629–634. 10.1136/bmj.315.7109.629 [PubMed: 9310563]

Eunice Kennedy Shriver National Institute of Child Health and Human Development. (n.d.). What are reading disorders? Retrieved from https://www.nichd.nih.gov/health/topics/reading/conditioninfo/pages/disorders.aspx

Every Student Succeeds Act of 2015, Pub. L. No. 114–95 § 114 Stat. 1177 (2015–2016).

Fisher Z, & Tipton E (2015). Robumeta: An R-package for robust variance estimation in meta-analysis [Computer software]. http://cran.r-project.org/web/packages/robumeta//vignettes/robumetaVignette.pdf

Foorman BR, Francis DJ, Winikates D, Mehta P, Schatschneider C, & Fletcher JM (1997). Early interventions for children with reading disabilities. Scientific Studies of Reading, 1(3), 255–276. 10.1207/s1532799xssr0103_5

*. Fritts JL (2016). Direct instruction and Orton-Gillingham reading methodologies: Effectiveness of increasing reading achievement of elementary school students with learning disabilities (Publication No. 10168236) [Master's thesis, Northeastern University]. ProQuest Dissertations and Theses Global.

Gersten R, Fuchs LS, Compton D, Coyne M, Greenwood C, & Innocenti MS (2005). Quality indicators for group experimental and quasi-experimental research in special education. Exceptional Children, 71(2), 149–164. 10.1177/001440290507100202

*. Giess S (2005). Effectiveness of a multisensory, Orton-Gillingham-influenced approach to reading intervention for high school students with reading disability (Publication No. 3177972) [Master's thesis, University of Florida]. ProQuest Dissertations and Theses Global.

Graham S, & Santangelo T (2014). Does spelling instruction make students better spelling, readers, and writers? A meta-analytic review. Reading and Writing, 27, 1703–1743. 10.1007/s11145-014-9517-0

*. Gunn BK (1996). An investigation of three approaches to teaching phonological awareness to first-grade students and the effects on word recognition (Publication No. 9706736) [Doctoral dissertation, University of Oregon]. ProQuest Dissertations and Theses Global.

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

Hedges LV, Tipton E, & Johnson MC (2010). Robust variance estimation in meta-regression with dependent effect size estimates. Research Synthesis Methods, 1(1), 39–65. 10.1002/jrsm.5 [PubMed: 26056092]

Higgins JPT, Thompson SG, Deeks JJ, & Altman DG (2003). Measuring inconsistency in meta-analyses. BMJ (Clinical Research Ed.), 327(7414), 557–560. 10.1136/bmj.327.7414.557

*. Hook PE, Macaruso P, & Jones S (2001). Efficacy of Fast ForWord training on facilitating acquisition of reading skills by children with reading difficulties: A longitudinal study. Annals of Dyslexia, 51(1), 73–96. 10.1007/s11881-001-0006-1

Individuals With Disabilities Education Improvement Act, 20 U.S.C. §§ 1400 *et seq*. (2004).

Institute for Multi-Sensory Education. (2020a). Orton-Gillingham. https://www.orton-gillingham.com/about-us/orton-gillingham/

Institute for Multi-Sensory Education. (2020b). Three components of multi-sensory instruction. https://journal.imse.com/three-components-of-multi-sensory-instruction/

International Dyslexia Association. (2002). Definition of dyslexia. https://dyslexiaida.org/definition-of-dyslexia/

International Dyslexia Association. (2020a). Multisensory structured language teaching fact sheet. https://dyslexiaida.org/multisensory-structured-language-teaching-fact-sheet/

International Dyslexia Association. (2020b). Structured literacy: Effective instruction for students with dyslexia and related reading difficulties. https://dyslexiaida.org/structured-literacy-effective-instruction-for-students-with-dyslexia-and-related-reading-difficulties/

*. Kutrumbos BM (1993). The effect of phonemic training on unskilled readers: A school-based study (Publication No. 9333368) [Master's thesis, University of Denver]. ProQuest Dissertations and Theses Global.

*. Kuveke SH (1996). Effecting instructional change: A collaborative approach. https://files.eric.ed.gov/fulltext/ED392029.pdf

*. Laub CM (1997). Effectiveness of Project Read on word attack skills and comprehension for third and fourth grade students with learning disabilities (Publication No. 1386289) [Master's thesis, California State University–Fresno]. ProQuest Dissertations and Theses Global.

Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JP, Clarke M, Devereaux PJ, Kleijnen J, & Moher D (2009). The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. PLOS Medicine, 6(7), e1000100. 10.1371/journal.pmed.1000100 [PubMed: 19621070]

Lim L, & Oei AC (2015). Reading and spelling gains following one year of Orton- Gillingham intervention in Singaporean students with dyslexia. British Journal of Special Education, 42(4), 374–389. 10.1111/1467-8578.12104

*. Litcher JH, & Roberge LP (1979). First grade intervention for reading achievement of high risk children. Bulletin of the Orton Society, 29, 238–244. 10.1007/bf02653745

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

Moore AE (1998). The effects of teaching reading to learning disabled students using an explicit phonics program combined with the Carbo Recorded Book Method (Publication No. 1391450) [Doctoral dissertation, Texas Woman's University]. ProQuest Dissertations and Theses Global.

National Center for Education Statistics. (2017). NAEP: The nation's report cards. An overview of NAEL.

National Center for Education Statistics. (2019). NAEP: The nation's report cards. An overview of NAEL.

National Center on Improving Literacy. (2021). State of dyslexia: Explore dyslexia legislation and related initiatives in the United States of America. https://improvingliteracy.org/state-of-dyslexia

Nelson RJ, Benner GJ, & Gonzalez J (2003). Learner characteristics that influence the treatment effectiveness of early literacy interventions: A meta-analytic review. Learning Disabilities Research & Practice, 18(4), 255–267. 10.1111/1540-5826.00080

*. Oakland T, Black JL, Stanford G, Nussbaum NL, & Balise RR (1998). An evaluation of the dyslexia training program: A multisensory method for promoting reading in students with reading disabilities. Journal of Learning Disabilities, 31(2), 140–147. 10.1177/002221949803100204 [PubMed: 9529784]

Odegard TN, Farris EA, Middleton AE, Oslund E, & Rimrodt-Frierson S (2020). Characteristics of students identified with dyslexia within the context of state legislation. Journal of Learning Disabilities, 53(5), 366–379. 10.1177/0022219420914551 [PubMed: 32338146]

Orton-Gillingham Academy. (2020). What is the Orton-Gillingham approach? https://www.ortonacademy.org/resources/what-is-the-orton-gillingham-approach/

*. Rauch ALI (2017). An analysis of two dyslexia interventions (Publication No. 10288068) [Doctoral dissertation, Texas Woman's University]. ProQuest Dissertations and Theses Global.

*. Reed MS (2013). A comparison of computer based and multisensory interventions on at-risk students' reading performance [Master's thesis, Indiana University of Pennsylvania].

Rehabilitation Act of 1973, 29 U.S.C. § 701 *et seq.* (1973).

*. Reuter HB (2006). Phonological awareness instruction for middle school students with disabilities: A scripted multisensory intervention (Publication No. 3251867) [Master's thesis, University of Oregon]. ProQuest Dissertations and Theses Global.

Ring J, Avrit K, Black J, Ring JJ, Avrit KJ, & Black JL (2017). Take Flight: The evolution of an Orton Gillingham-based curriculum. Annals of Dyslexia, 67(3), 383–400. 10.1007/s11881-017-0151-9 [PubMed: 29134479]

Ritchey KD, & Goeke JL (2006). Orton-Gillingham and Orton-Gillingham-based reading instruction: A review of the literature. The Journal of Special Education, 40(3), 171–183. 10.1177/00224669060400030501

Sayeski KL, Earle GA, Davis R, & Calamari J (2019). Orton Gillingham: Who, what, and how. TEACHING Exceptional Children, 51(3), 240–249. 10.1177/0040059918816996

Scammacca N, Roberts G, Vaughn S, & Stuebing K (2013). A meta-analysis of interventions for struggling readers in grades 4–12: 1980–2011. Journal of Learning Disabilities, 48(4), 369–390. 10.1177/0022219413504995 [PubMed: 24092916]

*. Simpson SB, Swanson JM, & Kunkel K (1992). The impact of an intensive multisensory reading program on a population of learning-disabled delinquents. Annals of Dyslexia, 42(1), 54–66. 10.1007/BF02654938 [PubMed: 24233868]

Spear-Swerling L (2007). The research-practice divide in beginning reading. Theory Into Practice, 46(4), 301–308. 10.1080/00405840701593881

Stevens EA, Park S, & Vaughn S (2018). A review of summarizing and main idea interventions for struggling readers in Grades 3 through 12: 1978–2016. Remedial and Special Education, 40(3), 131–149. 10.1177/0741932517749940

*. Stewart ED (2011). The impact of systematic multisensory phonics instructional design on the decoding skills of struggling readers (Publication No. 3443911) [Master's thesis, Walden University). ProQuest Dissertations and Theses Global.

Thorwarth C (2014). Debunking the myths of Dyslexia. Leadership and Research in Education, 1, 51–66.

Tipton E (2015). Small sample adjustments for robust variance estimation with meta-regression. Psychological Methods, 20(3), 375–393. 10.1037/met0000011 [PubMed: 24773356]

Tipton E, & Pustejovsky JE (2015). Small-sample adjustments for tests of moderators and model fit using robust variance estimation in meta-regression. Journal of Educational and Behavioral Statistics, 40(6), 604–634. 10.3102/1076998615606099

Torgesen JK, Alexander AW, Wagner RK, Rashotte CA, Voeller K, & Conway T (2001). Intensive remedial instruction for children with severe reading disabilities: Immediate and long-term outcomes from two instructional approaches. Journal of Learning Disabilities, 34(1), 33–58. 10.1177/002221940103400104 [PubMed: 15497271]

*. Torgesen J, Schirm A, Castner L, Vartivarian S, Mansfield W, Myers D, Stancavage F, Durno D, Javorsky R, & Haan C (2007). National assessment of Title I: Final report. Volume II. Closing the reading gap: Findings from a randomized trial of four reading interventions for striving readers (NCEE 2008–4013). National Center for Education Evaluation and Regional Assistance. https://ies.ed.gov/ncee/pdf/20084013.pdf

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

Tran L, Sanchez T, Arellano B, & Swanson LH (2011). A meta-analysis of the RTI literature for children at risk for read ing disabilities. Journal of Learning Disabilities, 44(3), 283–295. 10.1177/0022219410378447 [PubMed: 21521870]

Uhry JK, & Clark DB (2005). Dyslexia: Theory and practice of instruction (3rd ed.). Pro-Ed.

U.S. Department of Education, Office of Special Education Programs. (2019). Individuals With Disabilities Education Act (IDEA) database. https://www2.ed.gov/programs/osepidea/618-data/state-level-data-files/index.html#bcc

Valentine JC, & Cooper H (2008). What Works Clearinghouse study design and implementation assessment device (Version 1.0). U.S. Department of Education.

Vaughn S, Denton C, & Fletcher J (2010). Why intensive interventions are necessary for students with severe reading difficulties. Psychology in the Schools, 47(5), 432–444. 10.1002/pits.20481 [PubMed: 21072127]

Vaughn S, Elbaum BE, Wanzek J, Scammacca N, & Walker MA (2014). Code sheet and guide for education related intervention study syntheses. Meadows Center for Preventing Educational Risk.

*. Wade JJ (1993). Project Read versus basal reading program with respect to reading achievement, attitude toward school and self-concept (Publication No. 9417849) [Master's thesis, The University of Southern Mississippi]. ProQuest Dissertations and Theses Global.

*. Wanzek J, & Roberts G (2012). Reading interventions with varying instructional emphases for fourth graders with reading difficulties. Learning Disability Quarterly, 35(2), 90–101. 10.1177/0731948711434047

*. Westrich-Bond A (1993). The effect of direct instruction of a synthetic sequential phonics program on the decoding abilities of elementary school learning disabled students (Publication No. 9415253) [Doctoral dissertation, Rutgers University]. ProQuest Dissertations and Theses Global.

What Works Clearinghouse. (2010a). Alphabetic Phonics. U.S. Department of Education, Institute of Education Sciences. https://ies.ed.gov/ncee/wwc/Docs/InterventionReports/wwc_alpha_phonics_070110.pdf

What Works Clearinghouse. (2010b). Barton Reading & Spelling System. U.S. Department of Education, Institute of Education Sciences. https://ies.ed.gov/ncee/wwc/Docs/InterventionReports/wwc_barton_070110.pdf

What Works Clearinghouse. (2010c). Dyslexia Training Program. U.S. Department of Education, Institute of Education Sciences. https://ies.ed.gov/ncee/wwc/Docs/InterventionReports/wwc_dyslexia_070110.pdf

What Works Clearinghouse. (2010d). Fundations. U.S. Department of Education, Institute of Education Sciences. https://ies.ed.gov/ncee/wwc/Docs/InterventionReports/wwc_fundations_070110.pdf

What Works Clearinghouse. (2010e). Herman Method. U.S. Department of Education, Institute of Education Sciences. https://ies.ed.gov/ncee/wwc/Docs/InterventionReports/wwc_herman_070110.pdf

What Works Clearinghouse. (2013). LANGUAGE! U.S. Department of Education, Institute of Education Sciences. https://ies.ed.gov/ncee/wwc/Docs/InterventionReports/wwc_language_021213.pdf

What Works Clearinghouse. (2010f). Lindamood Phoneme Sequencing (LiPS). U.S. Department of Education, Institute of Education Sciences. https://ies.ed.gov/ncee/wwc/Docs/InterventionReports/wwc_lindamood_031610.pdf

What Works Clearinghouse. (2010g). Orton-Gillingham-based strategies (unbranded). U.S. Department of Education, Institute of Education Sciences. https://ies.ed.gov/ncee/wwc/Docs/InterventionReports/wwc_orton-gill_070110.pdf

What Works Clearinghouse. (2010h). Project Read Phonology. U.S. Department of Education, Institute of Education Sciences. https://ies.ed.gov/ncee/wwc/Docs/InterventionReports/wwc_project_read_070110.pdf

What Works Clearinghouse. (2010i). Wilson Reading System. U.S. Department of Education, Institute of Education Sciences. https://ies.ed.gov/ncee/wwc/Docs/InterventionReports/wwc_wilson_070110.pdf

What Works Clearinghouse. (2012). The Spalding Method. U.S. Department of Education, Institute of Education Sciences. https://ies.ed.gov/ncee/wwc/Docs/InterventionReports/wwc_spalding_method_101612.pdf

*. Wille GA (1993). Project Read as an early intervention program (Publication No. 1353782) [Master's thesis, California State University–Fullerton]. ProQuest Dissertations and Theses Global.

*. Young CA (2001). Comparing effects of tracing to writing when combined with Orton-Gillingham methods on spelling achievement among high school students with reading disabilities (Publication No. 3064692) [Doctoral dissertation, University of Texas at Austin]. ProQuest Dissertations and Theses Global.

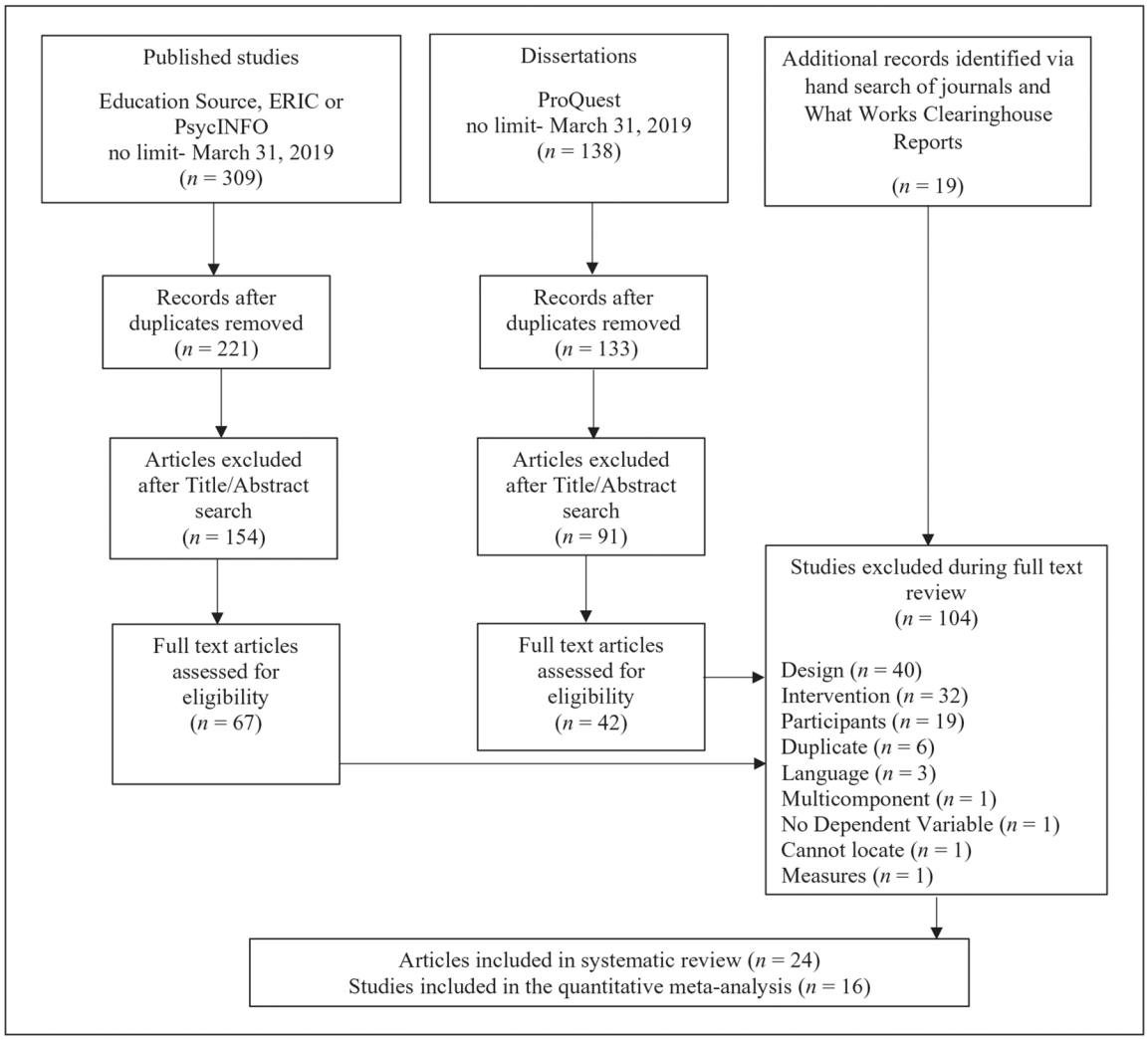

Figure 1.
PRISMA diagram.

Table 1.

State Legislation Mandating the Use of Orton-Gillingham (O-G).

State	Year	Legislation	Description of legislation
Arkansas	2019	Senate Bill 153	Mandated the DOE to create an approved list of materials, resources, and curriculum programs supported by the science of reading and based on instruction that is explicit, systematic, cumulative, and diagnostic, including dyslexia programs that are evidence based and grounded in the OG methodology
Minnesota	2019	Senate File Number 733	Permitted a district to use staff development funds for teachers to take courses from accredited providers, including providers accredited by the International Multisensory Structured Language Education Council and the Academy of OG Practitioners and Educators
Mississippi	2019	House Bill 1046; Senate Bill 2029; House Bell 496	Defined dyslexia therapy as “a program delivered by a licensed dyslexia therapist that is scientific, research-based, OG based, and offered in a small-group setting”; defined a dyslexia therapist as “a professional who has completed training in a department approved OG based dyslexia therapy training program”; required that each district employ a dyslexia coordinator trained in OG based dyslexia therapy
Missouri		House Bill 2379	Required the use of evidence-based reading instruction, with consideration of the National Reading Panel report and OG methodology principles; mandated that a task force be created including members with training and experience in early literacy development and effective research-based intervention techniques, including OG remediation programs
North Dakota		House Bill 1461	Required that district dyslexia specialists be trained in OG
Rhode Island	2019	House Bill 5426; House Bill 7968	Mandated (a) at-risk students to immediately receive intervention using an OG intervention provided by an individual who possesses a Level 1 certification in OG, (b) all teachers to receive professional development provided by the OG Practitioners and Educator, (c) districts to develop and publish reading support resource guides utilizing advice of the Academy of OG Practitioners and Educators, (d) the General Assembly to provide $50,000 annually for teacher training in the OG Classroom Educator Program, (e) a position at the DOE to include a reading specialist certified in OG, and (f) state universities to require preservice teachers to complete OG classroom educator programs
Wisconsin	2019	Assembly Bill 50; Assembly Bill 595; Senate Bill 555	Mandated that the state superintendent employ a dyslexia specialist certified by the Academy of OG Practitioners and Educators; provided grants to teachers who earn dyslexia-related certifications from the Academy of OG Practitioners and Educators

Note. DOE = Department of Education.

Table 2.

Experimental Study Information.

Study	Design	*n*	Grade	Risk type	Description of conditions	Total sessions	Total hours
Bisplinghoff (2015)	Quasiexperimental	21	1	SR	T: Barton (explicit instruction of phonemic awareness and phonics with practice opportunities for decoding and spelling) C: Houghton Mifflin Reading Tool Kit (phonemic and phonological awareness, phonics and decoding, and oral reading fluency)	72	36
Christodoulou et al. (2017)	Experimental	47	*M* = 1.4	RD, SR	T: Lindamood-Bell Seeing Stars (multisensory teaching of phonological and orthographic awareness, sight word recognition, and comprehension) C: No intervention	30	100–120
Dooley (1994) [a]	Experimental	151	7	SR	T: Multisensory Integrated Reading and Composition (adapted Alphabetic Phonics) C: Traditional basal approach	85–90	71–75
Fritts (2016)	Experimental	86	1–4	Dyslexia, RD	T1: Corrective Reading (word attack, group reading, workbook exercises) T2: Wilson Fundations or Wilson Reading C: Teacher-selected curriculum	50	33.3
Giess (2005) [b]	Quasiexperimental	18	10, 11	LD, SR, OHI	T: Barton Reading and Spelling System (Orton-Gillingham program) C: NR	NR	NR
Gunn (1996)	Experimental	34	1	AR	T1: Complete Auditory in Depth Discrimination (phonological awareness instruction and spelling and reading practice) T2: Modified Auditory in Depth Discrimination (phonological awareness instruction only) C: Basal instruction	40	20
Hook et al. (2001) [a,b]	Quasiexperimental	20	Ages 7–12	SR	T1: Orton-Gillingham T2: Fast ForWord (computer-based phonemic and phonics instruction) C: Business as usual	25	25
Kutrumbos (1993)	Quasiexperimental	40	Ages 9–14	Dyslexia, RD	T: Lindamood Auditory Discrimination In-Depth Program and Orton-Gillingham C: Remedial reading curriculum	48–60	36–45
Kuveke (1996) [b]	Quasiexperimental	12	Year 1: 2, 3 Year 2: 3,4	AR, SR	T: Alphabetic Phonics C: Business as usual	NR	NR
Laub (1997)	Quasiexperimental	48	3, 4	LD	T: Project Read (direct, multisensory, systematic, sequential) C: NR	80	66.67
Litcher and Roberge (1979) [a]	Quasiexperimental	40	1	AR	T: Orton-Gillingham C: Business as usual	NR (1 year in duration)	NR
Oakland et al. (1998) [a,b]	Quasiexperimental	48	*M* = 4.3	Dyslexia	T: Dyslexia Training Program (adaptation of Alphabetic Phonics) C: Typical school-provided reading instruction	350	350
Rauch (2017)	Quasiexperimental	72	2–5	School-identified dyslexic tendencies	T: Take Flight (multisensory phonemic awareness, spelling practice, comprehension strategies, fluency instruction) C: Rite Flight (district-developed intervention focusing on fluency, comprehension, and phonemic awareness; Fountas and Pinnell's Leveled Literacy Intervention)	80–472 (*Mdn* = 240)	80–472 (*Mdn* = 110)

Study	Design	*n*	Grade	Risk type	Description of conditions	Total sessions	Total hours
Reed (2013)	Quasiexperimental	87	1–3	SR	T1: Sonday (adaptation of Orton-Gillingham) T2: Fast ForWord (advances from p re reading skills to phonics, decoding, spelling, vocabulary, and comprehension) T3: Both Sonday and Fast ForWord C: No intervention	NR	NR
Reuter (2006)	Experimental	26	6–8	RD	T: Wilson Reading C: Typical reading instruction	70	52.5
Simpson et al. (1992) [a,b]	Quasiexperimental	63	Ages 13–18	LD	T: Orton-Gillingham C: Typically provided instruction	NR	*M* = 51.9
Stewart (2011)	Quasiexperimental	51	1	AR	T: Orton-Gillingham C: Trophies Program (traditional basal phonics instruction)	60	45
Torgesen (2007)	Experimental	335	3	SR	Tl: Failure Free Reading (combination of computer-based lessons, workbook exercises, and teacher-led instruction to teach sight words, fluency, and comprehension) T2: SpellRead Phonological Auditory Training (systematic and explicit instruction in phonics and phonemic awareness) T3: Wilson Reading T4: Corrective Reading (explicit and systematic scripted instruction aimed to improve word identification and fluency) C: Typical school instruction (delivered mostly in individualized and small-group settings)	NR	90
Torgesen (2007)	Experimental	407	5	SR	Tl: Failure Free Reading (combination of computer-based lessons, workbook exercises, and teacher-led instruction to teach sight words, fluency, and comprehension) T2: SpellRead Phonological Auditory Training (systematic and explicit instruction in phonics and phonemic awareness) T3: Wilson Reading T4: Corrective Reading (explicit and systematic scripted instruction aimed to improve word identification and fluency) C: Typical school instruction (delivered mostly in individualized and small-group settings)	NR	90
Torgesen et al. (1999)	Experimental	180	K	AR	Tl: Lindamood Bell Phonological Awareness plus synthetic phonics T2: Embedded phonics T3: Regular classroom instruction C: No treatment control	270	90
Wade (1993) [b]	Quasiexperimental	36	1, 2	AR	T: Project Read (direct, multisensory, systematic, sequential) C: Traditional basal reading	NR	NR
Wanzek and Roberts (2012)	Experimental	44	4	LD, RD, SR	T: Wilson Reading C: Business-as-usual school-provided intervention	85–114	42–57
Westrich-Bond (1993) [a]	Quasiexperimental	39	1–5	LD	T: Orton-Gillingham sequential synthetic phonics C: Ginn reading program (remedial reading instruction using basal)	NR (range 1–35 months)	NR
Wille (1993) [b]	Quasiexperimental	10	1	SR	T: Project Read (direct, multisensory, systematic, sequential) C: Typical classroom reading instruction	40	33.33

Study	Design	*n*	Grade	Risk type	Description of conditions	Total sessions	Total hours
Young (2001) [b]	Experimental	20	9–12	RD	Tl: Sight word instruction with tracing and Orton-Gillingham T2: Sight word instruction with writing and Orton-Gillingham C: Typically provided instruction	28	11.7

Note. Due to the inconsistent application of the term "dyslexia" and identification of students with dyslexia across the literature, we included studies with participants formally diagnosed with dyslexia and those without a diagnosis but who exhibited word-level reading difficulties (i.e., students at risk for dyslexia, students with a learning disability in reading, or struggling readers performing in the bottom quartile on a standardized reading measure). AR = at risk; C = comparison; LD = learning disability; NR = not reported; OHI = other health impaired; RD = reading disability; SR = struggling reader; T = treatment.

[a] Study included in Ritchey and Goeke (2006).

[b] Study not included in the meta-analysis.

Table 3.

Study Quality.

Study	Design	Comparison group	Implementation fidelity	*M* rating
Bisplinghoff (2015)	⦿	❍	•	1.00
Christodoulou et al. (2017)	❍	•	•	0.67
Dooley (1994) [a]	⦿	❍	•	1.00
Fritts (2016)	⦿	❍	⦿	1.33
Giess (2005) [b]	•	•	⦿	0.33
Gunn (1996)	⦿	❍	•	1.00
Hook et al. (2001) [a,b]	•	❍	•	0.33
Kutrumbros (1993)	⦿	❍	•	1.00
Kuveke (1996) [b]	•	•	•	0.00
Laub (1997)	⦿	•	•	0.33
Litcher and Roberge (1979) [a]	⦿	•	•	0.33
Oakland et al. (1998) [a,b]	⦿	•	•	0.33
Rauch (2017)	⦿	❍	•	1.00
Reed (2013)	⦿	•	•	0.33
Reuter (2006)	❍	⦿	•	1.00
Simpson et al. (1992) [a,b]	⦿	•	•	0.33
Stewart (2011)	⦿	❍	•	1.00
Torgesen et al. (1999)				
OG vs. no intervention	❍	•	•	0.67
OG vs. regular classroom support	❍	❍	•	1.33
Torgesen et al. (2007)	❍	❍	❍	2.00
Wade (1993) [b]	•	•	•	0.00
Wanzek and Roberts (2012)	❍	•	⦿	1.00
Westrich-Bond (1993) [a]	⦿	❍	•	1.00
Wille (1993) [b]	•	•	•	0.00

Study	Design	Comparison group	Implementation fidelity	*M* rating
Young (2001) [b]	⊙	❍	•	1.00
Average score by indicator	0.95	1.00	0.17	0.76

Note. M rating for each study provided on a scale of 0 to 2. ❍ = exemplary (2); ⊙ = acceptable (1); • = unacceptable (0); OG = Orton-Gillingham; BAU = business as usual.

[a] Study included in Ritchey and Goeke (2006).

[b] Study not included in the meta-analysis.