Providing Effective Reading Instruction to Elementary Students with or at Risk for Reading Difficulties

---

A Dissertation Presented to
the Faculty of the School of Education
and Human Development
University of Virginia

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

---

by
Katlynn Dahl-Leonard, M.A.
May 2024

University of Virginia
School of Education and Human Development Registrar
Office of Admissions and Student Affairs

Ehd-registrar@virginia.edu
Ridley Hall 102D
417 Emmet Street
Charlottesville, VA 22903

Dissertation Approval Form

Student Full Name: Dahl-Leonard,Katlynn [redacted]

Department Name: Curriculum, Instruction & Special Education

Degree Program: Education (PhD) Reading Education

Date of Defense: 3/21/2024

This doctoral dissertation has been approved by the Graduate Faculty of the School of Education and Human Development in partial fulfillment for the degree of Doctor of Philosophy.

Approved Title of Doctoral Dissertation:

Providing Effective Reading Instruction to Elementary Students with or at Risk for Reading Difficulties

	Name	Department/University	Signature
Chair	Colby Hall	CISE, EHD, UVA	DocuSigned by: Colby Hall
Co-Chair (if applicable)			
Committee Member	Emily Solari	CISE, EHD, UVA	DocuSigned by: Emily Solari
Committee Member	William Therrien	CISE, EHD, UVA	DocuSigned by: William Therrien
Committee Member	Jamie DeCoster	EDLF, EHD, UVA	DocuSigned by: Jamie DeCoster
Committee Member	Philip Capin	Department of Special Education, College of Education, University of Texas Austin	DocuSigned by: Philip Capin
Committee Member			
Committee Member			
Student	Katlynn Dahl-Leonard	Curriculum, Instruction & Special Education	DocuSigned by: Katlynn Dahl-Leonard

Table of Contents

Examining Fidelity Reporting within Studies of Foundational Reading Interventions for Elementary Students with or at Risk for Dyslexia

Katlynn Dahl-Leonard[1], Colby Hall[1], Philip Capin[2], Emily J. Solari [1], Alisha Demchak [1], and William J. Therrien[1]

[1]University of Virginia

[2]University of Texas at Austin

Katlynn Dahl-Leonard: awr8qt@virginia.edu; https://orcid.org/0000-0001-6124-5990

Colby Hall: csh4t@virginia.edu; https://orcid.org/0000-0002-0779-1322

Philip Capin: pcapin@utexas.edu; https://orcid.org/0000-0001-8305-5549

Emily J. Solari: ejs9ea@virginia.edu; https://orcid.org/0000-0002-8498-9834

Alisha Demchak: and3u@virginia.edu; https://orcid.org/0000-0002-2496-5996

William J. Therrien: therrien@virginia.edu; https://orcid.org/0000-0003-0594-5129

The authors have no conflicts of interest to disclose.

This work was supported by the Harrison Family Foundation. The content is solely the responsibility of the authors and does not necessarily represent the official views of the Harrison Family Foundation.

Correspondence concerning this article should be addressed to Katlynn Dahl-Leonard at awr8qt@virginia.edu.

Abstract

Early access to evidence-based reading intervention improves outcomes for students with or at risk for reading difficulties. Additionally, teacher implementation of reading interventions plays a key role in the efficacy of reading interventions. Previous research suggests the influence of intervention implementation fidelity on student language and literacy outcomes is more significant for lower-performing students and students with disabilities, such as dyslexia. However, recent syntheses have suggested that less than half of reading intervention studies report treatment fidelity data. This meta-analysis examined fidelity reporting within reading intervention studies for students with or at risk for dyslexia in Grades K-5. We aimed to record the frequency and extent of fidelity reporting, explore associations between study or intervention features and fidelity reporting, and compare mean intervention effect sizes for studies reporting fidelity and those that did not. A total of 51 studies were included. Results indicated that 75% of studies reported fidelity data. Studies reporting fidelity primarily focused on adherence and dosage data with little to no information reported for other dimensions of fidelity (i.e., quality, responsiveness, differentiation). Suggestions for improving reporting of treatment fidelity data are discussed.

Keywords: reading, intervention, dyslexia, fidelity, elementary

Examining Fidelity Reporting within Studies of Foundational Reading Interventions for Elementary Students with or at Risk for Dyslexia

Dyslexia is a specific learning disability characterized by difficulties with accurate or fluent word recognition (Fletcher et al., 2019). Estimates of the prevalence of dyslexia vary depending on the cut-point applied to a continuum of severity (Fletcher et al., 2019; Shaywitz et al., 1992). Most prevalence estimates range from 3-7% when applying a cut point of 1.5 standard deviations or more below the mean on measures of reading (Peterson & Pennington, 2012; Snowling & Melby-Lervåg, 2016). Students with significant reading difficulties are likely to demonstrate academic difficulties throughout their school careers and are more likely to drop out of school (Boscardin et al., 2008; Daniel et al., 2006; Francis et al., 1996; Hernandez, 2011). They are less likely to enroll in postsecondary education programs and more likely to be incarcerated (Greenberg et al., 2007; Horn & Berktold, 1999). They are also at greater risk for internalizing symptoms associated with anxiety and depression (Dahle & Knivsberg, 2014; Jordan et al., 2014; Mugnaini et al., 2009) as well as for externalizing behavior disorders (Willcutt et al., 2010).

Research has shown that access to evidence-based core reading instruction and intensive interventions within multi-tiered systems of support can reduce the severity of reading difficulties experienced by students, including students with or at risk for dyslexia (Al Otaiba et al., 2009; Torgesen, 2004; VanDerHeydan et al., 2007). Meta-analyses of intervention research (e.g., Gersten et al., 2009; Scammacca et al., 2007; Swanson et al., 1999; Vaughn et al., 2012) indicate that effective reading interventions for elementary students with reading difficulties typically focus on multiple reading component skills, including phonological awareness (PA), knowledge of grapheme-phoneme correspondences, decoding, encoding, and text reading. These

meta-analyses also indicate that effective instruction for students with reading difficulties is explicit and systematic. Effective instruction for students with reading difficulties includes teacher modeling with explanations and guided practice opportunities that offer frequent opportunities for students to respond and receive feedback, as well as scaffolding that can be gradually reduced to support students' acquisition of new skills and knowledge (Gersten et al., 2009; Vaughn et al., 2012; Swanson et al., 1999).

Treatment Fidelity in Intervention Research

Perhaps as important as intervention components and teaching methods is the degree to which an intervention can be implemented with fidelity. Fidelity is an overarching term encompassing multiple components of implementation (Dane & Schneider, 1998). Dane and Schneider identified five dimensions of fidelity: adherence, quality of delivery, dosage, participant responsiveness, and program differentiation. *Adherence* refers to the extent to which critical components of the intervention are implemented as intended. *Quality* of implementation is measured separately from adherence and describes qualitative aspects of implementation, such as implementer enthusiasm, preparedness, pacing, and behavior management. *Dosage* refers to the amount of instruction provided or how much exposure to an intervention students received. Participant *responsiveness* is the extent to which students responded to the intervention (e.g., level of engagement or independent use of skills taught). Program *differentiation* is the extent to which the intervention varies from another treatment or comparison condition.

Prior research suggests a strong, positive association between fidelity of implementation of evidence-based instruction and academic outcomes in students (Kretlow & Bartholomew, 2010). Moreover, findings indicate fidelity is most strongly associated with reading outcomes for low-performing students (e.g., Capin et al., 2022; Connor et al., 2007; Hamre et al., 2010),

including children with learning disabilities, such as dyslexia (Boardman et al., 2016). For instance, Connor and colleagues (2007) found that third-grade children with low initial reading performance disproportionately benefited from high-fidelity implementation of an explicit instruction intervention when compared to their higher performing peers. This finding—that fidelity really matters for lower-achieving students—has been replicated in language and literacy interventions with preschool children (e.g., Hamre et al., 2010), elementary students (e.g., Neugebauer et al., 2017), and secondary students (e.g., Capin et al., 2022). Consistent with the Connor et al. (2007) finding, Boardman and colleagues (2016) revealed a significant interaction between instructional fidelity and disability status. Higher implementation fidelity was associated with higher reading comprehension scores for students with learning disabilities in the middle grades. Given (a) the important role of fidelity in improving outcomes for students with learning difficulties and (b) the importance of identifying both effective and feasible-to-implement interventions for students with or at risk for dyslexia, the purpose of this study is to understand how fidelity is reported in studies of reading interventions designed to improve reading outcomes for elementary students with or at risk for dyslexia.

Research assessing an intervention's implementation fidelity is vital in showing that the intervention is feasible for teachers to implement. It determines the practicality of scaling up the intervention and generalizing it to other populations and settings (Moncher & Prinz, 1991; Nelson et al., 2012; O'Donnell, 2008). Further, for researchers aiming to evaluate interventions, implementation fidelity is a crucial methodological consideration because it increases the internal validity of a study evaluating an intervention's effects (i.e., when an intervention is implemented with fidelity, it is more possible to infer that the effects reported are indeed due to the intervention being evaluated; Shadish et al., 2002).

Past Research on Treatment Fidelity Reporting

Unfortunately, past syntheses show that treatment fidelity is frequently unreported (e.g., Capin et al., 2018; Gresham et al., 2000; Swanson et al., 2011). In 2000, Gresham et al. reviewed treatment fidelity reporting in academic intervention studies published in select special education journals from 1995 to 1999. They found that only 18% of studies reported treatment fidelity. Swanson et al. (2011) reviewed articles published in select general and special education journals from 2005 to 2009 and found that 47% of studies reported treatment fidelity data. Capin et al. (2018) examined treatment fidelity in K-3 reading intervention research. They also found that less than half (47%) of the reading intervention studies reported treatment fidelity. Therefore, although there was an increase in fidelity reporting between the Gresham et al. (2000) synthesis and the Swanson et al. (2011) synthesis, the proportion of studies reporting treatment fidelity seems to have remained relatively constant given the end date of the Swanson et al. search and the end date of the Capin et al. search.

Additionally, past reviews of treatment fidelity reporting show that studies often do not collect or report data for all dimensions of this construct (e.g., Capin et al., 2018; Swanson et al., 2011). Swanson et al. (2011) examined treatment adherence, quality of implementation, and dosage, but did not report on treatment receipt and differentiation. They revealed that studies often only reported treatment adherence data. Capin et al. (2018) explored the inclusion of all five fidelity dimensions in their synthesis to address this deficit. They discovered that studies reporting treatment fidelity primarily reported treatment adherence, with other dimensions of treatment fidelity (i.e., quality, dosage, receipt, and differentiation) consistently absent from the corpus of reading intervention studies reviewed. Capin and colleagues also explored how fidelity data were reported within studies of K-2 reading interventions. They noted that 86% of studies

that reported fidelity presented numeric treatment fidelity data while the remaining studies presented qualitative descriptions.

Only one previous study analyzed how fidelity reporting varies according to study features. Capin et al. (2018) discovered that publication year, research design, treatment implementer, and participant grade were related to treatment fidelity reporting. In particular, (a) reporting of fidelity data increased over time, (b) single case design studies were more likely than experiments, quasi-experiments, or treatment-comparison designs to report treatment fidelity data, (c) studies with paraprofessional implementers reported treatment fidelity more frequently than studies with teachers, research team members, or other implementers, and (d) studies with participants in Grade 1 reported treatment fidelity more frequently than studies of students in kindergarten, Grade 2, or multiple grade levels.

Study Purpose and Research Questions

The present study was intended to build on this existing knowledge base regarding treatment fidelity reporting. As shown in Table 1, the present review differs from previous reviews in several ways. In particular, our purpose was to provide a systematic review of studies of foundational reading interventions for elementary students with or at risk for dyslexia to determine how often and to what extent treatment fidelity was addressed. We were interested in examining fidelity reporting in research that includes students with or at risk for dyslexia because of prior findings that implementation fidelity is strongly associated with student outcomes for children with learning difficulties (Boardman et al., 2016; Capin et al., 2022; Connor et al., 2004; Hamre et al., 2010). We also aimed to examine whether study features associated with fidelity reporting in past research (Capin et al., 2018) were associated with fidelity reporting in studies that included students with or at risk for dyslexia.

Finally, unlike previous research, we aimed to compare mean intervention effects in studies that reported fidelity data to effects in studies that did not. This question was not concerned with associations between fidelity of intervention implementation and intervention effect size, but rather the presence or absence of fidelity reporting as a marker for study rigor. Past research has reported associations between variables that are proxies for study rigor (e.g., publication year, research design, sample size, or some aggregate "study quality" variable that often includes presence or absence of fidelity reporting as one sub-component) and effect size (Hall et al., 2017; Scammacca et al., 2015; Slavin & Smith, 2009). However, we were unable to identify studies that examined the associations between fidelity reporting alone (i.e., one aspect of study rigor) and intervention effect size.

Therefore, we asked:

1) What proportion of foundational reading intervention studies for K-5 students with or at risk for dyslexia reported treatment fidelity data?
2) For studies that reported fidelity information, what dimensions of fidelity (i.e., adherence, quality, dosage, responsiveness, or differentiation) were reported and how were they reported (i.e., quantitatively or qualitatively)?
3) Was fidelity reporting associated with other study features (i.e., publication year, research design, student grade level, and intervention implementer)?
4) Was there a difference in the average effect size for studies that reported fidelity information relative to those that did not?

Methods

Identification of Studies

To examine the fidelity reporting within studies of foundational reading interventions for

students with or at risk for dyslexia in Grades K-5, we utilized a corpus of studies identified in a recently conducted meta-analysis (Hall et al., 2022). This previous meta-analysis investigated the effects of reading intervention studies conducted within the last 40 years for Grade K-5 students with or at risk for dyslexia. Hall et al. (2022) used systematic procedures for identifying studies for inclusion. They searched peer-reviewed articles and grey literature published in English on or after January 1, 1980 and before January 1, 2021 following Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines (Moher et al., 2009). The initial phases of the search process included a search of electronic databases (i.e., ERIC and PsycINFO) using key population identifiers (*dyslex*, reading difficult*, reading disabilit*, risk for reading difficult*, risk for reading disabilit*, reading delay*, reading disorder*, learning disab**) and terms describing reading interventions (*reading interven*, reading instruction, phon*, correspondence*, reading fluency*). After a round of abstract screening to ascertain eligibility for inclusion, 541 full texts were reviewed and included if they met the following criteria:

1) Published between January 1, 1980 and January 1, 2021.
2) Employed experimental or quasi-experimental research designs with at least 15 participants per group.
3) Included students in Grades K-5 (ages 5-11) or the sample mean age was within this range.
4) Included a screening criterion at or below the 25th percentile or a mean baseline score at or below the 16th percentile on a norm-referenced test of word reading, spelling, or foundational word reading skills. This criterion aligns with contemporary conceptions of dyslexia that refer to performance at the lower part of a continuous distribution of word reading and spelling skills. We operationally defined dyslexia in this way because other approaches to identification based on IQ, cognitive discrepancies, and markers of

unexpectedness of reading difficulties have shown insufficient validity (e.g., Miciak & Fletcher, 2020).

5) Included a school-based intervention that was implemented for more than one session, focused on PA, phonics, word reading, spelling, or a combination of these domains, and was delivered primarily in English.
6) Included at least one calculable effect size on a norm-referenced measure of PA, knowledge of grapheme-phoneme correspondences, pseudoword reading, real word reading, spelling, or passage reading.

Of the 541 full-texts reviewed by Hall and colleagues (2022), 38 met all inclusion criteria. Hall et al. also completed a hand search of the five journals that published the greatest number of included studies: *Exceptional Children, Journal of Educational Psychology, Journal of Learning Disabilities, Journal of Research on Educational Effectiveness, and Learning Disabilities Research & Practice.* Additionally, they conducted an ancestral search of articles included in recent, relevant literature reviews (Galuschka et al., 2020; Gersten et al., 2020; Stevens et al., 2021; Suggate, 2016; Wanzek et al., 2016, 2018). Fourteen additional peer-reviewed articles met inclusion criteria and were coded.

In the present analysis, two additional studies were excluded because they did not report posttest mean and standard deviation estimates by condition. Thus, for the present analysis, 51 studies within 50 publications (36 identified through the database search and 14 identified through the ancestral search) were identified for inclusion. Figure 1 represents the search procedure and results at each stage of the search process.

Coding Procedures

Coding occurred in two phases. In phase one, Hall and colleagues (2022) coded

characteristics of participants, interventions, outcome measures, and research methods. Eight coders participated in a three-hour training and independently coded articles until they obtained a minimum of 90% reliability with a gold standard in each code sheet section. Once coding began, all articles were independently double coded by two members of the author team, and the first author resolved any discrepancies.

In phase two, three members of the author team coded studies for information related to (a) type of fidelity reported (i.e., quantitative or qualitative) and (b) dimensions of fidelity reported (i.e., adherence, quality, dosage, responsiveness, or differentiation). Information about fidelity-related codes is provided in Table 2. Coders participated in a one-hour training and achieved >95% reliability with a gold standard before coding studies. All texts were independently reviewed by two reviewers and disagreements were resolved via discussion and consensus.

Data Analysis

We addressed RQ1 and RQ2 by conducting descriptive analyses. We calculated and reported the percentage of studies employing a particular treatment fidelity procedure (e.g., reporting adherence data) as the ratio of the number of studies that used the procedure to the total number of studies for which the procedure could have been reasonably included. To address RQ3, we examined the associations between a given study feature (i.e., publication year, research design, student grade level, or intervention implementer) coded as a categorical variable (e.g., research design was categorized as either “randomized control trial design” or “quasi-experimental design”) and treatment fidelity reporting using chi-square statistics. Given our intent to replicate and extend the Capin et al. (2018) synthesis, we coded, categorized, and analyzed our data using similar methods (i.e., conducted chi-square tests) to those employed by

Capin and colleagues. More specifically, chi-square analyses were used to examine the relations between fidelity reporting and study features in the Capin et al. study and in our study because it could be applied to *all* of the study features of interest, including ones in which the collected data was naturally categorical (i.e., research design, grade level, and implementer). Notably, analyses assessing the associations between study features and treatment fidelity reporting should be considered exploratory rather than conclusive due to the dependent nature of these data.

RQ4 was addressed by analyzing relevant data using the Comprehensive Meta Analysis software (Borenstein et al., 2005). To quantify the effects of interventions on reading outcomes for students with or at risk for dyslexia in Grades K-5, we used standardized mean differences between intervention and control groups estimated with Hedges' *g* (Hedges, 1981), using reported posttest mean and standard deviation estimates by condition. We used adjusted posttest means (i.e., posttest means adjusted for pretest scores) when they were available. When more than one outcome measure was reported for a study, effect size estimates were aggregated using the mean to avoid overrepresentation of multi-measure studies in the overall analyses (Rosenthal, 1991). Although there are alternative analyses that account for effect size dependency that do not involve aggregating scores to a mean score (e.g., robust variance estimation), the variance across effect sizes when we assume independence (i.e., when the correlation is set to 0.0) is 0.000; when the correlation is set to 1.0 (i.e., when effect size estimates are aggregated using the mean), the variance is 0.002. Due to the very small difference in these two extremes, it is unlikely that setting the correlation to a numerical value between 0.0 to 1.0 (e.g., 0.8) would produce different results from those currently reported in the manuscript. Further, given that all our effect sizes were calculated from norm-referenced outcome measures of related literacy skills (e.g., word reading, spelling), our effect sizes are likely to be highly correlated (i.e., near 1.0). To describe

the heterogeneity between studies, we calculated the restricted maximum likelihood estimate of the between-study variance (τ^2). We also calculated the Q and I^2 statistics to reveal the extent to which heterogeneity among true effect sizes contributes to the observed variation in the effect size estimates.

Results

Fidelity Data Reporting (RQs 1 & 2)

Of the 51 foundational reading intervention studies coded for this analysis, 38 (75%) reported treatment fidelity data. Table 3 addresses RQ1 by summarizing the key features (e.g., sample size, grade, research design) of the studies reviewed, alongside information about fidelity reporting. To address RQ2, we examined the procedures authors used to report fidelity data (see Table 4).

Dimensions of Fidelity Reported

Overall, of the 38 studies that reported treatment fidelity data, 68% reported an indicator of adherence, 32% reported quality, 58% reported dosage, 18% reported responsiveness, 11% reported differentiation, and 29% reported a combined indicator. Most of the 38 studies (k = 30) that reported some information about fidelity reported multiple dimensions of fidelity (e.g., adherence, quality, and dosage reported separately). Only one study reported data for all five dimensions of fidelity (Denton et al., 2010). Eight of the 38 studies only reported data on one dimension of fidelity, with four studies only reporting adherence data and four studies only reporting a combined indicator (e.g., adherence and quality in a single score). No studies reported only quality, dosage, responsiveness, or differentiation as fidelity data.

Type of Fidelity Data Reported

The majority of the 38 studies reported quantitative data (k = 34; 89%) for at least one

fidelity dimension, with four studies (11%) providing only qualitative data (Georgiou et al., 2020; Mayfield, 2000; Morris et al., 2012; Vadasy et al., 1997b). For the studies that reported adherence, quality, differentiation, or a combined indicator, data were reported quantitatively over 80% of the time. However, four of the seven studies that reported responsiveness provided quantitative data, and three provided only qualitative data.

Study Features Associated with Fidelity Reporting (RQ3)

The 51 included studies evaluated 68 treatment conditions. Table 5 summarizes the key characteristics of the studies reviewed and the interventions they evaluated in relation to fidelity reporting. In the following sections, we present the results of our exploratory analyses examining the relations between study features and fidelity reporting.

Publication Year

Although our search included studies published between 1980 and 2020, the oldest studies were published in 1997 (Vadasy et al., 1997a; Vadasy et al., 1997b). Of the 51 studies, 8% were published in 1995-1999, 22% in 2000-2004, 31% in 2005-2009, 24% in 2010-2015, and 16% in 2016-2020. Study publication year was not statistically significantly related to the occurrence of treatment fidelity reporting, X^2 (4, k = 51) = 8.03, p = .08.

Research Design

All qualifying studies had designs that compared treatment groups to business-as-usual or no-foundational-skills reading instruction comparison groups. Most included studies (k = 43; 84%) employed a randomized control trial (RCT) design. Eight studies (16%) employed quasi-experimental designs (QEDs). There was not a statistically significant relation between research design and prevalence of treatment fidelity reporting, X^2 (1, k = 51) = 3.00, p = .08.

Student Grade Level

Most studies (k = 43; 84%) included students in grades K-2, with only eight studies (16%) focusing on students in grades 3-5. Student grade level was also not significantly associated with treatment fidelity reporting, X^2 (1, k = 51) < 0.01, p = .97.

Intervention Implementer

Most of the 68 treatment interventions were implemented by either school personnel (47%) or a research team member (32%). Approximately 9% of interventions were computer-delivered or technology-supported. The remaining 12% of interventions were implemented by someone who did not fall into these categories (e.g., community volunteers). The intervention implementer was significantly related to the occurrence of fidelity reporting, X^2 (3, N = 68) = 9.95, p = .02. Studies of interventions implemented by school personnel reported treatment fidelity more frequently than did studies of interventions delivered by research team members, computer-delivered interventions, or other implementers.

Effect Sizes Associated with Fidelity Reporting (RQ4)

The average effect on combined outcomes for studies that reported treatment fidelity was estimated as g = 0.30 (95% CI [0.21, 0.39], p < .01). The average effect on combined outcomes for studies that did not report treatment fidelity was estimated as g = 0.50 (95% CI [0.35, 0.64], p < .01). As is illustrated in Table 6, results of the test of heterogeneity indicated a statistically significant difference in effect size between the two groups (i.e., studies that reported treatment fidelity and studies that did not; Q = 5.04, p = .03). Therefore, we can conclude that the differences between the two groups of studies (k = 38 for fidelity reported and k = 13 for fidelity not reported) exist because studies' membership within the "reported" or "not reported" category.

Discussion

The purpose of this meta-analysis was to examine fidelity reporting within studies of reading interventions for students with or at risk for dyslexia in Grades K-5. In particular, we aimed to record the frequency and extent of fidelity reporting, explore associations between study features and fidelity reporting, and compare mean intervention effect sizes for studies that reported information about implementation fidelity and those that did not.

Frequency of Fidelity Reporting in Dyslexia Research

Of the 51 studies, 38 (75%) reported treatment fidelity data. Given the low prevalence of fidelity reporting observed in previous syntheses of academic interventions (e.g., Capin et al., 2018; Gresham et al., 2000; Swanson et al., 2011), this was a welcome surprise. This result suggests that studies evaluating the effects of reading interventions for students with or at risk for dyslexia are more likely to report information about intervention implementation fidelity than studies that evaluate the effects of reading interventions for typically developing students or students with reading difficulties broadly defined (i.e., defined as including a wide range of reading and language difficulties). However, it is not entirely clear why this would be the case. As noted previously, fidelity of intervention implementation has been shown to make a larger contribution to language and literacy outcomes for students with learning difficulties than for their typically developing peers (e.g., Boardman et al., 2016; Connor et al., 2007; Hamre et al., 2010; Neugebauer et al., 2017), but research does not suggest that fidelity of intervention implementation is more important in the remediation of word-reading difficulties than other types of language or reading comprehension difficulties.

We hypothesize that the reason for the relatively high rate of fidelity reporting in the studies in the present review relative to studies in previous reviews is as much (or more) due to our stringent criteria for inclusion than the population of students (i.e., students with or at risk for

dyslexia) included in reviewed studies. To determine that students had word reading difficulties prior to receiving the intervention, we required that studies report a screening threshold or a pretest mean on a norm-referenced measure of word reading or skills foundational to word reading. We also only included studies that reported intervention effects on norm-referenced outcome measures. Because studies that employ norm-referenced measures may be more rigorous in other research methods (i.e., including the measurement and reporting of treatment fidelity), it is perhaps unsurprising that the rate of fidelity reporting was higher in the studies included in the present review. The rate of fidelity reporting also may have been affected by the fact that our studies were published, on average, later than those examined in previous reviews. This was both a result of the fact that our search was conducted more recently (e.g., 2015 was the end date for the search conducted during the Capin et al. [2018] review), and the fact that older studies were less likely to report screening, pretest, and outcome scores on norm-referenced measures (and thus less likely to be included). Previous meta-analyses have reported an association between study publication date and study rigor, including measurement of implementation fidelity (Hall et al., 2017; Scammacca et al., 2015; Slavin & Smith, 2009). We hypothesize that our search dates and stringent criteria for inclusion were thus the most important factors explaining greater frequency of fidelity reporting in the studies included in the present review.

How Fidelity is Measured

Although nearly 80% of studies (30 out of 38) reported measuring multiple aspects of fidelity, only one study provided information about all five dimensions (i.e., adherence, quality, dosage, student responsiveness, and differentiation). Adherence was the dimension of fidelity that was most often the focus of fidelity reporting. Student responsiveness and differentiation

were much less frequently reported as being measured. It is noteworthy that, of the seven studies that reported responsiveness, four provided quantitative data and three provided only qualitative data (e.g., “during our observations of tutoring, students were engaged, and they responded frequently” [Al Otaiba et al., 2005, p. 205]). Regarding differentiation, 20 out of 38 studies provided information about what the comparison received, but only 11% (4 studies) compared treatment and control conditions with one another using common protocols. It is also worthy to note that combined indicators of fidelity (e.g., adherence and quality in a single score) were common.

These findings related to fidelity dimensions were different from those by Gresham et al. (2000), who only examined the adherence dimension of fidelity and found that only 18% of studies of academic interventions conducted with students with learning disabilities reported treatment adherence. They also differed from findings reported by Swanson et al. (2011), who examined the adherence, quality, and dosage dimensions of fidelity reported in studies of academic interventions published in general education or special education journals and found that adherence was reported in 47% of studies, quality reported in 7% of studies, and dosage reported in 5% of studies. In the review conducted by Capin et al. (2018), which was most comparable to the present review (because it included studies of reading interventions with students with or at risk for reading difficulties, examined all five dimensions of fidelity, and had a more recent search end date), adherence (90%) and quality of implementation (17%) were the dimensions of fidelity reported most frequently. In contrast, we found adherence (68%) and dosage (58%) to be the most prevalent dimensions of fidelity reported. As was true in the Capin et al. review, most studies included in the present review reported quantitative data for at least one fidelity dimension. However, even quantitative data intended to measure the same dimension

of fidelity (e.g., adherence) was not always collected or reported in a consistent manner.

Associations Between Study Characteristics and Fidelity Reporting

Although we explored the degree to which several variables (i.e., study publication year, research design, student grade level, and intervention implementer) were associated with fidelity reporting, only intervention implementer was statistically significantly associated with measuring and reporting fidelity data. Specifically, interventions implemented by school personnel reported treatment fidelity more frequently than interventions delivered by research team members, computer-delivered interventions, or other implementers. Capin et al. (2018) reported a similar finding in that, in their review, studies with paraprofessional implementers reported treatment fidelity more frequently (78%) than studies with research team members (45%), teachers (34%), or other implementers (e.g., community volunteers; 39%); note that they distinguished between teacher-delivered and paraprofessional-delivered interventions whereas we collapsed those two categories into an overarching "school personnel" category. However, Capin and colleagues observed that "there appears to be a threat to the validity of this finding. Patricia Vadasy was the first author on nine of the 22 studies conducted with paraprofessionals, and all of these studies reported treatment fidelity…if these studies were removed, the percentage of studies with paraprofessional implementers reporting treatment fidelity (46%) would have approximated the level of treatment fidelity reporting across all studies (47%)" (p. 10). The same confound may have impacted findings in the present review, within which 14 of the 68 treatment groups were part of studies led by Vadasy. Of those 14 treatment groups, 11 were implemented by school personnel, and all except for one reported fidelity. If these Vadasy-led studies were excluded from the review, then a much smaller proportion of studies that examined interventions implemented by school personnel might have been found to report information about fidelity of

implementation.

Although publication year (i.e., fidelity reporting in studies published prior to 2001, between 2001-2010, and after 2010) was not a statistically significant moderator of intervention effects at the $p \leq 0.05$ level, the difference between those groups of studies approached statistical significance ($p = .06$). Because of the previously discussed association between study rigor and publication date (Hall et al., 2017; Scammacca et al., 2015; Slavin & Smith, 2009), it may be surprising that there was not a bigger difference between fidelity reporting in earlier relative to later studies. Capin et al. (2018) found a statistically significant difference in fidelity reporting based on publication year, with the proportion of studies reporting treatment fidelity data increasing steadily over time. An explanation for the finding in the present review is that our inclusion criteria ensured that even earlier studies employed relatively rigorous methods, which again may have been associated with a higher rate of implementation fidelity measurement.

Unexpectedly, research design was not statistically significantly associated with fidelity reporting. Studies with more rigorous research designs might be expected to employ more rigorous methods across the board, including measuring and reporting fidelity of intervention implementation. It is worth noting that only eight of the 51 studies included in this meta-analysis employed QEDs, which impacted degrees of freedom and power to discover group differences for this variable. Descriptively speaking, a large number of the studies that did not report fidelity employed QEDs. Specifically, of the 13 studies that did not report information about implementation fidelity, nine were RCTs and four used QEDs. Of the 38 studies that reported information about implementation fidelity, 34 were RCTs, and only four used QEDs. In contrast to the present finding, Capin et al. (2018) found that fidelity reporting in their corpus of studies was associated with study design, with single case experimental design studies more likely to

report information about implementation fidelity than group design studies. We did not include studies that employed single case experimental designs.

Fidelity Reporting and Intervention Effect Size

Finally, we compared mean effects for studies that reported fidelity data with mean effects for those that did not. To be clear, this question did not seek to compare mean effects associated with interventions that were implemented with high levels of fidelity compared with effects associated with interventions that were implemented with lower levels of fidelity. Due to inconsistent methods for measuring and reporting information about the five dimensions of fidelity, this analysis was not possible. Instead, our research question aimed to explore the degree to which fidelity reporting, as a marker for study rigor, was associated with intervention effects.

We found that the average effect on combined literacy outcomes for studies that reported treatment fidelity was estimated as $g = 0.30$, whereas the average effect on combined outcomes for studies that did not report treatment fidelity was estimated as $g = 0.50$. The difference in mean effect size between the two groups of studies was statistically significant. It is unlikely that this effect size difference reflects the influence of fidelity measurement and reporting on intervention effectiveness (i.e., it is unlikely that interventions are less likely to be effective if researchers measure and report information on fidelity of implementation). Instead, the differences in mean effects are likely because studies that reported treatment fidelity were more rigorous overall, and numerous meta-analyses have reported an association between variables that are proxies for study rigor and effect size, with more rigorous studies being associated with smaller effects (Hall et al., 2017; Scammacca et al., 2015; Slavin & Smith, 2009). Superrealization bias is one explanation (Cheung & Slavin, 2012; Slavin & Smith, 2009); higher-quality studies (i.e., ones with larger samples and/or more rigorous research designs) may tend to

be conducted in less controlled settings in which experimenters are unable to provide the same degree of support to implementers or ensure highest-quality intervention implementation. Studies conducted in more tightly controlled experimental conditions can rarely be replicated at a large scale but are more likely to yield large effects. In addition, publication bias may explain the association between higher-quality studies and smaller effects: not only do smaller-sample studies need to achieve much larger differences between groups to achieve statistically significant findings (which are too often a prerequisite for publication), but journal editors are also more likely to publish high-quality studies with small or null effects than they are to publish lower-quality studies reporting small or null effects. All these factors may explain why studies that reported measures of implementation fidelity tended to yield smaller effect sizes.

Limitations

Relative to other meta-analyses that have examined fidelity reporting within studies of academic interventions, the present meta-analysis included a small sample of studies (k = 53). Although focusing on intervention research for students with or at risk for dyslexia provides a unique opportunity to inform practice, we did not have enough power to draw conclusions for some analyses. For example, 32 of the 38 studies that reported fidelity included students in Grades K-2, with only six including students in Grades 3-5. Similarly, most of these 38 studies were RCTs (89%); only four studies employed QEDs (11%). Our analyses would have been better powered had our corpus of included studies been larger. Additionally, one potential methodological limitation of this study was the use of chi-square analyses to explore associations between study features and treatment fidelity reporting. Chi-square tests only allow for categorical variables to be examined. Given that three of our four variables of interest (i.e., research design, grade level, and implementer) were naturally categorical, we chose to examine

publication year as a categorical variable (rather than a continuous variable) as well so that we could apply the same analytic methods to all our study features of interest. As a result of our small sample size and use of chi-square analyses for examining all associations, reported associations between study features and treatment fidelity reporting should be considered exploratory rather than conclusive.

Another limitation is that 14 of the 68 treatment groups included in the present review were part of studies led by Vadasy. All except for one of these Vadasy-led studies reported fidelity. If these Vadasy-led studies were removed from the present corpus of studies, the overall percentage of studies reporting information about implementation fidelity may have differed. Additionally, this meta-analysis was limited by the information provided in the included studies. In particular, it is feasible that researchers may have measured treatment fidelity but not reported it within their manuscripts. Therefore, the present review may not fully reflect the amount of treatment fidelity measured in K-5 reading intervention studies. Further, there were times when an author briefly mentioned an aspect of fidelity that was measured, but in such a way that it was unclear whether it should be described as adherence data or information about the quality of delivery. Thus, in some cases, it was difficult to make decisions about how to code certain reported aspects of fidelity.

Finally, as noted previously, it had originally been our intent to examine whether there was an association between fidelity of implementation scores and intervention effects. However, this analysis was not possible due to the widely varying methods for reporting information about the five dimensions of fidelity. One key finding of the present review is the overall inconsistency in fidelity reporting; it is also a major limitation. Although 75% of studies reported treatment fidelity data, only one study (Denton et al., 2010) reported information for all five dimensions of

fidelity. The most reported dimension of fidelity was adherence (with 68% of studies reporting adherence data), but even methods for reporting adherence are widely varied. Typically, authors reported an adherence percentage based on the number of required steps or components to be completed within an activity. However, in some instances, adherence was measured based purely on whether the step or component was completed; in other instances, adherence was measured based on how often the step or component was present (e.g., on a 4-point Likert scale where 1 = not present, 2 = inconsistently present, 3 = mostly present, 4 = always present). It would be beneficial if authors of future research could report fidelity data with enough precision and comparability that future meta-analysts could answer such a question.

Implications and Future Directions

It is good news that a relatively high proportion of studies that include students with or at risk for dyslexia report information about treatment fidelity (i.e., a higher proportion than has been reported in previous research with more diverse populations of students; Capin et al., 2018; Gresham et al., 2000; Swanson et al., 2011). Note that treatment fidelity has been shown to be a more important contributor to reading outcomes for students with learning difficulties, such as dyslexia (Boardman et al., 2016), so care to ensure that treatments can be implemented faithfully is important in the context of this group. Nevertheless, there were still a number of studies in our corpus that did not report fidelity data or did not report on all five dimensions of fidelity. Each dimension of fidelity provides valuable information about the extent to which the intervention was implemented as intended. These fidelity dimensions also provide information about the conditions under which the effects were obtained and can inform decisions about how to scale up interventions. Therefore, further research is needed to understand how each dimension of fidelity is associated with intervention effects. It is also important that researchers measure and report

fidelity data in a precise and consistent method. Without measuring all five dimensions of fidelity and reporting disaggregated fidelity scores by dimension, it is difficult to take a nuanced view of fidelity and explore which dimensions are most important for students with or at risk for dyslexia. As a result, we suggest that researchers conducting future reading intervention studies take a few key steps: (a) quantitatively measure all five dimensions of fidelity (Dane & Schneider, 1998) using valid and reliable methods and report disaggregated data, (b) clearly document and report all data collection procedures to support future replication, (c) analyze treatment fidelity and their relation to student outcomes to inform under what conditions interventions are successful, and (d) consider analyzing the interactions between dimensions of fidelity (e.g., adherence and quality) and their impact on treatment effects (Capin et al., 2022). Measurement research can support these efforts by, for instance, developing feasible methods for collecting fidelity information that have strong psychometric evidence (Sanetti & Kratochwill, 2009). By taking these steps, researchers can better support study internal validity, document the degree to which an intervention is feasible to implement, and provide information about which dimensions of implementation fidelity are the biggest contributors to intervention effects.

References

Al Otaiba, S., Connor, C. M., Foorman, B., Schatschneider, C., Greulich, L., & Sidler, J. F. (2009). Identifying and intervening with beginning readers who are at-risk for dyslexia: Advances in individualized classroom instruction. *Perspectives on Language and Literacy*, *35*(4), 13-19. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4296731/

*Al Otaiba, S., Schatschneider, C., & Silverman, E. (2005). Tutor-assisted intensive learning strategies in kindergarten: How much is enough?. *Exceptionality*, *13*(4), 195-208. https://doi.org/10.1207/s15327035ex1304_2

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

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

Boardman, A. G., Buckley, P., Vaughn, S., Roberts, G., Scornavacco, K., & Klingner, J. K. (2016). Relationship between implementation of collaborative strategic reading and student outcomes for adolescents with disabilities. *Journal of Learning Disabilities*, *49*(6), 644-657. https://doi.org/10.1177/0022219416640784

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2013). Comprehensive Meta-Analysis [software] Version 3. Biostat.

Boscardin, C., Muthén, B., Francis, D., & Baker, E. (2008). Early identification of reading difficulties using heterogeneous developmental trajectories. *Journal of Educational*

*Psychology, 100(*1), 192–208. https://doi.org/10.1037/0022-0663.100.1.192

*Burns, D. A. (2011). *Examining the effect of an overt transition intervention on the reading development of at-risk English-language learners in first grade*. University of Oregon.

Capin, P., Roberts, G., Clemens, N. H., & Vaughn, S. (2022). When treatment adherence matters: Interactions among treatment adherence, instructional quality, and student characteristics on reading outcomes. *Reading Research Quarterly, 57*(2), 753-774. https://doi.org/10.1002/rrq.442

Capin, P., Walker, M. A., Vaughn, S., & Wanzek, J. (2018). Examining how treatment fidelity is supported, measured, and reported in K–3 reading intervention research. *Educational Psychology Review, 30*(3), 885-919. https://doi.org/10.1007/s10648-017-9429-z

Cheung, A. C., & Slavin, R. E. (2012). How features of educational technology applications affect student reading outcomes: A meta-analysis. *Educational Research Review, 7*(3), 198–215. https://doi.org/10.1016/j.edurev.2012.05.002

*Christodoulou, J. A., Cyr, A., Murtagh, J., Chang, P., Lin, J., Guarino, A. J., Hook, P., & Gabrieli, J. D. (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. https://doi.org/10.1177/0022219415617163

Connor, C. M., Morrison, F. J., Fishman, B. J., Schatschneider, C., & Underwood, P. (2007). Algorithm-guided individualized reading instruction. *Science, 315*(5811), 464-465. https://doi.org/10.1126/science.1134513

*Coyne, M. D., Little, M., Rawlinson, D. A., Simmons, D., Kwok, O. M., Kim, M., Hagan-Burke, S., & Civetelli, C. (2013). Replicating the impact of a supplemental beginning reading intervention: The role of instructional context. *Journal of Research on*

Educational Effectiveness, 6(1), 1-23. https://doi.org/10.1080/19345747.2012.706694

Dahle, A. E., & Knivsberg, A.-M. (2014). Internalizing, externalizing and attention problems in dyslexia. *Scandinavian Journal of Disability Research, 16*(2), 179–193. https://doi.org/10.1080/15017419.2013.781953

Dane, A. V., & Schneider, B. H. (1998). Program integrity in primary and early secondary prevention: Are implementation effects out of control? *Clinical Psychology Review, 18*(1), 23-45. https://doi.org/10.1080/15017419.2013.781953

Daniel, S. S., Walsh, A. K., Goldston, D. B., Arnold, E. M., Reboussin, B. A., & Wood, F. B. (2006). Suicidality, school dropout, and reading problems among adolescents. *Journal of Learning Disabilities, 39*(6), 507–514. https://doi.org/10.1177/00222194060390060301

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

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

*Donegan, R. E., Wanzek, J., & Al Otaiba, S. (2020). Effects of a reading intervention implemented at differing intensities for upper elementary students. *Learning Disabilities Research & Practice, 35*(2), 62-71. https://doi.org/10.1111/ldrp.12218

*Duff, F. J., Hulme, C., Grainger, K., Hardwick, S. J., Miles, J. N., & Snowling, M. J. (2014). Reading and language intervention for children at risk of dyslexia: A randomised

controlled trial. *Journal of Child Psychology and Psychiatry*, *55*(11), 1234-1243. https://doi.org/10.1111/jcpp.12257

*Fawcett, A. J., Nicolson, R. I., Moss, H., Nicolson, M. K., & Reason, R. (2001). Effectiveness of reading intervention in junior school. *Educational Psychology*, *21*(3), 299-31. https://doi.org/10.1080/01443410120065496

Fletcher, J. M., Lyon, G. R., Fuchs, L. S., & Barnes, M. A. (2019). *Learning disabilities: From identification to intervention* (2nd ed.). Guilford Press.

Francis, D., Shaywitz, S., Stuebing, K., Shaywitz, B., & Fletcher, J. (1996). Developmental lag versus deficit models of reading disability: A longitudinal, individual growth curves analysis. *Journal of Educational Psychology, 88*(1), 3–17. https://doi.org/10.1037/0022-0663.88.1.3

Fuchs, D., Fuchs, L. S., & Stecker, P. M. (2010). The "blurring" of special education in a new continuum of general education placements and services. *Exceptional Children, 76*(3), 301-323. https://doi.org/10.1177/001440291007600304

*Frantz, S. (2000). *Effectiveness of the infusion of Reading Component Model based remedial reading instruction on the reading achievement of students in learning disabilities and Title I remedial reading programs.* Indiana State University.

Galuschka, K., Görgen, R., Kalmar, J., Haberstroh, S., Schmalz, X., & Schulte-Körne, G. (2020). Effectiveness of spelling interventions for learners with dyslexia: A meta-analysis and systematic review. *Educational Psychologist, 55*(1), 1-20. https://doi.org/10.1080/00461520.2019.1659794

*Georgiou, G. K., Savage, R., Dunn, K., Bowers, P., & Parrila, R. (2020). Examining the effects of Structured Word Inquiry on the reading and spelling skills of persistently poor Grade 3

readers. *Journal of Research in Reading, 44*(1), 131-153. https://doi.org/10.1111/1467-9817.12325

Gersten, R., Compton, D., Connor, C.M., Dimino, J., Santoro, L., Linan-Thompson, S., and Tilly, W.D. (2009). *Assisting students struggling with reading: Response to Intervention and multi-tier intervention for reading in the primary grades. A practice guide.* (NCEE 2009-4045). National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. https://ies.ed.gov/ncee/wwc/practiceguide/3

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

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

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

Greenberg, E., Dunleavy, E., & Kutner, M. (2007). Literacy behind bars: Results from the 2003 national assessment of adult literacy prison survey. NCES 2007-473. *National Center for Education Statistics.* https://eric.ed.gov/?id=ED496564

Gresham, F. M., MacMillan, D. L., Beebe-Frankenberger, M. E., & Bocian, K. M. (2000). Treatment integrity in learning disabilities intervention research: Do we really know how treatments are implemented? *Learning Disabilities Research & Practice, 15*(4), 198-205.

https://doi.org/10.1207/SLDRP1504_4

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

*Hagans, K. S., & Good, R. H. (2013). Decreasing reading differences in children from disadvantaged backgrounds: The effects of an early literacy intervention. *Contemporary School Psychology*, *17*(1), 103-117. https://doi.org/10.1007/BF03340992

Hall, C., Dahl-Leonard, K., Cho, E., Solari, E. J., Capin, P., Conner, C., Henry, A. R., Cook, L., Hayes, L., Richmond, C. L., Vargas, I., & Kehoe, K. F. (2022). Forty years of reading intervention research for elementary students with or at risk for dyslexia: A systematic review and meta-analysis. *Reading Research Quarterly*. Advance online publication. https://doi.org/10.1002/rrq.477

Hall, C., Roberts, G. J., Cho, E., McCulley, L. V., Carroll, M., & Vaughn, S. (2017). Reading instruction for English learners in the middle grades: A meta-analysis. *Educational Psychology Review*, *29*(4), 763-794. https://doi.org/10.1007/s10648-016-9372-4

Hamre, B. K., Justice, L. M., Pianta, R. C., Kilday, C., Sweeney, B., Downer, J. T., & Leach, A. (2010). Implementation fidelity of MyTeachingPartner literacy and language activities: Association with preschoolers' language and literacy growth. *Early Childhood Research Quarterly*, *25*(3), 329-347. https://doi.org/10.1016/j.ecresq.2009.07.002

*Hatcher, P. J., Hulme, C., Miles, J. N., Carroll, J. M., Hatcher, J., Gibbs, S., Smith, G., Bower-Crane, C., & Snowling, M. J. (2006). Efficacy of small group reading intervention for beginning readers with reading-delay: A randomised controlled trial. *Journal of Child*

Psychology and Psychiatry, 47(8), 820-827. https://doi.org/10.1111/j.1469-7610.2005.01559.x

Hedges, L. V. (1981). Distribution theory for Glass's estimator of effect size and related estimators. *Journal of Educational Statistics, 6*(2), 107-128. https://doi.org/10.3102/10769986006002107

Hernandez, D. J. (2011). Double jeopardy: How third-grade reading skills and poverty influence high school graduation. *Annie E. Casey Foundation.* https://eric.ed.gov/?id=ED518818

Horn, L., & Berktold, J. (1999). Students with disabilities in postsecondary education: A profile of preparation, participation, and outcomes. Postsecondary education descriptive analysis reports. Statistical analysis report. *ERIC*. https://eric.ed.gov/?id=ED4312688

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

Jordan, J. A., McGladdery, G., & Dyer, K. (2014). Dyslexia in higher education: Implications for maths anxiety, statistics anxiety and psychological well-being. *Dyslexia, 20*(3), 225-240. https://doi.org/10.1002/dys.1478

Kretlow, A. G., & Bartholomew, C. C. (2010). Using coaching to improve the fidelity of evidence-based practices: A review of studies. *Teacher Education and Special Education, 33*(4), 279-299. https://doi.org/10.1177/0888406410371643

Lembke, E. S., McMaster, K. L., & Stecker, P. M. (2010). The prevention science of reading research within a response-to-intervention model. *Psychology in the Schools, 47*(1), 22-35. https://doi.org/10.1002/pits.20449

*Little, M. E., Rawlinson, D. A., Simmons, D. C., Kim, M., Kwok, O. M., Hagan-Burke, S.,

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

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

*Mayfield, L. G. (2000). *The effects of structured one-on-one tutoring in sight word recognition of first-grade students at-risk for reading failure.* Louisiana Tech University.

Miciak, J., & Fletcher, J. M. (2020). The critical role of instructional response for identifying dyslexia and other learning disabilities. *Journal of Learning Disabilities, 53*(5), 343-353. https://doi.org/10.1177/0022219420906801

*Miciak, J., Roberts, G., Taylor, W. P., Solis, M., Ahmed, Y., Vaughn, S., & Fletcher, J. M. (2018). The effects of one versus two years of intensive reading intervention implemented with late elementary struggling readers. *Learning Disabilities Research & Practice, 33*(1), 24-36. https://doi.org/10.1111/ldrp.12159

Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *Annals of Internal Medicine, 151*(4), 264-269. https://doi.org/10.7326/0003-4819-151-4-200908180-00135

Moncher, F. J., & Prinz, R. J. (1991). Treatment fidelity in outcome studies. *Clinical Psychology Review, 11*(3), 247-266. https://doi.org/10.1016/0272-7358(91)90103-2

*Morris, R. D., Lovett, M. W., Wolf, M., Sevcik, R. A., Steinbach, K. A., Frijters, J. C., &

Shapiro, M. B. (2012). Multiple-component remediation for developmental reading disabilities: IQ, socioeconomic status, and race as factors in remedial outcome. *Journal of Learning Disabilities, 45*(2), 99–127. https://doi.org/10.1177/0022219409355472

Mugnaini, D., Lassi, S., La Malfa, G., & Albertini, G. (2009). Internalizing correlates of dyslexia. *World Journal of Pediatrics, 5*(4), 255–264. https://doi.org/10.1007/s12519-009-0049-7

Nelson, M. C., Cordray, D. S., Hulleman, C. S., Darrow, C. L., & Sommer, E. C. (2012). A procedure for assessing intervention fidelity in experiments testing educational and behavioral interventions. *The Journal of Behavioral Health Services & Research, 39*(4), 374-396. https://doi.org/10.1007/s11414-012-9295-x

Neugebauer, S., Coyne, M., McCoach, B., & Ware, S. (2017). Teaching beyond the intervention: The contribution of teacher language extensions to vocabulary learning in urban kindergarten classrooms. *Reading and Writing, 30*(3), 543-567. https://doi.org/10.1007/s11145-016-9689-x

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

*O'Callaghan, P., McIvor, A., McVeigh, C., & Rushe, T. (2016). A randomized controlled trial of an early-intervention, computer-based literacy program to boost phonological skills in 4-to 6-year-old children. *British Journal of Educational Psychology, 86*(4), 546-558. https://doi.org/10.1111/bjep.12122

*O'Connor, R. E., Bell, K. M., Harty, K. R., Larkin, L. K., Sackor, S. M., & Zigmond, N. (2002). Teaching reading to poor readers in the intermediate grades: A comparison of text

difficulty. *Journal of Educational Psychology, 94*(3), 474. https://doi.org/10.1037/0022-0663.94.3.474

O'Donnell, C. L. (2008). Defining, conceptualizing, and measuring fidelity of implementation and its relationship to outcomes in K–12 curriculum intervention research. *Review of Educational Research, 78*(1), 33-84. https://doi.org/10.3102/0034654307313793

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

Peterson, R. L., & Pennington, B. F. (2012). Developmental dyslexia. *The Lancet, 379*(9830), 1997–2007. https://doi.org/10.1016/S0140-6736(12)60198-6

Rosenthal, R. (1991). *Meta-analytic procedures for social science research.* Sage.

Sanetti, L. M. H., & Kratochwill, T. R. (2009). Toward developing a science of treatment integrity: Introduction to the special series. *School Psychology Review, 38*(4), 445.

Scammacca, N., 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

Scammacca, N., Vaughn, S., Roberts, G., Wanzek, J., & Torgesen, J. K. (2007). Extensive reading interventions in grades K–3: From research to practice. *RMC Research Corporation, Center on Instruction.* https://eric.ed.gov/?id=ED521573

*Scanlon, D. M., Vellutino, F. R., Small, S. G., Fanuele, D. P., & Sweeney, J. M. (2005). Severe reading difficulties—Can they be prevented? A comparison of prevention and intervention approaches. *Exceptionality, 13*(4), 209-227. https://doi.org/10.1207/s15327035ex1304_3

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference.* Houghton Mifflin.

Shaywitz, S. E., Escobar, M. D., Shaywitz, B. A., Fletcher, J. M., & Makuch, R. (1992). Evidence that dyslexia may represent the lower tail of a normal distribution of reading ability. *New England Journal of Medicine, 326*(3), 145-150. https://doi.org/10.1056/nejm199201163260301

*Simmons, D. C., Coyne, M. D., Hagan-Burke, S., Kwok, O. M., Simmons, L., Johnson, C., Zou, Y., Taylor, A. B., Mcalenney A. L., Ruby, M., & Crevecoeur, Y. C. (2011). Effects of supplemental reading interventions in authentic contexts: A comparison of kindergarteners' response. *Exceptional Children, 77*(2), 207-228. https://doi.org/10.1177/001440291107700204

Slavin, R., & Smith, D. (2009). The relationship between sample sizes and effect sizes in systematic reviews in education. *Educational Evaluation and Policy Analysis, 31*(4), 500-506. https://doi.org/10.3102/0162373709352369

Snowling, M. J., & Melby-Lervåg, M. (2016). Oral language deficits in familial dyslexia: A meta-analysis and review. *Psychological Bulletin, 142*(5), 498. https://doi.org/10.1037/bul0000037

Stevens, E. A., Austin, C., Moore, C., Scammacca, N., Boucher, A. N., & Vaughn, S. (2021). Current state of the evidence: Examining the effects of Orton-Gillingham reading interventions for students with or at risk for word-level reading disabilities. *Exceptional Children, 87*(4), 397-417. https://doi.org/10.1177/0014402921993406

*Storey, C., McDowell, C., & Leslie, J. C. (2020). Headsprout early reading for specific literacy difficulty: A comparison study. *Journal of Behavioral Education, 29*(3), 619-633.

https://doi.org/10.1007/s10864-019-09336-7

Suggate, S. P. (2016). A meta-analysis of the long-term effects of phonemic awareness, phonics, fluency, and reading comprehension interventions. *Journal of Learning Disabilities, 49*(1), 77-96. https://doi.org/10.1177/0022219414528540

Swanson, E., Wanzek, J., Haring, C., Ciullo, S., & McCulley, L. (2011). Intervention fidelity in special and general education research journals. *Journal of Special Education, 47*(1), 3-13. https://doi.org/10.1177/0022466911419516

Swanson, H. L., Hoskyn, M., & Lee, C. (1999). *Interventions for students with learning disabilities: A meta-analysis of treatment outcomes*. Guilford Press.

Torgesen, J. K. (2004). Lessons learned from research on interventions for students who have difficulty learning to read. In P. McCardle & V. Chhabra (Eds.), *The voice of evidence in reading research* (pp. 355–382). Brookes Publishing Co.

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

*Toste, J. R., Capin, P., Williams, K. J., Cho, E., & Vaughn, S. (2019). Replication of an experimental study investigating the efficacy of a multisyllabic word reading intervention with and without motivational beliefs training for struggling readers. *Journal of Learning Disabilities, 52*(1), 45-58. https://doi.org/10.1177/0022219418775114

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

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

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

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

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

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

*Vadasy, P. F., Jenkins, J. R., Antil, L. R., & Wayne, S. K. (1997a). Community-based early reading intervention for at-risk first graders. *Learning Disabilities Research and Practice, 12*(1), 29-39.

*Vadasy, P. F., Jenkins, J. R., Antil, L. R., Wayne, S. K., & O'Connor, R. E. (1997b). The effectiveness of one-to-one tutoring by community tutors for at-risk beginning readers. *Learning Disability Quarterly, 20*(2), 126-139. https://doi.org/10.2307/1511219

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

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

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

VanDerHeyden, A. M., Witt, J. C., & Gilbertson, D. (2007). A multi-year evaluation of the effects of a Response to Intervention (RTI) model on identification of children for special education. *Journal of School Psychology, 45*(2), 225–256. https://doi.org/10.1016/j.jsp.2006.11.004

*Vaughn, S., Cirino, P. T., Linan-Thompson, S., Mathes, P. G., Carlson, C. D., Cardenas-Hagan, E., Pollard-Durodola, S., Fletcher, J. M., & Francis, D. J. (2006b). Effectiveness of a Spanish intervention and an English intervention for English-language learners at risk for reading problems. *American Educational Research Journal, 43*(3), 449-487. https://doi.org/10.3102/00028312043003449

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

Vaughn, S., Wanzek, J., & Fletcher, J. M. (2007). Multiple tiers of intervention: A framework for prevention and identification of students with reading/learning disabilities. In B. M. Taylor & J. E. Ysseldyke (Eds.), *Effective instruction for struggling readers, K-6* (pp. 173-195). Teacher's College Press.

Vaughn, S., Wanzek, J., Murray, C. S., & Roberts, G. (2012). *Intensive interventions for students struggling in reading and mathematics: A practice guide.* Center on Instruction. https://eric.ed.gov/?id=ED531907

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

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

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

*Wanzek, J., Otaiba, S. A., Schatschneider, C., Donegan, R. E., Rivas, B., Jones, F., & Petscher, Y. (2020). Intensive intervention for upper elementary students with severe reading comprehension difficulties. *Journal of Research on Educational Effectiveness, 13*(3), 408-429. https://doi.org/10.1080/19345747.2019.1710886

Wanzek, J., Stevens, E. A., Williams, K. J., Scammacca, N., Vaughn, S., & Sargent, K. (2018). Current evidence on the effects of intensive early reading interventions. *Journal of*

Learning Disabilities, *51*(6), 612-624. https://doi.org/10.1177/0022219418775110

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

Willcutt, E. G., Betjemann, R. S., McGrath, L. M., Chhabildas, N. A., Olson, R. K., DeFries, J. C., & Pennington, B. F. (2010). Etiology and neuropsychology of comorbidity between RD and ADHD: The case for multiple-deficit models. *Cortex, 46*(10), 1345–1361. https://doi.org/10.1016/j.cortex.2010.06.009

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

Figure 1

PRISMA Search Flow Diagram

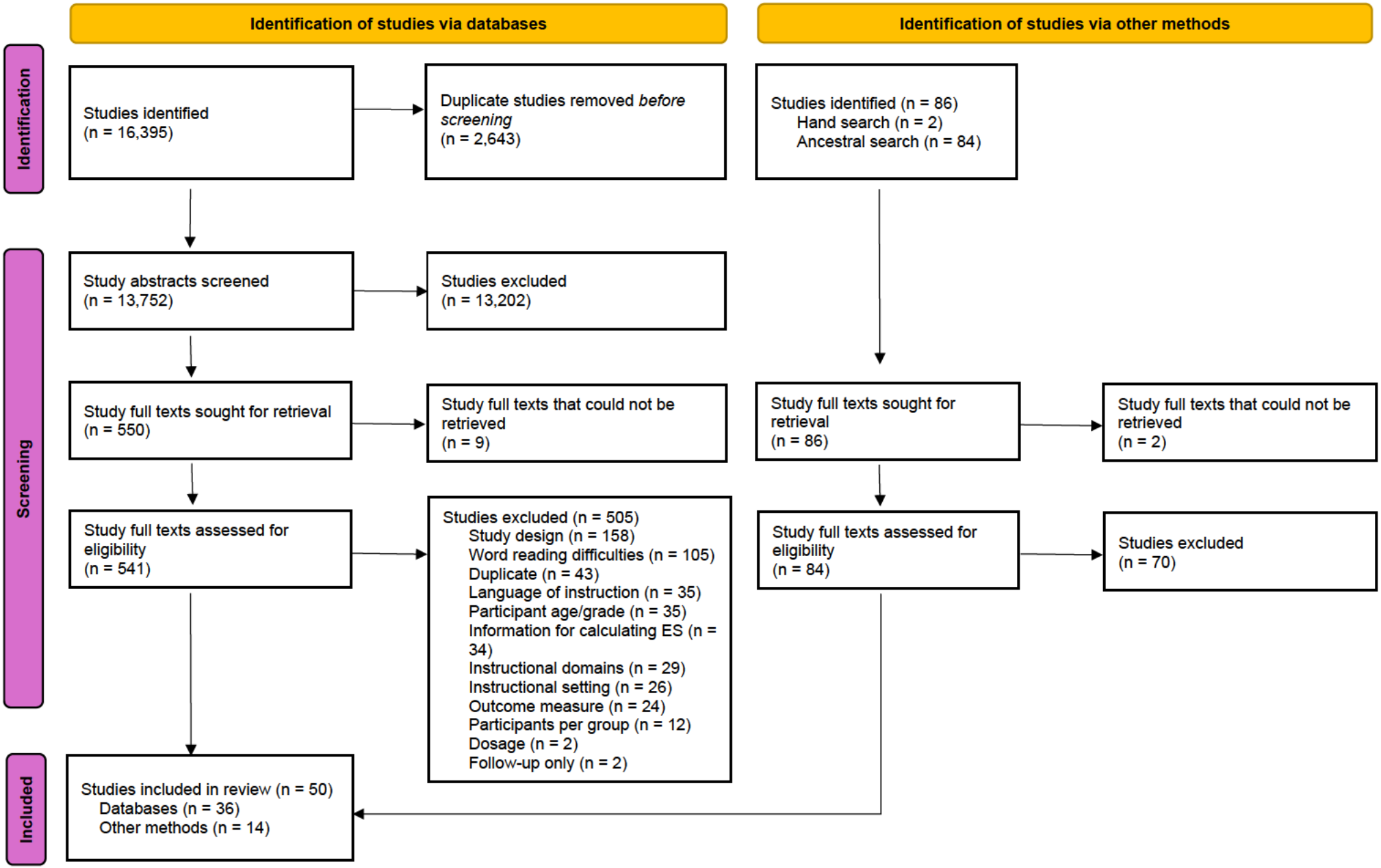

Table 1

Comparing Characteristics of Previous Reviews of Fidelity Reporting

	Swanson et al. (2011)	**Capin et al. (2018)**	**Present Meta-Analysis**
Search Dates	2005-2009	1995-2015	1980-2020
Total Studies	76	175	53
Study Designs	experimental, quasi-experimental, or single case designs	experimental, quasi-experimental, or single case designs	experimental or quasi-experimental (with at least 15 participants per group)
Interventions	"student-level reading, mathematics, or writing intervention" (p. 6)	"intervention targeted early literacy in English (i.e., phonics, fluency, phonemic awareness, vocabulary, reading comprehension, or spelling)" (p. 7)	intervention focused on PA, phonics, word reading, spelling, or a combination of these domains
Grades	PK-12	K-3	K-5
Population of Interest	N/A[1]	"identified with a learning disability, reading difficulty, or as at risk for reading difficulties (e.g., students with low achievement, low phonemic awareness, language disorders)" (p. 7)	a screening criterion at or below the 25th percentile, or a baseline score at or below the 16th percentile, on a norm-referenced test of word reading, spelling, or foundational word reading skills
Research Questions	"Research Question 1: What proportion of intervention studies published in peer-reviewed journals reported fidelity? Research Question 2: What components of fidelity measurement were reported in published intervention studies? Research Question 3: Did reported components differ	"RQ1. What proportion of K–3 reading intervention studies published in dissertations or peer-reviewed journals reported treatment fidelity data, and how did the proportion of fidelity reporting vary according to study features (e.g., publication year, research design, journal impact factor)? RQ2. What procedures did authors report using to train and support treatment implementers?	RQ 1: What proportion of foundational reading intervention studies for K–5 students with or at risk for dyslexia reported treatment fidelity data? RQ 2: For studies that reported fidelity information, what dimensions of fidelity (i.e., adherence, quality, dosage, responsiveness, or differentiation) were reported and how were they reported (i.e., quantitatively or qualitatively)?

	according to research design, intervention type, or number of intervention sessions?" (p. 6)	RQ3. What procedures did authors report using to collect observations of fidelity and measure treatment fidelity? RQ4. What dimensions of treatment fidelity (i.e., adherence, quality, receipt, dosage, and differentiation) and what levels of treatment fidelity were presented in published intervention studies? RQ5. To what extent were treatment fidelity scores used in the analysis of treatment effects?" (p. 6)	RQ 3: Was fidelity reporting associated with any other study/intervention features (i.e., publication year, research design, student grade level, and intervention implementer)? RQ 4: Was there a difference in the average effect size for studies that reported fidelity information relative to those that did not?

[1]Swanson et al. (2011) examined treatment fidelity reporting practices described in select high-impact factor journals that published general and special education intervention research (i.e., they did not require students to have or be at risk for any difficulties/disabilities).

Table 2

Definitions of Variables Used in Fidelity-Related Coding

Category	Variable	Definition
Type of Data	Quantitative	Reported numeric data (e.g., "teachers implemented 4 of the 5 steps [90%] on average; all teachers scored >85% on a measure of adherence).
	Qualitative	Reported narrative/descriptive information (e.g., the research team noticed high fidelity).
Dimensions	Adherence	Described the extent to which components are implemented as intended.
	Quality	Reported a measure of instructional quality separate from adherence to components (e.g., data related to pacing, corrective feedback, behavior management).
	Dosage	Reported the amount of instruction provided in relation to fidelity (e.g., data related to the number and length of sessions or number of students in attendance during each session).
	Receipt	Described the extent to which participants responded to the intended treatment (e.g., data collected based on student behavior, such as time on task or engagement).
	Differentiation	Described the extent to which the treatment condition varied from the comparison condition using a common protocol.
	Combined	Reported two or more dimensions as one combined score (e.g., adherence and quality are reported as one "fidelity" score).

Table 3

Study Characteristics and Fidelity Dimensions Reported

Study	*N*	Grade	Design	Adherence	Quality	Dosage	Responsiveness	Differentiation	Combined
Al Otaiba et al. (2005)	73	K	RCT	X			X		
Baker et al. (2000)	84	1-2	RCT						
Blachman et al. (2004)	69	2-3	RCT	X		X			
Burns (2011)	78	1	RCT	X				X	
Christodoulou et al. (2017)	47	1-4	RCT						
Coyne et al. (2013)	162	K	RCT	X	X	X		X	
Denton et al. (2010)	422	1	RCT	X	X	X	X	X	
Denton et al. (2014)	206	1	RCT	X	X	X			
Donegan et al. (2020) Study 1	153	4	RCT	X	X		X		
Duff et al. (2014)	145	1	QED		X	X			
Fawcett et al. (2001)	87	2	QED						
Frantz (2000)	78	2-6	QED						
Georgiou et al. (2020)	48	3	RCT						X
Graham et al. (2002)	54	2	RCT	X					
Gunn et al. (2005)	245	K-3	RCT	X					
Hagans and Good (2013)	50	1	RCT						
Hatcher et al. (2006)	77	K	RCT						
Jenkins et al. (2004)	99	1	RCT			X			X
Little et al. (2012)	90	K	RCT	X		X			
Mathes et al. (2003)	89	1	QED	X					
Mayfield (2000)	60	1	RCT	X			X		
Miciak et al. (2018)	270	4	RCT	X	X	X			
Morris et al. (2012)	279	2-3	RCT						X
Nicolson et al. (1999)	102	1	QED						
O'Callaghan et al. (2016)	98	PK-K	RCT						
O'Connor et al. (2002)	46	3-5	RCT	X		X			

O'Shaughnessy and Swanson (2000)	45	2	RCT	X						
Scanlon et al. (2005)	84	K-1	RCT							
Simmons et al. (2011)	206	K	RCT	X	X			X		
Storey et al. (2020)	32	1-4	RCT							
Torgesen et al. (2010)	108	1	RCT	X		X				
Toste et al. (2019)	108	4-5	RCT	X	X					
Vadasy and Sanders (2008a)	86	K	QED	X		X				
Vadasy and Sanders (2008b)	162	2-3	RCT	X	X	X	X			
Vadasy and Sanders (2009)	202	2-3	RCT	X	X	X				
Vadasy and Sanders (2010)	84	K	RCT	X		X				
Vadasy and Sanders (2011)	98	1	RCT	X		X				
Vadasy et al. (1997a)	35	1	RCT							
Vadasy et al. (1997b)	40	1	RCT	X	X					
Vadasy et al. (2000)	46	1	RCT			X				X
Vadasy et al. (2005)	57	1	QED			X				X
Vadasy et al. (2006)	67	K	RCT			X				X
Vadasy et al. (2007)	43	2-3	RCT	X		X				
Vaughn et al. (2006a)	40	1	RCT			X				X
Vaughn et al. (2006b)	91	1	RCT			X				X
Vellutino et al. (2008)	113	K	RCT							
Wang and Algozzine (2008)	139	1	RCT							X
Wanzek and Vaughn (2008) Study 1	50	1	RCT							X
Wanzek and Vaughn (2008) Study 2	36	1	RCT				X			X
Wanzek et al. (2020)	260	4	RCT	X	X	X	X			
Wise et al. (1999)	153	2-5	QED							

Note. RCT = randomized controlled trial; QED = quasi-experimental design.

Table 4

Fidelity Data Collection and Reporting

	Studies Reporting Fidelity (*k* = 38)	
Variable	*k*	%
Dimension of Fidelity Reported		
Adherence	26	68
Quality	12	32
Dosage	58	22
Responsiveness	7	18
Differentiation	4	11
Combined	11	29
Type of Data Reported		
Quantitative	34	89
Qualitative Only	4	11

Table 5

Characteristics of the Reviewed Studies and Interventions

	All Studies (*k* = 51)		**Studies Reporting Fidelity (*k* = 38)**		**Studies Not Reporting Fidelity (*k* = 13)**	
Variable	*k*	%	*k*	%	*k*	%
Publication Year						
1995-1999	4	8	1	3	3	23
2000-2004	11	22	8	21	3	23
2005-2009	16	31	13	34	3	23
2010-2015	12	24	11	29	1	8
2016-2020	8	16	5	13	3	23
Research Design						
RCT	43	84	34	89	9	69
QED	8	16	4	11	4	31
Grade Level						
K-2	43	84	32	84	11	85
3-5	8	16	6	16	2	15
	All Interventions (*N* = 68)		**Interventions Reporting Fidelity (*N* = 53)**		**Interventions Not Reporting Fidelity (*N* = 15)**	
Variable	*N*	%	*N*	%	*N*	%
Implementer						
School Personnel	32	47	28	53	4	27
Researcher-Delivered	22	32	18	34	4	27
Computer-Delivered	6	9	2	4	4	27
Other Implementer	8	12	5	9	3	20
Group Size						
One-on-One	25	37	19	36	6	40
Small-Group	43	63	34	64	9	60
Sessions						
Less Than 100	44	65	34	64	10	67
100 or More	24	35	19	36	5	33

Table 6

Effect Sizes Associated with Fidelity Reporting

Variable	***k***	***g***	***SE***	**95% CI**	***p***	***df***	I^2	τ^2	***Q***	***p***
Fidelity									5.04	0.03
Reported	38	0.30	0.05	0.21, 0.39	< .01	37	45.16	0.03		
Not Reported	13	0.50	0.07	0.35, 0.64	< .01	12	14.03	0.01		

Elementary Educator Self-Efficacy and Knowledge to Teach Reading

Katlynn Dahl-Leonard, Colby Hall, Jamie DeCoster, and Emily J. Solari

University of Virginia

Katlynn Dahl-Leonard: awr8qt@virginia.edu; https://orcid.org/0000-0001-6124-5990

Colby Hall: csh4t@virginia.edu; https://orcid.org/0000-0002-0779-1322

Jamie DeCoster: jamied@virginia.edu; https://orcid.org/0000-0001-9751-6051

Emily J. Solari: ejs9ea@virginia.edu; https://orcid.org/0000-0002-8498-9834

The authors have no conflicts of interest to disclose.

This work was supported by the Harrison Family Foundation. The content is solely the responsibility of the authors and does not necessarily represent the official views of the Harrison Family Foundation.

Correspondence concerning this article should be addressed to Katlynn Dahl-Leonard at awr8qt@virginia.edu.

Abstract

The present study aimed to better understand teacher self-efficacy for teaching students with or at risk for reading difficulties and its relation to teacher knowledge to teach reading. We developed and administered surveys to 313 elementary-grade teachers. They rated their self-efficacy for providing reading instruction to students with or at risk for reading difficulties as moderately high. However, there was a statistically significant negative association of teacher knowledge to teach reading with self-efficacy for providing reading instruction to students with or at risk for reading difficulties. Implications for practice and directions for future research are discussed.

Keywords: teacher self-efficacy, teacher knowledge, reading instruction, elementary

Elementary Educator Self-Efficacy and Knowledge to Teach Reading

In the United States, a recent National Assessment of Educational Progress (NAEP) reading report indicated that 68% of fourth-grade students read below proficient levels (NAEP, 2022). Students who read below proficient levels are likely to have difficulties throughout their academic careers and are at an increased risk of dropping out of school (Boscardin et al., 2008; Daniel et al., 2006). They are also at greater risk for emotional issues, such as anxiety and depression, and behavioral disorders (Dahle & Knivsberg, 2014; Mugnaini et al., 2009; Willcutt et al., 2010). Access to evidence-based core and supplemental reading instruction within multi-tiered systems of support can reduce the incidence and severity of reading difficulties (RDs) experienced by students (Al Otaiba et al., 2009; VanDerHeydan et al., 2007). Thus, providing effective reading instruction is vital not only to facilitate students' development of proficient reading but also to support their overall academic performance and wellbeing.

Teacher knowledge to teach reading and self-efficacy for teaching reading are two key factors related to a teacher's ability to provide effective reading instruction. In particular, both are associated with instructional quality and student reading performance (Guo et al., 2012; Jordan & Bratsch-Hines, 2020; Podhajski et al., 2009; Varghese et al., 2016). The present study aims to better understand teacher self-efficacy for providing reading instruction to students with or at risk for RDs and its relation to teacher knowledge to teach reading.

Teacher Knowledge of Reading Development and Instruction

Decades of reading-related research yield broad consensus as to how teachers can facilitate the acquisition of reading skills (Castles et al., 2018). The simple view of reading (Gough & Tunmer, 1986) is an empirically validated framework positing that reading comprehension is the product of word recognition and language comprehension. Both

components comprise multiple subcomponents that contribute to skilled reading and reading comprehension. Word recognition depends on phonological awareness, phonics knowledge, orthographic knowledge, and decoding skill, whereas language comprehension depends on background and vocabulary knowledge, syntactical knowledge, text genre knowledge, and verbal reasoning skill (Scarborough, 2001). Research indicates that effective reading instruction addresses both components (Donegan & Wanzek, 2021; Foorman et al., 2016). Research also shows that effective reading instruction is explicit and systematic (Gersten et al., 2009; Vaughn et al., 2012; Swanson et al., 1999). In particular, it (a) orients students to the lesson objective, (b) sequences topics and activities so that they build on prior knowledge and provide the right level of challenge, (c) breaks down complex tasks into small, manageable steps, (d) includes teacher modeling with explanations, (e) allows frequent opportunities for students to respond and receive feedback, and (f) incorporates scaffolding that can be gradually reduced to support students' acquisition of new skills and knowledge.

Much research has investigated elementary teachers' knowledge of reading development and instruction. In particular, research has frequently examined associations of teachers' reading-related knowledge with their instructional practices (e.g., Cohen et al., 2017; Foorman & Moats, 2004; McCutchen et al., 2009). These studies reveal mixed findings. For example, Spear-Swerling and Zibulsky (2014) revealed that elementary teachers' knowledge of phonological awareness and phonics was positively related to the amount of instructional time allocated to instruction in phonological awareness and phonics, but their knowledge of vocabulary and comprehension was not correlated with the amount of time allocated to vocabulary or comprehension. Jordan and Bratsch-Hines (2020) demonstrated that Kindergarten and Grade 1 teachers' knowledge of teaching word reading was not significantly associated with instructional

variables such as discrete skills, print awareness, active learning, or collaborative learning. However, teacher knowledge was significantly associated with comprehensive instruction in that it predicted teacher use of 'elements of high-quality reading instruction documented as necessary for reading achievement' (p. 282). McCutchen et al. (2002) similarly revealed that Kindergarten and Grade 1 teachers' knowledge about teaching reading was associated with increased use of explicit instruction during phonological awareness and comprehension instruction. However, Piasta et al. (2009) determined that Grade 1 teachers' knowledge of literacy concepts was not associated with the amount of explicit decoding instruction they provided.

Previous research has also demonstrated mixed associations between teacher reading-related knowledge and student reading outcomes. Several empirical studies revealed positive associations between teacher knowledge and student reading-related outcomes (e.g., Al Otaiba et al., 2012; Foorman & Moats, 2004; McCutchen et al., 2009; Spear-Swerling & Brucker, 2004). For example, Podhajski et al. (2009) showed that teacher knowledge was positively associated with gains in phoneme segmentation, letter naming, nonsense word reading, and oral reading fluency among Grade 1 students. Porter et al. (2023) revealed that kindergarten and Grade 1 teachers' knowledge of language and literacy reliably predicted students' foundational reading skills (i.e., print awareness, phonological awareness, phonics/decoding) scores, but not their reading comprehension scores. Other studies have demonstrated only small effects of teacher reading-related knowledge on student literacy skills (Carlisle et al., 2011; Piasta et al., 2009). Taken together, this work signals the need for further research identifying factors interacting with knowledge to produce effective teacher instructional practices and improvements in student reading outcomes.

Teacher Self-Efficacy for Providing Reading Instruction

Bandura's Social Cognitive Theory (1977; 1986; 1997) defines self-efficacy as an assessment of one's capabilities to attain a desired level of performance. Bandura posited that one's behaviors are greatly influenced by their perceived capabilities, or self-efficacy, to perform those behaviors. In fact, he theorised that self-efficacy beliefs, rather than one's actual abilities, have the most powerful impact on behavior. Bandura further speculated that a cyclical relationship between self-efficacy beliefs and behavior exists (i.e., the more confident one is in their abilities, the more likely they are to succeed, which provides them with experiences that further enhance their self-efficacy).

In the teaching context, self-efficacy is the extent to which teachers believe they can influence a student's performance through their instruction (Khan et al., 2015; Tschannen-Moran & Johnson, 2011). Bandura's Social Cognitive Theory (1977; 1986; 1997) suggests that to provide effective instruction, teachers need to believe they have the necessary knowledge and skills to impact student performance. When a teacher believes that they can provide effective instruction, it increases their ability to engage in behaviors that align with that goal. In alignment with this theory, the Rand Corporation conducted several studies in the 1970s that focused on reading instruction in search of variables that would explain differences in teacher effectiveness (Armor et al., 1976). They discovered that teacher self-efficacy was positively related to reading achievement among students. More specifically, students taught by teachers who believed they could impact students' learning had higher reading achievement than students whose teachers believed there was little they could do considering the environmental barriers to learning.

More recent empirical research has also demonstrated positive associations of teachers' self-efficacy for teaching reading with the quality of reading instruction they provide and the reading achievement of students in their classrooms (Guo et al., 2012; Maloch et al., 2003;

Varghese et al., 2016). For example, Guo et al. (2012) examined the effects of teacher self-efficacy, education, and years of experience on observed classroom practices as they related to Grade 5 students' literacy skills. They revealed that teachers with a higher sense of self-efficacy supported students more (e.g., via scaffolding and feedback) and provided a more positive classroom environment than teachers with lower self-efficacy. Further, teacher self-efficacy was positively related to student literacy outcomes.

Limited research has examined teacher feelings of self-efficacy focused on providing reading instruction to students with or at risks for RDs. In the research that does exist, teachers often do not feel prepared to instruct such students. For example, Bos et al. (2001) measured elementary teachers' perceived preparedness to teach reading on a scale of 1 (not prepared) to 4 (well prepared). They demonstrated that although teachers' perceived preparedness to teach reading to typically developing students fell in the moderate range (M = 2.6), they felt less prepared to teach students with RDs (M = 2.3). Similarly, Washburn et al. (2011) measured elementary teachers' perceived teaching ability on a scale of 1 (minimal) to 4 (expert). They reported that teachers perceived their abilities to teach reading to typically developing students as good (M = 2.79 out of 4), but slightly less so for teaching students with RDs (M = 2.64).

Given the paucity of research on this topic, it is unsurprising that a valid and reliable measure assessing teachers' self-efficacy for teaching reading to students with or at risk for RDs has not emerged. A commonly used measure of teacher self-efficacy for providing literacy instruction in general is the 22-item Teachers' Sense of Efficacy for Literacy Instruction (TSELI) scale, which has demonstrated a Cronbach's alpha reliability of .96 (Tschannen-Moran & Johnson, 2011). However, only one item on the TSELI addresses providing instruction for students with or at risk for RDs (i.e., 'How much can you do to meet the needs of struggling

readers?'). Therefore, there is a need to develop a valid and reliable scale intended to measure teachers' self-efficacy for teaching reading to students with or at risk for RDs.

Relation Between Teacher Knowledge and Self-Efficacy

Although knowledge to teach reading and self-efficacy to teach reading are both associated with instructional quality and student reading performance (Guo et al., 2012; Jordan & Bratsch-Hines, 2020; Podhajski et al., 2009; Varghese et al., 2016), research on the relation between knowledge and self-efficacy is scarce, and the little research that exists has frequently yielded null results. For example, Sharp et al. (2016) found that prospective elementary teachers' confidence in teaching specific literacy skills was not significantly related to knowledge required to teach those skills. Stark et al. (2016) similarly found no relationship between teacher self-rated ability to teach literacy and their literacy knowledge. Additionally, Ciampa and Gallagher (2018) found no relation between pre-service teachers' self-efficacy and their knowledge to provide literacy instruction at the end of an elementary literacy methods course.

Present Study

Teacher knowledge to teach reading and teacher self-efficacy for teaching reading have both been associated with improved instructional quality and student reading performance (Guo et al., 2012; Jordan & Bratsch-Hines, 2020; Podhajski et al., 2009; Varghese et al., 2016). However, there is little research on the relation of teacher knowledge with self-efficacy (Ciampa & Gallagher, 2018; Sharp et al., 2016; Stark et al., 2016). Further, previous research has focused on investigating teachers' self-efficacy for providing reading instruction to all students (Guo et al., 2012; Varghese et al., 2016), with limited research exploring teachers' self-efficacy for teaching reading to students with or at risk for RDs (Bos et al., 2001; Washburn et al., 2011). Thus, the present study developed a measure to better understand teacher self-efficacy for

teaching students with or at risk for RDs and examined its relation with teacher knowledge to teach reading. Our research questions were:

1) What is the factor structure and reliability of the self-efficacy scale?
2) What levels of self-efficacy are demonstrated by elementary teachers in the current sample?
3) To what extent is self-efficacy associated with knowledge to teach reading?

Method

Study Context

Our research team recently reported on the development and validation of a measure for assessing elementary teachers' knowledge to teach reading (Hall et al., 2023). In addition to the knowledge measure, our research team also developed and administered a self-efficacy scale for teaching students with or at risk for RDs. However, the data associated with the self-efficacy scale were not published in the Hall et al. (2023) manuscript. Thus, the present study focused on assessing the factor structure and reliability of the self-efficacy scale as well as determining the levels of self-efficacy demonstrated by elementary teachers. To answer our third research question (i.e., to evaluate the extent to which self-efficacy is associated with knowledge to teach reading), we drew on the knowledge survey data used to conduct analyses in the previously-reported study (Hall et al., 2023). The Institutional Review Board at the research team's university approved the study procedures.

Participants and Data Collection Procedures

The participants and data collection procedures for the present study were identical to those described in the Hall et al. (2023) publication. That said, in this section we highlight some key information about these procedures. Participants were recruited via Qualtrics Panel, a private research software company specializing in web-based data collection. Eligible participants were

(a) residing in the United States, (b) working as professional educators of K-5 students, and (c) teaching English language arts, literacy, or reading. Participants who matched the eligibility criteria and consented to participate in the research study were asked to complete the survey online. Those who completed the survey received approximately $5 or the equivalent in redeemable points directly from Qualtrics. Qualtrics established security checks for bots and ensured that participants were only able to take the survey one time. Table 1 displays demographic information for the final sample of 313 participants.

Participants completed the online survey (including the knowledge survey and the self-efficacy scale) during the summer of 2022. The knowledge survey, called the Teacher Understanding of Literacy Constructs and Evidence-Based Instructional Practices (TULIP) survey, consists of 55 multiple-choice items measuring teacher content knowledge within the domains of (a) phonological awareness (12 items), (b) phonics, decoding, and encoding (15 items), (c) reading fluency (7 items), (d) oral language (9 items), and (e) reading comprehension (12 items). Note that Hall et al. (2023) describes in detail the development and validation of the TULIP survey.

Participants also responded to six self-efficacy items on a five-point Likert-scale for which higher values indicated a greater degree of perceived self-efficacy. The self-efficacy items were developed in conjunction with the TULIP survey items and underwent an iterative development process that included expert review (as described in Hall et al., 2023). The final self-efficacy items focused on teachers' perceived levels of preparedness, confidence, and ability for teaching students with or at risk for RDs. Figure 1 shows the self-efficacy items as they were presented to the participants.

All survey items were presented in English. Participants were required to choose at least

one answer option for each item before moving to the next item but could end the survey at any time. A total of 313 participants provided complete data for our analyses. The average time to complete the entire survey was 19 minutes.

Data Analysis

We performed confirmatory factor analyses (CFAs) to understand the constructs that underlay responses to the self-efficacy scale. We believed that responses would be organised around two domains: beliefs about past preparation (items 2 and 3) and beliefs about current preparedness, confidence, and ability (items 1, 4, 5, and 6). To determine the appropriateness of this model, we compared the fit of a one-factor model in which all items loaded on a single overall factor to a two-factor model in which the items were organised around the proposed domains. The latent domain scores were allowed to co-vary in the two-factor CFA. We also examined the item correlations to better understand the dimensionality of the scale and performed reliability analyses to understand the internal consistency of the scale. To describe the self-efficacy of elementary teachers in our sample, we performed descriptive analyses of the self-efficacy scale (e.g., calculating teachers average self-efficacy rating). To examine whether self-efficacy was associated with knowledge, we used regression models predicting teacher knowledge from teacher self-efficacy. CFAs were conducted using Stata version 17; all other analyses were conducted using SPSS version 28.

Results

Research Question 1: Factor Structure and Reliability

We performed CFAs to compare the fit of a one-factor model to a two-factor model. The one-factor CFA (chi-square [9] = 144.76, $p < 0.001$, RMSEA = 0.22, CFI = 0.85, TLI = 0.74, SRMR = 0.07) showed evidence of unsatisfactory fit, whereas the two-factor CFA (chi-

square [8] = 28.56, $p < 0.001$, RMSEA = 0.09, CFI = 0.98, TLI = 0.96, SRMR = 0.03) showed evidence of acceptable fit. A model comparison test indicated the two-factor model fit the data better than the one-factor model (chi-square [1] = 116.23, $p < 0.001$). Table 2 presents the reliability of the full scale and the two subscales as estimated by Cronbach's alpha and McDonald's omega. The full scale and the two subscales had good reliability.

Table 3 represents the correlations among the items. All between-item correlations would be described as moderate to large (range: 0.36-0.77) and were significant ($p < .01$). Analysis of the discrimination of the items revealed that most items had good discrimination in terms of their associations with the overall ratings: 83% of the items had item-total correlations greater than 0.70. The average item-total correlation was 0.77.

Research Question 2: Ratings of Self-Efficacy

Table 4 provides descriptive statistics for the teacher responses to the self-efficacy items. The full scale average self-efficacy rating was 3.86 (*SD* = 0.79) out of 5. The average ratings for the Beliefs about Past Preparation and Beliefs about Current Ability factors were 3.52 (*SD* = 1.13) and 4.03 (*SD* = 0.75), respectively. The average ratings for each of the six items ranged from 3.46 (*SD* = 1.24) to 4.19 (*SD* = 0.84).

Research Question 3: Association with Knowledge to Teach Reading

Teachers answered an average of 27.79 out of 55 (50.52%) of items correctly on the knowledge survey (see Hall et al., 2023 for more information about knowledge survey results). Regression analyses were conducted to determine the degree to which self-efficacy may have contributed to knowledge. For the full scale, results indicated that 2% of the total variation in knowledge score was explained by overall self-efficacy rating ($R^2 = 0.02$) and that the full self-efficacy scale had a small, statistically significant negative relation with knowledge ($r = -.13$, p

= .02). The Beliefs about Current Ability factor also had a statistically significant negative relation with knowledge (r = -.29, p < .001), whereas the Beliefs about Current Ability factor did not have a statistically significant relation with knowledge (r = .01, p = .88).

Given that Hall et al. (2023) found that teacher knowledge was significantly related to education level (i.e., teachers with more education had higher knowledge scores), we added teachers' highest level of education (coded as [a] high school degree/GED/associate degree, [b] bachelor's degree, or [c] master's degree or beyond) as a covariate in our regression models. Results indicated a similar pattern of findings. The full scale (r = -.13, p = .02) and Beliefs about Current Ability factor (r = -.27, p < .001) maintained statistically significant negative relations with knowledge, whereas the relation between the Beliefs about Current Ability factor and knowledge remained non-significant.

Discussion

Extant research shows that both teacher knowledge to teach reading and teacher self-efficacy for teaching reading are associated with improved instructional quality and student reading performance (Guo et al., 2012; Jordan & Bratsch-Hines, 2020; Podhajski et al., 2009; Varghese et al., 2016). However, there is not much prior research on the relation between teacher knowledge and self-efficacy, and the research that does exist has often revealed nonsignificant relations (Ciampa & Gallagher, 2018; Sharp et al., 2016; Stark et al., 2016). It is also noteworthy that research in this area has mostly focused on investigating teachers' self-efficacy for providing reading instruction to all students (Guo et al., 2012; Varghese et al., 2016). The purpose of the present study was to develop a measure to evaluate elementary teachers' self-efficacy for teaching reading to students with or at risk for RDs. Further, this study aimed to explore the relation between teacher self-efficacy for teaching students with or at risk for RDs and teacher

knowledge to teach reading.

Self-Efficacy Scale Validity and Reliability

The factor analysis provided preliminary evidence that the current scale may serve as a valid two-factor measure of teacher self-efficacy for providing reading instruction to students with or at risk for RDs. That said, further validity evidence is needed (i.e., an assessment of convergent validity with similar measures of teacher self-efficacy for teaching reading). The full scale and both factors were shown to reliably assess self-efficacy for teaching students with or at risk for RDs. Therefore, the current scale shows promise as a valid and reliable measure of elementary teacher self-efficacy for teaching such students.

Levels of Self-Efficacy for Teaching Students with or at Risk for RDs

The average self-efficacy rating for the elementary teachers in our sample was 3.86 out of 5. In other words, teachers rated their self-efficacy for providing reading instruction to students with or at risk for RDs as moderately high (i.e., they mostly agreed with the statements on the scale). This finding is somewhat inconsistent with previous survey studies that examined teachers' perceptions about teaching reading to students with RDs. For example, Bos et al. (2001) revealed that elementary teachers only felt somewhat prepared to teach students with RDs (M = 2.3 out of 4). Similarly, Washburn et al. (2011) determined that elementary teachers' perceived abilities to teach students with RDs was moderate (M = 2.64 out of 4). Notably, there appeared to be an increase in teacher self-efficacy ratings from Bos et al.'s 2001 study to Washburn et al.'s 2011 study; this increasing trend appears to have continued between Washburn's 2011 study and the present study conducted in 2022. Further, our finding that teachers rated their self-efficacy for teaching students with or at risk for RDs as moderate to high on average was consistent with a recent study that utilised teacher interviews instead of surveys.

Dahl-Leonard et al. (2023) interviewed kindergarten teachers about their self-efficacy in providing reading instruction to students with RDs. Teachers in their sample described moderate to high levels of self-efficacy in teaching reading to students with RDs, with only one teacher sharing feelings of low self-efficacy. The authors also noted that teachers often described feeling more confident teaching reading to students with RDs than their typically developing peers.

Changes in teacher preparation and professional development may explain this recent increase in teacher self-efficacy for providing reading instruction to students with or at risk for RDs. More than ever before, teachers are now expected to meet the individual needs of diverse learners (Whitaker & Valtierra, 2018). As such, teacher educators are responsible for preparing future teachers to teach diverse groups of students in their classrooms (Theoharis & Causton-Theoharis, 2011). Recent standards for teachers of reading outlined by the International Literacy Association (ILA; ILA, 2017) and other international, national, and state-level organizations emphasize the importance of understanding reading development and knowledge of evidence-based instructional methods that support learning for students with RDs. Research shows that self-efficacy related to providing reading instruction typically increases after pre-service teachers complete reading methods coursework and participate in reading-related field experiences (Helfrich & Clark, 2016). Additionally, in the last decade, numerous states have passed laws or implemented policies requiring increased professional development for teachers of students with learning difficulties (Gearin et al., 2022). Thus, it is likely that increased coursework and field experiences as well as professional learning opportunities focused on diverse learners have resulted in increased self-efficacy for teaching students with or at risk for RDs (Al-Otaiba et al., 2012; Leader-Janssen & Rankin-Erickson, 2013).

Relation Between Teacher Knowledge and Self-Efficacy

Our study revealed that teacher knowledge to teach reading was negatively related to self-efficacy for providing reading instruction to students with or at risk for RDs. That is, for teachers in our sample, the more self-efficacious teachers felt to teach students with or at risk for RDs, the less knowledgeable for providing reading instruction they actually were. Notably, Tschannen-Moran and Johnson (2011) speculated that 'slightly overestimating one's actual abilities may be useful if it leads to greater effort, persistence, and resilience, and because it is difficult for a person to invest fully in an activity while fighting self-doubt' (p. 753). Therefore, some overestimation of abilities may actually be beneficial to lower-knowledge teachers. For example, these teachers may believe their teaching matters for students' reading development, accept responsibility for students' learning, and make instructional adjustments when students are not meeting expected goals. However, it has also been argued that 'an overestimation of one's skills may lead to complacency and a failure to pursue professional development opportunities to overcome deficiencies' (Tschannen-Moran & Johnson, 2011, p. 753).

There was a statistically significant negative relation between the Beliefs about Past Preparation factor and knowledge. The two items in the Beliefs about Past Preparation factor focused on teacher education coursework and student teaching experiences. One item stated, 'My teacher education coursework provided me with the knowledge I need to effectively teach students with or at risk for RDs' and the other item stated, 'My teacher education coursework/student teaching provided me with the hands-on practice I need to effectively teach students with or at risk for RDs.' Thus, the more prepared teachers felt by their teacher preparation coursework and experiences to teach students with or at risk for RDs, the less knowledgeable for providing reading instruction they actually were. Conversely, more knowledgeable teachers tended to perceive their teacher preparation coursework and experiences

as having been inadequate. We posit that this may be because the teachers who had experienced the most knowledge-building professional development opportunities on the job (and had thus accumulated the largest amounts of up-to-date knowledge of the science of reading) were the ones who were most aware of shortfalls in their own teacher preparation coursework and experiences.

Limitations

A few limitations of this study should be noted. First, we only assessed the content validity (via expert feedback) and the construct validity (via factor analysis) of the current self-efficacy scale. As a result, we do not know whether the scale demonstrates convergent validity with similar measures of teacher self-efficacy for teaching reading, or whether it predicts classroom reading instruction or reading outcomes for students with or at risk for RDs. Also, the current scale does not address all aspects of teacher self-efficacy. For example, Tschannen-Moran and Woolfolk-Hoy (2001) conceptualised teacher self-efficacy as having three dimensions: instructional strategies, classroom management, and student engagement. The current self-efficacy scale only included items that would be categorised into the instructional strategies dimension. Additionally, quantitative measures of self-efficacy have received criticism for not fully capturing the complexities of teacher self-efficacy; researchers have called for the integration of multiple sources of data, including qualitative data, when exploring teacher self-efficacy (Glackin & Hohenstein, 2018). Further, research suggests that five-point Likert scales can sway respondents to restrict their preferences and avoid extreme statements, such as strongly agree and strongly disagree (Cohen et al., 2007). Lastly, it is important to note that the knowledge survey used in the present study did not measure every aspect of knowledge to teach reading (e.g., it focused on content knowledge rather than instructional methods).

Implications and Directions for Future Research

Our findings revealed that the current scale can reliably assess elementary teachers' self-efficacy for providing reading instruction to students with or at risk for RDs. In practice, teacher educators and school administrators may use the current scale as a screening tool to measure elementary teachers' self-efficacy for teaching reading to such students. Our findings also indicated a negative relation between teacher self-efficacy for teaching students with or at risk for RDs and teacher knowledge to teach reading. This may be problematic because teachers overestimating their abilities are less likely to seek out or be receptive of professional learning opportunities. Teachers able to accurately discriminate between what they know and do not know are likely to allocate greater attention to areas in which they truly lack knowledge (Bostock & Boon, 2012; Ciampa & Gallagher, 2018). It may be beneficial for school administrators to measure both teacher self-efficacy and knowledge to determine professional learning opportunities that are appropriate for individual teachers. For example, when providing professional development to a teacher with high self-efficacy and low knowledge, there may be a need to address misconceptions about effective reading instruction.

Although the current scale shows promise as a valid and reliable measure of elementary teacher self-efficacy for teaching reading to students with or at risk for RDs, there is value in conducting future research to determine whether the scale demonstrates convergent validity with similar measures of teacher self-efficacy for providing literacy instruction, such as the TSELI (Tschannen-Moran & Johnson, 2011). It is also important to note that, even if the current scale is a valid and reliable measure of teacher self-efficacy for teaching reading to students with or at risk for RDs, it should be viewed as only one source of data. We echo other researchers in the field who acknowledge that 'it is essential that traditional quantitative approaches are better

triangulated and integrated with other sources of data' (Glackin & Hohenstein, 2018, p. 271). We therefore encourage researchers to incorporate qualitative measures, such as interviews, into studies of teacher self-efficacy.

Unlike prior research, the current study demonstrated a statistically significant negative relation between self-efficacy and knowledge to teach reading. We would therefore encourage researchers to further explore the associations of self-efficacy with knowledge, as well as factors that may be causing this relation. For example, some research suggests that teacher knowledge may need to research a certain threshold before it can be implemented in practice (Piasta et al., 2020). It follows that there may be a level of knowledge at which teachers are aware of what they know and are able to more accurately assess their abilities, aligning their knowledge and self-efficacy. Lastly, given that prior research has focused on examining teachers' self-efficacy for providing reading instruction to all students, it may be worthwhile to further examine the impact of elementary teacher self-efficacy for providing reading instruction to students with or at risk for RDs on teaching practices and student outcomes.

Conclusion

The present study aimed to better understand teacher self-efficacy for teaching students with or at risk for RDs and its relation to teacher knowledge to teach reading. The current scale shows promise as a valid and reliable measure of elementary teacher self-efficacy for teaching reading to students with or at risk for RDs. Thus, teacher educators and school administrators may use the current scale as a screening tool to measure elementary teachers' self-efficacy for teaching reading to such students. Teachers in our sample rated their self-efficacy as moderately high, but there was a statistically significant negative association of teacher knowledge to teach reading with self-efficacy. This finding demonstrates a need for future research on this topic.

References

Al Otaiba, S., Connor, C. M., Foorman, B., Schatschneider, C., Greulich, L., & Sidler, J. F. (2009). Identifying and intervening with beginning readers who are at-risk for dyslexia: Advances in individualized classroom instruction. *Perspectives on Language and Literacy, 35*(4), 13-19. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4296731/

Al Otaiba, S., Lake, V. E., Greulich, L., Folsom, J. S., & Guidry, L. (2012). Preparing beginning reading teachers: An experimental comparison of initial early literacy field experiences. *Reading and Writing, 25*(1), 109-129. https://doi.org/10.1007/s11145-010-9250-2

Armor, D., Conroy-Oseguera, P., Cox, M., King, N., McDonell, L., Pascal, A., Pauly, E., & Zellman, G. (1976). *Analysis of the school preferred reading programs in selected Los Angeles minority schools.* Report No. R2007-LAUSD. Rand Corporation. https://files.eric.ed.gov/fulltext/ED130243.pdf

Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review, 84*(2), 191–215. https://doi.org/10.1037/0033-295X.84.2.191

Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory. Prentice-Hall series in social learning theory.* Prentice-Hall.

Bandura, A. (1997). *Self-Efficacy: The exercise of control.* W.H. Freeman and Company.

Bos, C., Mather, N., Dickson, S., Podhajski, B., & Chard, D. (2001). Perceptions and knowledge of preservice and inservice educators about early reading instruction. *Annals of Dyslexia, 51*(1), 97-120. https://doi.org/10.1007/s11881-001-0007-0

Boscardin, C., Muthén, B., Francis, D., & Baker, E. (2008). Early identification of reading difficulties using heterogeneous developmental trajectories. *Journal of Educational Psychology, 100(*1), 192–208. https://doi.org/10.1037/0022-0663.100.1.192

Bostock, L. & Boon, H. (2012). Pre-service teachers' literacy self-efficacy and literacy competence. *Australian and International Journal of Rural Education, 22*(1), 19–37.

Carlisle, J. F., Kelcey, B., Rowan, B., & Phelps, G. (2011). Teachers' knowledge about early reading: Effects on students' gains in reading achievement. *Journal of Research on Educational Effectiveness, 4*(4), 289-321. https://doi.org/10.1080/19345747.2010.539297

Castles, A., Rastle, K., & Nation, K. (2018). Ending the reading wars: Reading acquisition from novice to expert. *Psychological Science in the Public Interest, 19*(1), 5-51. https://doi.org/10.1177/1529100618772271

Ciampa, K., & Gallagher, T. L. (2018). A comparative examination of Canadian and American pre-service teachers' self-efficacy beliefs for literacy instruction. *Reading and Writing, 31*(2), 457-481. https://doi.org/10.1007/s11145-017-9793-6

Cohen, L., L. Manion, & K. Morrison. (2007). *Research Methods in Education.* Routledge.

Cohen, R. A., Mather, N., Schneider, D. A., & White, J. M. (2017). A comparison of schools: Teacher knowledge of explicit code-based reading instruction. *Reading and Writing, 30*(4), 653-690. https://doi.org/10.1007/s11145-016-9694-0

Dahl-Leonard, K., Mesa, M. P., Hall, C., & Zucker, T. A. (2023). Understanding kindergarten teacher self-efficacy for providing reading instruction to students with reading difficulties. *Teachers and Teaching: Theory and Practice.* Advance online publication. https://doi.org/10.1080/13540602.2023.2201424

Dahle, A. E., & Knivsberg, A.-M. (2014). Internalizing, externalizing and attention problems in dyslexia. *Scandinavian Journal of Disability Research, 16*(2), 179–193. https://doi.org/10.1080/15017419.2013.781953

Daniel, S. S., Walsh, A. K., Goldston, D. B., Arnold, E. M., Reboussin, B. A., & Wood, F. B.

(2006). Suicidality, school dropout, and reading problems among adolescents. *Journal of Learning Disabilities, 39*(6), 507–514. https://doi.org/10.1177/00222194060390060301

Donegan, R. E., & Wanzek, J. (2021). Effects of reading interventions implemented for upper elementary struggling readers: A look at recent research. *Reading and Writing, 34*(8), 1943-1977. https://doi.org/10.1007/s11145-021-10123-y

Foorman, B. R., & Moats, L. C. (2004). Conditions for sustaining research-based practices in early reading instruction. *Remedial and Special Education, 25*(1), 51-60. https://doi.org/10.1177/07419325040250010601

Foorman, B., Beyler, N., Borradaile, K., Coyne, M., Denton, C. A., Dimino, J., Furgeson, J., Hayes, L., Henke, J., Justice, L., Keating, B., Lewis, W., Sattar, S., Streke, A., Wagner, R., & Wissel, S. (2016). *Foundational skills to support reading for understanding in kindergarten through 3rd grade* (NCEE 2016-4008). National Center for Education Evaluation and Regional Assistance (NCEE), Institute of Education Sciences, U.S. Department of Education. https://ies.ed.gov/ncee/wwc/practiceguide/21

Gearin, B., Petscher, Y., Stanley, C., Nelson, N. J., & Fien, H. (2022). Document analysis of state dyslexia legislation suggests likely heterogeneous effects on student and school outcomes. *Learning Disability Quarterly, 45*(4), 267-279. https://doi.org/10.1177/0731948721991549

Gersten, R., Compton, D., Connor, C.M., Dimino, J., Santoro, L., Linan-Thompson, S., and Tilly, W.D. (2009). *Assisting students struggling with reading: Response to Intervention and multi-tier intervention for reading in the primary grades. A practice guide.* (NCEE 2009-4045). National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

https://ies.ed.gov/ncee/wwc/practiceguide/3

Glackin, M., & Hohenstein, J. (2018). Teachers' self-efficacy: Progressing qualitative analysis. *International Journal of Research & Method in Education, 41*(3), 271-290. https://doi.org/10.1080/1743727X.2017.1295940

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

Guo, Y., Connor, C. M., Yang, Y., Roehrig, A. D., & Morrison, F. J. (2012). The effects of teacher qualification, teacher self-efficacy, and classroom practices on fifth graders' literacy outcomes. *The Elementary School Journal*, *113*(1), 3-24. https://doi.org/10.1086/665816

Hall, C., Solari, E. J., Hayes, L., Dahl-Leonard, K., DeCoster, J., Kehoe, K. F., Conner, C., Henry, A. R., Demchak, A., Richmond, C. L., & Vargas, I. (2023). Validation of an instrument for assessing elementary-grade educators' knowledge to teach reading. *Reading and Writing.* Advance online publication. https://doi.org/10.1007/s11145-023-10456-w

Helfrich, S. R., & Clark, S. K. (2016). A comparative examination of pre-service teacher self-efficacy related to literacy instruction. *Reading Psychology*, *37*(7), 943-961. https://doi.org/10.1080/02702711.2015.1133466

International Literacy Association. (2017). *Standards for the Preparation of Literacy Professionals.* https://www.literacyworldwide.org/get-resources/standards/standards-2017

Jordan, R. L., & Bratsch-Hines, M. (2020). Associations of reading knowledge with kindergarten and first grade teachers' reported instructional strategies. *Literacy Research and*

Instruction, 59(4), 277-297. https://doi.org/10.1080/19388071.2020.1774689

Khan, A., Fleva, E., & Qazi, T. (2015). Role of self-esteem and general self-efficacy in teachers' efficacy in primary schools. *Psychology, 6*(1), 117. https://doi.org/10.4236/psych.2015.61010

Leader-Janssen, E. M., & Rankin-Erickson, J. L. (2013). Preservice teachers' content knowledge and self-efficacy for teaching reading. *Literacy Research and Instruction, 52*(3), 204–229. https://doi.org/10.1080/19388071.2013.781253

Maloch, B., Flint, A., Eldridge, D., Harmon, J., Loven, R., Fine, J., et al. (2003). Understanding, beliefs, and reported decision making of first-year teachers from different reading teacher preparation programs. *The Elementary School Journal, 103*(5), 431-459. https://doi.org/10.1086/499734

McCutchen, D., Abbott, R. D., Green, L. B., Beretvas, S. N., Cox, S., Potter, N. S., Quiroga, T., & Gray, A. L. (2002). Beginning literacy: Links among teacher knowledge, teacher practice, and student learning. *Journal of Learning Disabilities, 35*(1), 69-86. https://doi.org/10.1177/002221940203500106

McCutchen, D., Green, L., Abbott, R. D., & Sanders, E. A. (2009). Further evidence for teacher knowledge: Supporting struggling readers in grades three through five. *Reading and Writing, 22*(4), 401-423. https://doi.org/10.1007/s11145-009-9163-0

Mugnaini, D., Lassi, S., La Malfa, G., & Albertini, G. (2009). Internalizing correlates of dyslexia. *World Journal of Pediatrics, 5*(4), 255–264. https://doi.org/10.1007/s12519-009-0049-7

National Assessment of Education Progress. (2022). NAEP Reading 2022 Highlights. The Nation's Report Card. https://www.nationsreportcard.gov/highlights/reading/202/.

Piasta, S. B., Connor, C. M., Fishman, B. J., & Morrison, F. J. (2009). Teachers' knowledge of literacy concepts, classroom practices, and student reading growth. *Scientific Studies of Reading*, *13*(3), 224-248. https://doi.org/10.1080/10888430902851364

Piasta, S. B., Soto Ramirez, P., Farley, K. S., Justice, L. M., & Park, S. (2020). Exploring the nature of associations between educators' knowledge and their emergent literacy classroom practices. *Reading and Writing*, *33*(6), 1399-1422. https://doi.org/10.1007/s11145-019-10013-4

Podhajski, B., Mather, N., Nathan, J., & Sammons, J. (2009). Professional development in scientifically-based reading instruction: Teacher knowledge and reading outcomes. *Journal of Learning Disabilities, 42*(5), 403-417. https://doi.org/10.1177/0022219409338737

Porter, S. B., Odegard, T. N., Farris, E. A., & Oslund, E. L. (2023). Effects of teacher knowledge of early reading on students' gains in reading foundational skills and comprehension. *Reading and Writing.* Advance online publication. https://doi.org/10.1007/s11145-023-10448-w

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

Sharp, A. C., Brandt, L., Tuft, E. A., & Jay, S. (2016). Relationship of self-efficacy and teacher knowledge for prospective elementary education teachers. *Universal Journal of Educational Research*, *4*(10), 2432-2439. https://doi.org/10.13189/ujer.2016.041022

Spear-Swerling, L., & Brucker, P. O. (2004). Preparing novice teachers to develop basic reading and spelling skills in children. *Annals of Dyslexia*, *54*(2), 332-364.

https://doi.org/10.1007/s11881-004-0016-x

Spear-Swerling, L., & Zibulsky, J. (2014). Making time for literacy: Teacher knowledge and time allocation in instructional planning. *Reading and Writing, 27*(8), 1353-1378. https://doi.org/10.1007/s11145-013-9491-y

Stark, H. L., Snow, P. C., Eadie, P. A., & Goldfeld, S. R. (2016). Language and reading instruction in early years' classrooms: The knowledge and self-rated ability of Australian teachers. *Annals of Dyslexia, 66*(1), 28-54. https://doi.org/10.1007/s11881-015-0112-0

Swanson, H. L., Hoskyn, M., & Lee, C. (1999). *Interventions for students with learning disabilities: A meta-analysis of treatment outcomes.* Guilford Press.

Theoharis, G., & Causton-Theoharis, J. (2011). Preparing pre-service teachers for inclusive classrooms: Revising lesson-planning expectations. *International Journal of Inclusive Education, 15*(7), 743–761. https://doi.org/10.1080/13603110903350321

Timperley, H. S., & Phillips, G. (2003). Changing and sustaining teachers' expectations through professional development in literacy. *Teaching and Teacher Education, 19*(6), 627-641. https://doi.org/10.1016/S0742-051X(03)00058-1

Tschannen-Moran, M., & Hoy, A. W. (2001). Teacher efficacy: Capturing an elusive construct. *Teaching and Teacher Education, 17*(7), 783-805. https://doi.org/10.1016/S0742-051X(01)00036-1

Tschannen-Moran, M., & Johnson, D. (2011). Exploring literacy teachers' self-efficacy beliefs: Potential sources at play. *Teaching and Teacher Education, 27*(4), 751-761. https://doi.org/10.1016/j.tate.2010.12.005

VanDerHeyden, A. M., Witt, J. C., & Gilbertson, D. (2007). A multi-year evaluation of the effects of a Response to Intervention (RTI) model on identification of children for special

education. *Journal of School Psychology, 45*(2), 225–256. https://doi.org/10.1016/j.jsp.2006.11.004

Varghese, C., Garwood, J. D., Bratsch-Hines, M., & Vernon-Feagans, L. (2016). Exploring magnitude of change in teacher efficacy and implications for students' literacy growth. *Teaching and Teacher Education, 55*(1), 228-239. https://doi.org/10.1016/j.tate.2016.01.011

Vaughn, S., Wanzek, J., Murray, C. S., & Roberts, G. (2012). *Intensive interventions for students struggling in reading and mathematics: A practice guide.* Center on Instruction. https://files.eric.ed.gov/fulltext/ED531907.pdf

Washburn, E. K., Joshi, R. M., & Binks-Cantrell, E. S. (2011). Teacher knowledge of basic language concepts and dyslexia. *Dyslexia, 17*(2), 165-183. https://doi.org/10.1002/dys.426

Whitaker, M.C. & Valtierra, K.M. (2018). Enhancing preservice teachers' motivation to teach diverse learners. *Teaching and Teacher Education, 73*(1), 171–182. https://doi.org/10.1016/j.tate.2018.04.004

Willcutt, E. G., Betjemann, R. S., McGrath, L. M., Chhabildas, N. A., Olson, R. K., DeFries, J. C., & Pennington, B. F. (2010). Etiology and neuropsychology of comorbidity between RD and ADHD: The case for multiple-deficit models. *Cortex, 46*(10), 1345–1361. https://doi.org/10.1016/j.cortex.2010.06.009

Figure 1

Self-Efficacy Items

For the below items, select the number from 1-5 (where 1 is “strongly disagree” and 5 is “strongly agree”) that best describes how you feel.

	1 (strongly disagree)	2	3 (neutral)	4	5 (strongly agree)
I feel prepared to teach my students with or at risk for reading difficulties.	○	○	○	○	○
My teacher education coursework provided me with the knowledge I need to effectively teach students with or at risk for reading difficulties.	○	○	○	○	○
My teacher education coursework/student teaching provided me with the hands-on practice I need to effectively teach students with or at risk for reading difficulties.	○	○	○	○	○
I feel confident that I can help my students with reading difficulties become better readers.	○	○	○	○	○
Most of my students with reading difficulties can learn what I teach them.	○	○	○	○	○
By trying different methods, I can help my students with reading difficulties become better readers.	○	○	○	○	○

Table 1

Teacher Demographic Information

	%
Gender	
Female	80
Male	19
Non-Binary/Non-Conforming	1
Race	
American Indian or Alaska Native	1
Asian	3
Black or African American	10
Multi-Racial	4
Native Hawaiian or Other Pacific Islander	<1
White	80
Other	2
Hispanic	15
Highest Level of Education	
High School Diploma, GED, or Associate's Degree	11
Bachelor's Degree	41
Master's Degree or Beyond	48
Certification Type	
Regular	80
Alternative	12
Temporary	5
Other	3
Position	
General Education Teacher	74
Special Education Teacher	16
Reading Specialist	6
Other	4
Grade Level Taught	
K, 1, and/or 2	73
Not K, 1, or 2	27

Note. Total number of participants = 313.

Table 2

Internal Consistency Reliability

	Number of Items	Cronbach's Alpha	McDonald's Omega
Full Scale	6	.86	.86
Beliefs about Past Preparation	2	.89	N/A
Beliefs about Current Ability	4	.82	.82

Note. McDonald's Omega cannot be estimated for the Beliefs about Past Preparation factor because the number of items is less than 3.

Table 3

Item Correlations

	Full Scale	Item 1	Item 2	Item 3	Item 4	Item 5	Item 6
Full Scale		0.80	0.83	0.80	0.77	0.73	0.70
Item 1	0.80		0.55	0.53	0.64	0.52	0.46
Item 2	0.83	0.55		0.77	0.48	0.44	0.46
Item 3	0.80	0.53	0.77		0.48	0.42	0.36
Item 4	0.77	0.64	0.48	0.48		0.54	0.52
Item 5	0.73	0.52	0.44	0.42	0.54		0.55
Item 6	0.70	0.46	0.46	0.36	0.52	0.55	

Table 4
Descriptive Statistics

	Mean	SD
Full Scale	3.86	0.79
Beliefs about Past Preparation	3.52	1.13
Beliefs about Current Ability	4.03	0.75
Item 1	3.92	1.01
Item 2	3.46	1.24
Item 3	3.58	1.17
Item 4	4.08	0.94
Item 5	3.94	0.93
Item 6	4.19	0.84

Exploring the Feasibility of Implementing SPELL-Links to Reading and Writing

Katlynn Dahl-Leonard and Colby Hall

University of Virginia

Katlynn Dahl-Leonard: awr8qt@virginia.edu; https://orcid.org/0000-0001-6124-5990

Colby Hall: csh4t@virginia.edu; https://orcid.org/0000-0002-0779-1322

The authors have no conflicts of interest to disclose.

This work was supported by Learning by Design. The content is solely the responsibility of the authors and does not necessarily represent the official views of Learning by Design.

Correspondence concerning this article should be addressed to Katlynn Dahl-Leonard at awr8qt@virginia.edu.

Abstract

This study examined six reading intervention teachers' implementation of the SPELL-Links to Reading and Writing intervention with students in Grades 2 and 3. The purpose was to evaluate the extent to which teachers were able to implement the intervention with acceptable fidelity (i.e., dosage of 30 minutes a day, four days a week, with ≥80% adherence and quality). It also aimed to better understand the determinants (i.e., barriers and facilitators) of teachers' implementation and their perceptions related to the importance, feasibility, and effectiveness of the intervention. Data from four sources (implementation logs, implementation observations, interviews, and surveys) were collected and analyzed. Participating teachers did not implement the intervention at the intended dosage, but they did achieve acceptable adherence and quality on average. Several barriers to implementation were identified, including intervention content and structure, training, compatibility with existing practices, and alignment with goals. Notable facilitators of implementation were teacher capability and peer support. Teachers had varying perceptions of the intervention, with low satisfaction on average. These findings indicate a need for further research on determinants of intervention implementation as they may be crucial in supporting teachers' implementation fidelity.

Keywords: reading intervention, fidelity, determinants of implementation

Exploring the Feasibility of Implementing SPELL-Links to Reading and Writing

A recent National Assessment of Educational Progress (NAEP) reading report indicated that 68% of fourth-grade students read below proficient levels in the United States (NAEP, 2022). There are adverse consequences associated with reading difficulties, including high school dropout, incarceration, anxiety, and depression (Dahle & Knivsberg, 2014; Daniel et al., 2006; Greenberg et al., 2007; Jordan et al., 2014). Therefore, developing proficiency in reading is essential for students' overall success and wellbeing. Importantly, access to evidence-based cinmore and supplemental reading instruction can reduce the incidence and severity of reading difficulties experienced by students (Al Otaiba et al., 2009; VanDerHeydan et al., 2007).

Unfortunately, there is a profound gap between empirical findings about evidence-based reading instructional practices and typical practice in school settings (Solari et al., 2020). There are multiple reasons for this gap, including lack of access to science of reading-aligned curricula and instructional materials. However, even when teachers have access to evidence-based reading programs, successful implementation of such programs can be challenging. Implementation science research has identified several factors that may influence the feasibility of implementing evidence-based reading interventions in authentic school contexts (Damschroder et al., 2022; Proctor et al., 2011), including intervention-level (e.g., program attributes, such as cost and design quality), individual-level (e.g., the knowledge or motivation of the teacher implementing the intervention), and school-level factors (e.g., the priority placed on implementing the intervention). The present study focused on the feasibility of reading intervention teachers of students in Grades 2 and 3 implementing an evidence-based literacy intervention. In particular, we evaluated the extent to which teachers are able to implement the intervention with fidelity and explored the teacher-reported intervention-level, individual-level, and systems-level

determinants (i.e., barriers and facilitators) of implementation fidelity.

Implementation Fidelity

Conceptual and empirical work has indicated that implementation fidelity is a multidimensional construct (e.g., Dane & Schneider, 1998; Fogarty et al., 2014; O'Donnell, 2008). Although the conceptualization of fidelity and its core components varies across research disciplines, Dane and Schneider's five dimensions of fidelity is one of the most comprehensive examinations of fidelity as it relates to early literacy intervention research (Guo et al., 2016). According to Dane and Schneider, the five dimensions of fidelity are dosage, adherence, quality, responsiveness, and differentiation. *Dosage* refers to the amount of instruction provided or how much exposure to an intervention students received. *Adherence* refers to the extent to which critical components of the intervention are implemented as intended. *Quality* of instructional delivery describes qualitative aspects of implementation, such as implementer preparedness and pacing. *Responsiveness* is the extent to which students respond to the intervention. *Differentiation* is the extent to which the intervention varies from another treatment condition or a comparison condition.

Prior research suggests a positive association of dimensions of implementation fidelity with evidence-based instruction and academic outcomes for students (Kretlow & Bartholomew, 2010; Vadasy et al., 2015; Wolgemuth et al., 2014). For example, Wolgemuth et al. (2014) reported that intervention adherence and quality of delivery impacted student outcomes on measures of phonological awareness and word reading. Vadasy et al. (2015) revealed that greater intervention dosage was associated with kindergarten students' gains in vocabulary and decoding. They also found that greater intervention adherence was associated with student gains in vocabulary and spelling. Importantly, implementation fidelity is most strongly associated with

reading outcomes for low-performing students (Boardman et al., 2016; Capin et al., 2022; Connor et al., 2007). For example, Connor and colleagues (2007) found that third-grade students with low initial reading performance disproportionately benefited from high-fidelity implementation of an explicit reading intervention when compared to their higher performing peers. Therefore, how teachers implement interventions is an important part of understanding why interventions are or are not successful for students with reading difficulties.

However, reviews of fidelity reporting within intervention research show that fidelity is frequently unreported. Swanson et al. (2011) reviewed articles published in select general and special education journals from 2005 to 2009 and found that only 47% of studies reported fidelity data. Capin et al. (2018) examined treatment fidelity in K-3 reading intervention research and found that only 47% of the reading intervention studies reported fidelity data. More recently, Dahl-Leonard et al. (2023a) examined fidelity reporting within reading intervention studies for elementary students with or at risk for dyslexia and found that 75% of studies reported fidelity data. In all three syntheses (Capin et al., 2018; Dahl-Leonard et al., 2023a; Swanson et al., 2011), the authors discovered that studies reporting fidelity data primarily reported treatment adherence, with other dimensions of fidelity (i.e., dosage, quality, responsiveness, and differentiation) consistently absent from the corpus of studies reviewed.

The lack of fidelity reporting within these studies is concerning because for researchers aiming to evaluate interventions, implementation fidelity is a crucial methodological consideration. Measuring and reporting fidelity increases the internal validity of a study evaluating the intervention's effects (i.e., when an intervention is implemented with fidelity, it is more possible to infer that the effects reported are indeed due to the intervention being evaluated; Shadish et al., 2002). Additionally, research that assesses the implementation fidelity of an

intervention is vital in showing whether the intervention is feasible for teachers to implement in authentic classroom contexts. Overall, determining the feasibility of implementing an intervention is important because it informs the practicality of scaling up the intervention and generalizing to other populations and settings (Moncher and Prinz 1991; Nelson et al. 2012; O'Donnell 2008; Solari et al., 2020).

Determinants of Implementation

In addition to measuring and reporting fidelity of implementation, there is value in identifying factors that impact intervention feasibility. A commonly used implementation science framework, the Consolidated Framework for Implementation Research (CFIR; Damschroder et al., 2022), identifies characteristics that support or hinder implementation of evidence-based practices (i.e., determinants of implementation). Notably, the CFIR describes intervention-level, teacher-level, and school-level determinants.

According to the CFIR (Damschroder et al., 2022), intervention-level determinants of implementation include the intervention source, evidence base, relative advantage, adaptability, trialability, complexity, design, and cost. For example, in a study of the feasibility of a teacher-implemented intervention, Solari et al. (2018) concluded that intervention developers "must ensure that the interventions are streamlined and efficient as well as effective to increase the likelihood of consistent implementation" (p. 187). In other words, intervention-level determinants, such as the intervention design and complexity, play a major role in the feasibility of implementation. Teacher-level characteristics that may serve as determinants of implementation include need (i.e., well-being or personal fulfillment that will be addressed by implementing), capability (i.e., competence, knowledge, and skills to implement), opportunity (e.g., availability, scope, and power to implement), and motivation (e.g., commitment to

implementing). For example, there is evidence that teachers' reading content knowledge impacts their instructional practices (Piasta et al., 2020; Spear-Swerling & Zibulsky, 2014) and that teachers' self-efficacy is associated with their reading instruction (Guo et al., 2012; Varghese et al., 2016).

School-level determinants that may relate to the implementation of interventions include general characteristics of the school, such as structural characteristics, relational connections, communications, and culture. There are also school-level factors that are specific to the implementation of the intervention, including tension for change, compatibility, relative priority, incentive systems, mission alignment, available resources, and access to knowledge and information. For example, Mihai et al. (2017) conducted qualitative analysis to understand the factors that influenced Head Start teachers' implementation of a new literacy curriculum. They discovered that the teaching context influenced teachers' implementation of the curriculum, with teachers indicating that competing responsibilities and other expectations impeded their ability to consistently implement the literacy curriculum (i.e., the relative priority of implementing the new curriculum was low). Notably, a key purpose of the CFIR is to retrospectively explain implementation outcomes by assessing determinants within particular contexts, which then allows for the development and implementation of strategies that may best address specific contextual determinants (Damschroder et al., 2022).

Present Intervention and Study Purpose

The SPELL-Links to Reading and Writing intervention is an evidence-based structured literacy program that enables intervention teachers to provide reading and spelling instruction. The SPELL-Links to Reading and Writing intervention has been shown to improve decoding, encoding, and curriculum-based writing (Apel et al., 2004; Kelman & Apel, 2004; Wanzek et al.,

2017). However, previous studies of SPELL-Links to Reading and Writing have not measured implementation fidelity (Apel et al., 2004; Kelman & Apel, 2004) or have only measured fidelity when the intervention was implemented by interventionists hired and trained by the research team (Wanzek et al., 2017). Therefore, there is value in exploring the feasibility of intervention teachers implementing the SPELL-Links to Reading and Writing in an authentic classroom context. In the present study, we observed SPELL-Links to Reading and Writing instruction provided by reading intervention teachers of students in Grades 2 and 3 to evaluate the extent to which teachers are able to implement the intervention with fidelity. We also examined the teacher-reported determinants of intervention implementation. Our research questions were:

1. To what extent are reading intervention teachers able to implement SPELL-Links to Reading and Writing with acceptable fidelity (i.e., dosage of 30 minutes a day, four days a week, with $\geq$80% adherence and quality)?
    a. What qualitative differences are present in the varying levels of implementation?
2. What do teachers perceive as the determinants of implementation of SPELL-Links to Reading and Writing?
3. To what extent do teachers consider SPELL-Links to Reading and Writing to be important, feasible, and effective?

Method

Study Context and Participants

The study was approved by the research team's institutional review board and the school district's research department. The study was conducted during the 2023-2024 academic year. Six reading intervention teachers of students in Grades 2 and 3 from four schools in one school district in Virginia participated in the study. Students in this district are primarily White (60%)

and African American (25%); 64% of students in this district are economically disadvantaged (i.e., qualify for free or reduced-price lunch). Teachers were eligible to participate in the study if they provided reading intervention to small groups of students in Grades 2 and 3. All participating teachers identified as White and female. They had an average of 10.83 years of experience as reading interventionists (range: 1-21 years). Teacher demographic information is provided in Table 1.

Description of the Intervention

The SPELL-Links to Reading and Writing intervention is an evidence-based structured literacy program that enables intervention teachers to provide reading and spelling instruction across multiple linguistic skill areas. SPELL-Links uses a five-block model of instruction to functionally integrate phonological awareness, orthographic knowledge, semantics, morphological knowledge, and mental orthographic images. SPELL-Links' meta-linguistic approach allows students to understand the linguistic properties of words, gain cognitive flexibility with multiple linguistic strategies for spelling and decoding, and practice and self-evaluate performance. SPELL-Links incorporates a systematic format of explicit instruction and practice with SPELL-Links' word study strategies, which leads to the development of independent readers and writers. Learning by Design provided the materials needed for the participating reading intervention teachers to implement SPELL-Links to Reading and Writing with small groups of up to six students for 30 minutes a day, four days a week, for the duration of the school year. They also provided a six-hour virtual training session with information about SPELL-Links to Reading and Writing, including directions on how to implement the activities.

Data Collection and Analysis Plan

We utilized four data sources to answer our research questions: implementation logs,

implementation observations, interviews, and surveys.

Implementation Logs

Participating teachers completed logs in which they document their implementation *dosage* (i.e., daily implementation of SPELL-Links to Reading and Writing). They recorded the activity or activities taught each day and the amount of time spent on each activity. Teachers documented reasons they did not implement the intervention on some days (e.g., no school, field trip, teacher absence, testing). These logs allowed us to determine implementation *dosage* (i.e., whether teachers implement the intervention on the recommended schedule of four days per week). Teacher implementation log data indicating the number of activities implemented and amount of time spent implementing each activity was analyzed descriptively.

Implementation Observations

To measure *adherence* and *quality,* we videorecorded teacher implementation on three separate occasions during the approximately 20-week study period. We employed a researcher-developed SPELL-Links to Reading and Writing adherence and quality of implementation measure. Each SPELL-Links to Reading and Writing activity provides the teacher with step-by-step directions. The *adherence* section of the measure outlines these activity steps and each activity step was coded as 1 (implemented) or 0 (not implemented). The *quality* section of the measure included four quality of implementation indicators (i.e., modeling, scaffolding, pacing, and preparation/organization) for each activity. Each quality indicator was coded on a three-point Likert scale in which 3 = high quality, 2 = moderate quality, and 1 = low quality. The measure also included a section for a written summary or important notes about the activity. Supplemental Appendix A provides further details regarding the definitions, expectations, and coding guidance for each indicator.

Each observation was coded for adherence and quality using the researcher-developed form. To analyze adherence and quality, we calculated the number of points that would be awarded for perfect implementation and the percentage of this perfect score obtained by each teacher. For example, in an activity in which the teacher implemented all activity steps, except for Step 2, she received an adherence score of 11 out of 12, or 92% adherence. These scores were averaged across observations and across teachers to determine the extent to which teachers are able to implement SPELL-Links to Reading and Writing with $\geq$ 80% adherence and quality.

Prior to coding, coders participated in a one-hour training. The training was followed by a practice session during which coders watched a video, coded the session independently, and then discussed codes. Inter-observer agreement was established prior to coding video-recorded observations, with the lead observer or "gold standard" establishing a set of correct observation codes against which other observers' codes were compared. Percent agreement was calculated as the number of agreements divided by the total number of possible codes. Coders were required to reach 90% agreement prior to coding observations independently (M = 94%). Approximately 50% of observations were double coded. Average agreement across double coded observations was 94%. Discrepancies in coding were resolved via discussion and consensus.

Interviews

At the end of the intervention period, teachers participated in semi-structured videoconference interviews (see Supplemental Appendix B for the interview guide). Semi-structured interviews are commonly used as a flexible approach to gather in-depth information about participants' perceptions (Creswell & Poth, 2017). The use of semi-structured interviews allowed us to engage in conversations with teachers related to the determinants of implementing the intervention. They also enabled us to ask follow-up questions or provide probes that

encourage teachers to elaborate on their responses (e.g., the interviewer may ask, “Can you tell me more about that?”).

To identify determinants of implementation, the teacher interviews were audio recorded, transcribed, and analyzed thematically. Thematic analysis is commonly used to analyze experiences and perspectives of research participants (Braun & Clarke, 2006). We used a priori domains from the CFIR framework (Damschroder et al., 2022) to deductively code the teacher interviews. As previously mentioned, the CFIR framework is a comprehensive list of factors related to implementation (e.g., beliefs about capabilities, environmental context and resources). Two coders independently analyzed each interview and discrepancies were resolved via discussion and consensus. By comparing teachers’ responses to the questions, we could identify patterns, or themes, in the data and explore differences in teachers’ responses relating to intervention feasibility. Table 2 provides examples of how CFIR constructs were used to analyze teacher interview responses and identify themes.

Surveys

Prior to receiving training and intervention materials, teachers completed the Teacher Understanding of Evidence-Based Literacy Instructional Practices (TULIP) survey (Hall et al., 2023) and the Teachers’ Sense of Efficacy for Literacy Instruction (TSELI) survey (Tschannen-Moran & Johnson, 2011). The TULIP survey assesses teacher knowledge in the domains of (a) phonological awareness, (b) phonics, decoding, and encoding, (c) reading fluency, (d) oral language, and (e) reading comprehension. In their validation study, Hall et al. (2023) reported a Cronbach’s alpha reliability estimate of .93. The TSELI survey asks teachers to respond to 22 items on a nine-point Likert scale (1 = not at all; 9 = a great deal) by considering the combination of their current ability, resources, and opportunity in their present position. The TSELI survey

has a Cronbach's alpha reliability estimate of .96 (Tschannen-Moran & Johnson, 2011). At the end of the study period, teachers completed a survey to assess their perceptions of the importance, feasibility, and effectiveness of the intervention. Teachers were presented with 16 items and asked to rate each item on a five-point Likert scale (1 = strongly disagree; 5 = strongly agree). Teachers were also prompted to share any additional feedback through an open-ended item at the end of the survey. The data collected via these surveys allowed us to better describe our study sample and triangulate our findings. All survey data was analyzed descriptively.

Results

Recruitment and Retention

During the initial recruitment phase for this study in May 2023, 10 teachers consented to participate. Before the study began, two teachers withdrew because they were moving to new school districts and one additional teacher consented to participate. Nine teachers participated in the virtual SPELL-Links training in September 2023. After receiving the training but before beginning to implement the intervention, three teachers withdrew from the study. Teacher reasons for withdrawal included feeling like implementing the SPELL-Links program was "too much to handle" and not feeling like they can "teach this program with the energy and focus that it deserves." One teacher, who was new to the school and position, similarly noted that it was "not the best time for [her] to try a new program." Six teachers began implementing the intervention in October 2023. In January 2023, one teacher (T3) decided to discontinue implementation of SPELL-Links with her students because she noticed her students "have not grown since beginning this program." However, she agreed to participate in an interview and complete a survey so that we could learn about her experiences with the intervention.

Fidelity of Implementation

Individual teacher implementation fidelity data (i.e., dosage, adherence, and quality) is reported in Table 3.

Dosage

In terms of *dosage*, teachers were expected to implement SPELL-Links activities four days a week for 30 minutes per day. For the teachers who implemented the intervention for all 20 weeks of the study period, the average implementation dosage was 52 sessions (range: 44-65). Thus, they implemented 2.60 sessions per week on average. The teacher who only used the intervention for 11 weeks implemented 23 sessions, for an average of 2.09 sessions per week. In their implementation logs, teachers recorded when they did not implement the intervention and their reasons for not implementing it. The most reported reasons for not delivering the intervention were no school/early release (average of 24 instances across teachers), teacher was unavailable (e.g., due to training, teacher absence; average of 10 instances across teachers), and student testing (average of 8 instances across teachers). Regarding student testing, is worth noting that several teachers administered intervention-aligned assessments frequently (i.e., approximately every 2-3 weeks) to assess their students' mastery of the intervention content. Additionally, teachers reported that although they pulled their groups for 30 minutes, only about 25 minutes were spent on instruction. Based on the observation videos, the average session length was 24.55 minutes (range: 15.75-29.75). Overall, teachers were not able to implement the intervention at the intended dosage of 30 minutes per day, four days per week.

Adherence and Quality

Teachers were expected to *adhere* to $\geq$80% of the steps when implementing the activities. Data from observations demonstrated that teachers adhered to 85% of activity steps on average (range: 67%-100%). Across teachers, average adherence ratings ranged from 73% to 100%.

Average adherence ratings met or exceeded 80% for five of the six teachers. Teachers were also expected to achieve ≥80% *quality* of instructional delivery. The average quality of delivery rating across the observations was 95% (range: 83%-100%). Across teachers, average quality ratings ranged from 89% to 100%. Thus, average quality ratings exceeded 80% for all teachers. Taken together, teachers in our sample implemented the intervention with adequate adherence and quality.

Qualitative Differences

Based on our qualitative data (i.e., written summaries) from observation coding, we are able to construct a better understanding of teachers' fidelity scores. For example, one teacher (T5) whose average adherence across the three observations was 85% had individual observation adherence scores of 100%, 70%, and 86%. This large variance in adherence scores is unusual (especially in combination with an average quality score of 100%) and warranted further exploration. The qualitative data for this teacher's observations revealed that she frequently adapted activities based on the needs of her students. For example, in one activity, the coder noted, "The teacher skipped Steps 5-7, but it seemed intentional. Her students did not seem to need the intermediate steps of having the first letter present before spelling the whole word, and completing all the steps would have taken too much time. Therefore, this felt like a successful adaptation to the activity." However, making this adaptation involved skipping activity steps and ultimately resulted in a lower adherence score for this activity. T

We also explored how teachers with varying levels of adherence and quality implemented similar activities. For example, two teachers (T4 and T6) implemented the same activity (12.1 "Lines & Letters") during their second observation. T4 implemented the activity with 100% adherence and quality. The qualitative summary of her implementation of this activity stated,

“The teacher really seemed to understand the activity and provided additional information to support her students' learning. Before starting the activity, she did a review of letter sounds. During the activity, she modeled before students practiced (i.e., used explicit instruction). She also gave students additional ‘challenge’ words to practice (i.e., upward scaffolding).” On the other hand, T6 implemented the activity with 75% adherence and 83% quality. Her qualitative summary stated, “It felt like she rushed into the activity and did not provide much explanation about the objective of the activity as directed in Step 1. She delivered Steps 2-5 well but skipped Step 6 and only discussed one of the two strategies for Step 8. Together, these actions indicate that she was not well prepared to implement the activity.” These stark differences in activity adherence and quality make it worthwhile to explore the determinants of implementation.

Determinants of Implementation

The primary *barriers* to implementation that we identified from teacher interviews and surveys were related to intervention content and structure, training, compatibility with existing practices, and alignment with goals. The primary *facilitators* we identified were teacher capability (e.g., knowledge, skills, self-efficacy) and support from other teachers. We detail each of these determinants below, noting how these factors related to implementation fidelity.

Intervention Content and Structure

All six teachers indicated in their interviews that the intervention content and structure was not ideal. Notably, they perceived the intervention as complex to implement. A teacher with adequate adherence and very high quality of implementation explained, “It was an overwhelming amount of information, and even if you read ahead, prepped, and did what you were supposed to, and had everything out and read ahead of time, still you didn't feel like you knew what you were really doing” (T5). Another teacher with similar levels of adherence and quality elaborated,

“Even though I would read through the lessons multiple times before I did it, I found it wasn't easy to just move through. It wasn't always natural. ... I just always was not positive I was doing the right thing” (T1). Teachers also noted that the intervention did not focus on the skills they thought were most relevant. For example, one teacher mentioned, “I would like it if it had more reading because one passage for six lessons is not a whole lot of [reading] in context” (T1). Another teacher shared, “I didn't feel like there was enough reading involved. It was like a lot of spelling and some reading, but it wasn't reading in context at all” (T5). The same teacher also said, “I feel like it's too much of a focus on just one thing, instead of getting through multiple things in a day, being able to do warm up things, being able to read, being able to write and spell.” One teacher with very high adherence and quality but moderate dosage mentioned, “I did supplement some materials because I didn't feel like it was always enough for my intervention babies” (T4). This teacher also said, “I think my intervention babies need more explicit [instruction].” She further described, “It felt a little bit like they were doing a lot of discovery learning ... but I think with the intervention kids, they need all the support and scaffold that they can get, and just kind of asking them to discover it, I don't necessarily think is the best way to get there in intervention.”

Additionally, the intervention was designed for one-on-one implementation, but all teachers in our sample used the intervention with a small group of students and thus were required to adapt the intervention for small group instruction. One teacher with adequate adherence and high quality explained, “It made it a little more complicated when you read the directions to be like, ‘Okay, I'm not doing this with one kid, so now I just need to divvy things up equally, or I need to, like I said, we're not back and forth, back and forth, we're going around the table. It wasn’t super hard to do, but it's a harder lesson. It's not that hard to change it like

mentally in your head, but it was harder to just implement in general than if it had been one on one" (T5). Another teacher shared, "I would do kind of a mix between what it said for small group and what it said [for one-on-one]. My students don't always work well together because of their personalities. ... So, instead of putting words on cards, I would put it on the board, and then we talk about it on the board together, and we fix it together, and then we'd use their notebook. ... but we would do a lot using the board instead of individual cards" (T1). Notably, this teacher had rather low adherence despite having high quality of delivery. Based on the qualitative data from the observations, this teacher's adaptations for small group instruction occasionally led to low adherence to the activity steps.

Intervention Training

All six teachers indicated in their interviews that they perceived the intervention training as insufficient, and that more training would have helped them implement the intervention more successfully. When asked about the training, one teacher with low dosage, adequate adherence, and high quality shared, "I felt like it was not enough, and also crammed for one day, like it was really long, and I think it would have been better to have an in-person training versus a Zoom training, and that it could have been broken up over at least two days, maybe more." A teacher with lower adherence but high quality similarly mentioned, "It was a lot in one day ... if it was in smaller chunks, it probably would have been better so I could have tried something and then gone back for further training on it" (T1). Another teacher with low adherence further explained, "there is a lot packed into that one day, and it seems like there were some things that we could have spent more time on, and kind of spread out some of that" and "if we had time to actually do some of the activities together, and actually take more time to kind of investigate and figure all that out, because a lot of it we had to figure out on our own" (T6). Even a teacher with very high

adherence and quality shared similar sentiments. She noted that the training “probably should have been longer” and that “it would have been good to sleep on it and come back” (T4).

Compatibility with Existing Practices

Some teachers noted the lack of compatibility with their existing practices as a perceived barrier to implementation. For example, a teacher with moderate dosage who implemented the intervention with very high adherence and quality said, “It was so different from what we're already doing that it just felt like it was unknown. I was trying to jump in, and I did, but it was unknown, and it just felt like it took a lot of time to prep. ... It was just tough for this year, and the structure of our intervention, I didn't feel like it really fit with what we were already doing. And I think that's the part that I had the hardest time with” (T4). Another teacher with low dosage, adequate adherence, and high quality said, “It was just too much time and energy for me to devote for one 30-minute group when I have nine groups” (T5). She further explained, “It was helpful that our planning was right before this group... I would use part of that time every day to prep, which is more than I had to do for any of my other groups because [for those groups] I could do my lesson plans for the week, know what to print, know what to read, and everything is pretty much ready, whereas this I felt like it was too much to prep every single day.” Notably, even the teacher with the highest dosage said, “if you're really going to be committed to it... you need more time to have the commitment” (T6).

Alignment with Goals

A few teachers also expressed a lack of alignment between the intervention and their goals. One teacher with low implementation dosage shared, “I felt like it was going really slow. And I was never going to get to where I needed to get for my second graders to be able to pass that PALS spelling” (T3). Another teacher with low dosage further explained, “It felt like it was

really slow to get through what we're getting through. I'm still on digraphs and it's February. I've gotten through short vowels, and I'm not even through digraphs, and that's all we've done, when in other programs I could have knocked out that in November and been on to the next thing" (T5). A teacher with moderate dosage but very high adherence and quality similarly noted, "I would have liked for it to move a little bit faster" (T4).

Teacher Capability

We identified teacher capability (e.g., knowledge, skills, self-efficacy) as a facilitator of implementation. For our sample of teachers, the average percent correct on the teacher knowledge survey was 93% (range: 80%-100%), indicating that teachers demonstrated a high level of knowledge of literacy instruction. There is a noteworthy pattern in the fidelity data related to knowledge. The two teachers with the highest knowledge survey scores (T4, 100%; T2, 97%) received the highest average adherence scores (T4, 100%; T2, 88%) whereas the teacher with the lowest knowledge survey score (T3, 80%) received the lowest average adherence score (73%). This finding suggests that teachers with higher levels of knowledge may be able to better adhere to the activity steps for this intervention. Additionally, the average self-efficacy survey rating was 7.18 out of 9.00 (range: 6.64-7.82), indicating moderate to high levels of self-efficacy for providing literacy instruction. There is also a noteworthy pattern in the fidelity data related to self-efficacy. The two teachers with the lowest self-efficacy ratings (T2, 6.64; T6, 6.82) also received the lowest average quality of implementation scores (both 89%), suggesting that lower self-efficacy may lead to lower quality of implementation.

Our interviews revealed similar findings. For example, the teacher with the highest adherence (T4) noted she felt "pretty confident" implementing the intervention. She explained, "because doing reading intervention, we already know what they need, and we're kind of used to

breaking it down by skill, and I already had a lot of background about the skills I was teaching, so I think that helped." On the other hand, some of the lower adherence teachers said they were "only mildly" (T1) or "not real" confident (T6), or that their confidence "depended on the activity" (T3).

Peer Support

Three of our teachers were located at the same school. They described being able to discuss the intervention during lunch and planning periods together as "beneficial" to their implementation (T5). One teacher explained, "we got on a flow for a while where we would be within a couple of days of each other, so it would be like, 'okay, I made these cards, let me pass them to you' or you'd be like, 'okay, I looked at this, and I don't understand it, did you?'" (T4). Another teacher shared, "doing it together has helped with getting through it" and "it really does work better if you have that right there in your school" (T6). Overall, these three teachers implemented the intervention with moderate to high dosage. The average adherence scores for these teachers ranged from 80%-100% and their average quality scores ranged from 89%-100%.

Perceptions of Intervention Importance, Feasibility, and Effectiveness

Based on the teacher survey of the importance, feasibility, and effectiveness of the intervention, teachers had varying perceptions of the intervention (see Table 4 for individual teacher ratings). On a 5-point scale (1 = strongly disagree; 5 = strongly agree), all teachers indicated that they agreed that the literacy skills targeted in the intervention are important (M = 4.33; SD = 0.82). They also somewhat agreed that the activities in the intervention were appropriate for their students (M = 3.50; SD = 1.38). However, their overall satisfaction with the intervention was rather low (M = 2.67; SD = 1.03). Notably, teachers indicated that they did not feel like the training and online platform provided them with adequate support to implement the

intervention (*M* = 2.67; *SD* = 1.51). Accordingly, teachers were only somewhat confident in their abilities to implement the activities (*M* = 3.33; *SD* = 1.03). Regarding feasibility, teachers slightly agreed that the activities were easy to implement (*M* = 3.33; *SD* = 1.21), could be completed in a reasonable amount of time (*M* = 3.33; *SD* = 0.82), and could be implemented without extensive preparation (*M* = 3.17; *SD* = 0.98), but they did not agree that the activity materials easy to use (*M* = 2.67; *SD* = 0.52). In the open-ended response section of survey, one teacher with high dosage, but moderate adherence explained, "If the training was face to face and more explicit, I feel I would have felt better about implementing the program" (T6).

Teachers somewhat agreed that the intervention improved their students' abilities to read words (*M* = 3.33; *SD* = 0.52), but somewhat disagreed that it improved their students' abilities to write words (*M* = 2.67; *SD* = 1.03). They were relatively neutral on several other items related to their perceived effectiveness of the intervention: "continuing to use the activities with my students will make them better readers" (*M* = 3.17; *SD* = 1.17), "continuing to use the activities with my students will make them better writers" (*M* = 3.17; *SD* = 1.47), "the learning that occurred during this program will help my students be successful in the upcoming years of education" (*M* = 3.00; *SD* = 1.10), and "my students have generalized the skills they learned from the program to other environments and contexts" (*M* = 3.00; *SD* = 1.10). They did not agree that the intervention holds promise for other students (*M* = 2.67; *SD* = 1.21).

Several teachers provided more information about their perceptions of the intervention in the open-ended response section of survey. For example, one teacher shared, "I have a strong literacy background, so there are many phonics tools in my literacy toolbox. This program had some positive components, but I believe that there are more activities and strategies needed in our reading intervention setting" (T2). One teacher mentioned, "The program was not systematic

and explicit enough for my students, and too many ways to read or spell words were given at a time" (T3). Another teacher similarly noted, "I think that at times this program gives too much information at once" and "this program was not explicit enough in the spelling" (T1).

Discussion

The purpose of this study was to evaluate the extent to which teachers were able to implement the SPELL-Links to Reading and Writing intervention with acceptable fidelity (i.e., dosage of 30 minutes a day, four days a week, with ≥80% adherence and quality). It also aimed to better understand the determinants (i.e., barriers and facilitators) of teachers' implementation of the intervention as well as their perceptions of the importance, feasibility, and effectiveness of the intervention. Overall, teachers in the present study did not implement the intervention at the intended dosage, but they did achieve acceptable adherence and quality. That said, several barriers to implementation were identified, including intervention content and structure, training, compatibility with existing practices, and alignment with goals. Notable facilitators of intervention implementation were teacher capability and peer support. Lastly, teachers had varying perceptions of the intervention, with overall satisfaction with the intervention being rated rather low.

Exploring Determinants of Implementation Fidelity

On average, teachers implemented the SPELL-Links to Reading and Writing intervention with acceptable adherence to activity steps and delivered it with acceptable quality. However, as a group, they did not implement it at the intended dosage of four days per week. This finding is not unexpected given that intervention research has consistently demonstrated difficulties with achieving intended dosage (Denton et al., 2021; Solari et al., 2018). Yet it is problematic because current evidence suggests that interventions should be delivered with the intended dosage (e.g.,

four sessions per week) to impact student outcomes (Vadasy et al., 2015; Wolgemuth et al., 2014). Therefore, it is worthwhile to explore the determinants of implementation to better understand how we can support teachers' successful intervention implementation.

Intervention-Level Determinants

At the intervention level, our findings suggest that the content and structure of the intervention was a key barrier to implementation. This barrier was not surprising given that previous research has highlighted the importance of developing interventions that are streamlined and efficient (Solari et al., 2018). Further, in a study on teachers' perceived barriers to intervention implementation, Long et al. (2016) revealed that over half (58%) of the teacher-identified barriers to implementation were perceived to be related to the intervention itself.

It was also not surprising that the intervention training was identified as a barrier to implementation. In a meta-analysis on the effects of teacher training on intervention implementation, Brock and Carter (2016) estimated that training has a strong effect on implementation fidelity (g = 1.08). Research suggests that teacher training should build upon background knowledge, explicitly describe and illustrate content, provide opportunities to actively apply and generalize learned content in real-world contexts, and support metacognition (e.g., reflection and self-monitoring) throughout the training process (Trivette et al., 2009). Although the use of a one-day standalone training with limited follow-up is common, without structured, ongoing implementation support (e.g., coaching) teachers may show low levels of fidelity (Guskey & Yoon, 2009). The fact that participating teachers were only provided with a one-day training without ongoing coaching may be one reason for lower levels of fidelity.

It is worthy to note that the teachers in our sample, on average, were experienced reading interventionists, demonstrated high levels of knowledge to teach literacy, and reported moderate

to high levels of self-efficacy for teaching literacy. Therefore, we posit that they were often able to use their experience, knowledge, and skills to overcome intervention-level implementation barriers. For example, one teacher (T4), who described the intervention as not being explicit enough and not having enough practice and review for her students, explained that she was able to adapt the intervention and "supplement" with materials she thought were more appropriate for her students, while still providing the intervention as intended (i.e., average adherence and quality was 100% for this teacher).

School-Level Determinants

Three school-level determinants were identified in our data: compatibility with existing practices, alignment with goals, and peer support. Notably, two of these determinants (compatibility with existing practices and alignment with goals) served as barriers, whereas peer support served as a facilitator. The lack of compatibility with existing practices was primarily evident through the perceived large amount of planning and preparation required for implementing the intervention. This finding is in alignment with previous research on barriers to implementation, which has demonstrated adequate planning time is a prerequisite for high-level implementation (Long et al., 2016) and that competing priorities can make it challenging to find time to implement (Mihai et al., 2017). In other words, when teachers do not have enough time to plan and prepare, their implementation of the intervention suffers.

The lack of alignment with the teachers' goals was another significant school-level barrier to intervention implementation. Teachers often cited concerns regarding the inability of the intervention to adequately address the skills their students needed to master to meet state-specified benchmarks within the allotted timeframe. Given that prior research has shown that teachers are more motivated to implement an intervention when they believe it will improve

student outcomes (Cramer et al., 2023), it follows that the teachers in our sample did not implement the intervention at the intended dosage.

Peer support served as a facilitator of intervention implementation for the teachers in our study. They reported that having other teachers at their school who were also using the intervention was beneficial to their implementation. Previous research suggests that having a strong support system within their schools increases teachers' sense of efficacy (Dahl-Leonard et al., 2023b) and that teacher self-efficacy is associated with improved instructional quality and student performance (Guo et al., 2012; Varghese et al., 2016). Therefore, it is possible this perceived support influenced their implementation fidelity as well as student outcomes.

Teacher-Level Determinants

Based on findings from prior research (Guo et al., 2012; Piasta et al., 2020), we anticipated that teachers' capabilities (e.g., knowledge, skills, self-efficacy) would emerge as a determinant of implementation. We found that teacher knowledge and self-efficacy were both facilitators of high-fidelity implementation in this study. In particular, teacher knowledge appeared to align with adherence to activity steps and teacher self-efficacy appeared to be linked to quality of implementation. We also believe that our teachers used their experience, knowledge, and skills to overcome several barriers to implementation. It is worth noting that one teacher (T4) scored 100% correct on the knowledge survey and described feeling "pretty confident" implementing the intervention due to her background. She was the only teacher to achieve an average of 100% adherence and quality. As described previously, this teacher was able to make successful adaptations to the intervention when she felt it did not provide her students with the knowledge and skills that they needed.

Perceptions of Intervention Importance, Feasibility, and Effectiveness

The results from the teacher survey on the importance, feasibility, and effectiveness of the intervention mostly aligned with our observation and interview data. For example, on the survey, teachers somewhat agreed (M = 3.5) that the activities in the intervention were appropriate for their students, with ratings ranging from 1 (strongly disagree) to 5 (strongly agree), and the average rating for the item "I was satisfied with the program" was 2.67, with ratings ranging from 1 to 4, indicating no teachers strongly agreed with that statement. Our interviews similarly revealed that not all teachers were satisfied with the content and structure of the intervention. For example, in one interview, the teacher acknowledged, "I know that everything I'm saying is pretty negative. I overall have a negative feeling towards the program" (T5). It was also apparent from the survey and interview data that teachers were not satisfied with the intervention training they received, and as a result, they were only somewhat confident in their ability to implement the intervention successfully. As previously mentioned, this lack of confidence has implications for the quality of implementation.

On average, teachers' perceptions about the feasibility of implementing the intervention were relatively neutral. Survey results indicated teachers slightly agreed that the activities could be completed in a reasonable amount of time (M = 3.33) and without extensive preparation (M = 3.17). Interviews with teachers demonstrated similarly neutral perceptions but did reveal that the amount of preparation required to implement the intervention was difficult for some teachers to manage. For example, one teacher shared, "It was just too much time and energy for me to devote for one 30-minute group when I have nine groups" (T5). This lack of compatibility with existing school-level systems (e.g., planning time) was an identified barrier to high-fidelity implementation.

Lastly, teachers demonstrated mixed perceptions about the effectiveness of the

intervention. As previously noted, one teacher (T3) decided to stop using the intervention after 11 weeks because she was not seeing student growth. In her interview, this teacher expressed concerns that if she continued to use the intervention, then her students would not meet state-specified benchmarks, especially for spelling. Survey data similarly indicated that, on average, teachers did not see improvement in their students' abilities to write words. However, it also indicated that teachers did see improvement in their students' abilities to read words on average. That said, the average rating for the item "I think this program holds promise for other students" was 2.67, with ratings ranging from 1 to 4, indicating that teachers did not agree with this statement on average. Again, this is problematic because teachers are more motivated to implement an intervention when they believe it will improve student outcomes (Cramer et al., 2023).

Limitations

There are some limitations to this study. First, the participant sample only included six teachers from one school district. Thus, the reported findings are exploratory and have limited generalizability. Additionally, a primary data collection method utilized in the present study was semi-structured interviews, which has strengths and weaknesses. Although semi-structured interviews are commonly used as a flexible approach to gathering information about participants' perceptions, this interview method inherently lacks the rigor of structured interviews (Creswell & Poth, 2017). In the present study, we did not ask all participants the exact same questions or prompt them to elaborate on their responses in the exact same way. That said, interview data was paired with observation and survey data to triangulate our findings. It is also worth noting that some activities included as few as eight steps and all steps were weighted equally. Therefore, in some instances, missing one activity step could result in a much lower adherence score (e.g.,

implementing 7 out of 8 steps equals 88% adherence).

Implications and Future Directions

Although the teachers in our sample were able to implement the intervention with acceptable adherence and quality, several determinants of successful implementation were identified. Thus, it is important to explore the practical implications of our findings, especially related to overcoming implementation barriers. Results suggested barriers to implementation frequently related to the intervention content and structure. In particular, teachers reported that the intervention was "overwhelming" and "not easy to move through." Teachers also felt that the intervention was not always appropriate for their students. Therefore, intervention developers should consider ways to ensure that their intervention is straightforward and efficient for teachers to implement. They should also ensure that their intervention incorporates evidence-based instructional practices, such as explicit instruction and cumulative review. Further, administrators adopting new interventions should select interventions that are compatible with the existing practices at their school and that have sufficient alignment with the goals for their students. They should also strive to create environments in which teachers are implementing the same intervention have time to plan and discuss the intervention with one another as our findings show that teachers may benefit from having a support system within their schools. Also, it may be worthwhile for administrators to provide teachers with more opportunities to gain knowledge and skills, and increase their self-efficacy, via training and ongoing support when implementing a new intervention.

Overall, this study demonstrated that reading intervention teachers were able to implement the SPELL-Links to Reading and Writing intervention with acceptable adherence and quality, but not adequate dosage. Examination of the determinants of implementation may be

crucial in supporting teachers' implementation of such interventions. However, the present findings should be considered exploratory rather than conclusive. The field would benefit from future studies that incorporate a larger, more diverse sample of participants. Further, the addition of student-level data could allow for more sophisticated data analysis, such as exploring associations between teacher implementation fidelity and student reading achievement.

References

Al Otaiba, S., Connor, C. M., Foorman, B., Schatschneider, C., Greulich, L., & Sidler, J. F. (2009). Identifying and intervening with beginning readers who are at-risk for dyslexia: Advances in individualized classroom instruction. *Perspectives on Language and Literacy, 35*(4), 13-19. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4296731/

Al Otaiba, S., Lake, V. E., Greulich, L., Folsom, J. S., & Guidry, L. (2012). Preparing beginning reading teachers: An experimental comparison of initial early literacy field experiences. *Reading and Writing, 25*(1), 109-129. https://doi.org/10.1007/s11145-010-9250-2

Apel, K. E. N. N., Masterson, J., & Hart, P. (2004). Integration of language components in spelling. In E. R. Silliman & L. C. Wilkinson (Eds.), *Language and Literacy Learning in Schools* (pp. 292-318). Guilford Press.

Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review, 84*(2), 191–215. https://doi.org/10.1037/0033-295X.84.2.191

Boardman, A. G., Buckley, P., Vaughn, S., Roberts, G., Scornavacco, K., & Klingner, J. K. (2016). Relationship between implementation of collaborative strategic reading and student outcomes for adolescents with disabilities. *Journal of Learning Disabilities, 49*(6), 644-657. https://doi.org/10.1177/0022219416640784

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*(2), 77-101. https://doi.org/10.1191/1478088706qp063oa

Brock, M. E., & Carter, E. W. (2017). A meta-analysis of educator training to improve implementation of interventions for students with disabilities. *Remedial and Special Education, 38*(3), 131-144. https://doi.org/10.1177/0741932516653477

Capin, P., Roberts, G., Clemens, N. H., & Vaughn, S. (2022). When treatment adherence

matters: Interactions among treatment adherence, instructional quality, and student characteristics on reading outcomes. *Reading Research Quarterly, 57*(2), 753-774. https://doi.org/10.1002/rrq.442

Capin, P., Walker, M. A., Vaughn, S., & Wanzek, J. (2018). Examining how treatment fidelity is supported, measured, and reported in K–3 reading intervention research. *Educational Psychology Review, 30*(3), 885-919. https://doi.org/10.1007/s10648-017-9429-z

Castles, A., Rastle, K., & Nation, K. (2018). Ending the reading wars: Reading acquisition from novice to expert. *Psychological Science in the Public Interest, 19*(1), 5-51. https://doi.org/10.1177/1529100618772271

Connor, C. M., Morrison, F. J., Fishman, B. J., Schatschneider, C., & Underwood, P. (2007). Algorithm-guided individualized reading instruction. *Science, 315*(5811), 464-465. https://doi.org/10.1126/science.1134513

Cramer, T., Cappella, E., Morris, P., & Ganimian, A. (2023). Measuring and predicting teachers' commitment to implementing evidence-based programs. *Early Childhood Research Quarterly, 64*(1), 405-415. https://doi.org/10.1016/j.ecresq.2023.04.009

Creswell, J. W., & Poth, C. N. (2017). *Qualitative inquiry and research design: Choosing among five approaches.* Sage Publications.

Dahl-Leonard, K., Hall, C., Capin, P., Solari, E. J., Demchak, A., & Therrien, W. J. (2023a). Examining fidelity reporting within studies of foundational literacy skill interventions for elementary students with or at risk for dyslexia. *Annals of Dyslexia, 73*(2), 288-313. https://doi.org/10.1007/s11881-023-00279-3

Dahl-Leonard, K., Mesa, M. P., Hall, C., & Zucker, T. A. (2023b). Understanding kindergarten teacher self-efficacy for providing reading instruction to students with reading

difficulties. *Teachers and Teaching: Theory and Practice, 28*(7-8), 772-787. https://doi.org/10.1080/13540602.2023.2201424

Dahle, A. E., & Knivsberg, A.-M. (2014). Internalizing, externalizing and attention problems in dyslexia. *Scandinavian Journal of Disability Research, 16*(2), 179–193. https://doi.org/10.1080/15017419.2013.781953

Damschroder, L. J., Reardon, C. M., Widerquist, M. A. O., & Lowery, J. (2022). The updated Consolidated Framework for Implementation Research based on user feedback. *Implementation Science, 17*(1), 1-16. https://doi.org/10.1186/s13012-022-01245-0

Dane, A. V., & Schneider, B. H. (1998). Program integrity in primary and early secondary prevention: Are implementation effects out of control? *Clinical Psychology Review, 18*(1), 23-45. https://doi.org/10.1080/15017419.2013.781953

Daniel, S. S., Walsh, A. K., Goldston, D. B., Arnold, E. M., Reboussin, B. A., & Wood, F. B. (2006). Suicidality, school dropout, and reading problems among adolescents. *Journal of Learning Disabilities, 39*(6), 507–514. https://doi.org/10.1177/00222194060390060301

Denton, C. A., Montroy, J. J., Zucker, T. A., & Cannon, G. (2021). Designing an intervention in reading and self-regulation for students with significant reading difficulties, including dyslexia. *Learning Disability Quarterly, 44*(3), 170-182. https://doi.org/10.1177/0731948719899479

Fogarty, M., Oslund, E., Simmons, D., Davis, J., Simmons, L., Anderson, L., Clemens, N., & Roberts, G. (2014). Examining the effectiveness of a multicomponent reading comprehension intervention in middle schools: A focus on treatment fidelity. *Educational Psychology Review, 26*, 425-449. https://doi.org/10.1007/s10648-014-9270-6

Greenberg, E., Dunleavy, E., & Kutner, M. (2007). Literacy behind bars: Results from the 2003

national assessment of adult literacy prison survey. NCES 2007-473. *National Center for Education Statistics.* https://eric.ed.gov/?id=ED496564

Guo, Y., Connor, C. M., Yang, Y., Roehrig, A. D., & Morrison, F. J. (2012). The effects of teacher qualification, teacher self-efficacy, and classroom practices on fifth graders' literacy outcomes. *The Elementary School Journal, 113*(1), 3-24. https://doi.org/10.1086/665816

Guo, Y., Dynia, J. M., Logan, J. A., Justice, L. M., Breit-Smith, A., & Kaderavek, J. N. (2016). Fidelity of implementation for an early-literacy intervention: Dimensionality and contribution to children's intervention outcomes. *Early Childhood Research Quarterly, 37*, 165-174. https://doi.org/10.1016/j.ecresq.2016.06.001

Guskey, T. R., & Yoon, K. S. (2009). What works in professional development? *Phi Delta Kappan*, *90*(7), 495-500. https://doi.org/10.1177/003172170909000709

Hall, C., Solari, E. J., Hayes, L., Dahl-Leonard, K., DeCoster, J., Kehoe, K. F., Conner, C., Henry, A. R., Demchak, A., Richmond, C. L., & Vargas, I. (2023). Validation of an instrument for assessing elementary-grade educators' knowledge to teach reading. *Reading and Writing.* Advance online publication. https://doi.org/10.1007/s11145-023-10456-w

Jordan, J. A., McGladdery, G., & Dyer, K. (2014). Dyslexia in higher education: Implications for maths anxiety, statistics anxiety and psychological well-being. *Dyslexia, 20*(3), 225-240. https://doi.org/10.1002/dys.1478

Kelman, M. E., & Apel, K. (2004). Effects of a multiple linguistic and prescriptive approach to spelling instruction: A case study. *Communication Disorders Quarterly*, *25*(2), 56-66. https://doi.org/10.1177/15257401040250020301

Kretlow, A. G., & Bartholomew, C. C. (2010). Using coaching to improve the fidelity of evidence-based practices: A review of studies. *Teacher Education and Special Education, 33*(4), 279-299. https://doi.org/10.1177/0888406410371643

Long, A. C., Sanetti, L. M. H., Collier-Meek, M. A., Gallucci, J., Altschaefl, M., & Kratochwill, T. R. (2016). An exploratory investigation of teachers' intervention planning and perceived implementation barriers. *Journal of School Psychology, 55*(1), 1-26. https://doi.org/10.1016/j.jsp.2015.12.002

Mihai, A., Butera, G., & Friesen, A. (2017). Examining the use of curriculum to support early literacy instruction: A multiple case study of head start teachers. *Early Education and Development, 28*(3), 323-342. https://doi.org/10.1080/10409289.2016.1218729

Moncher, F. J., & Prinz, R. J. (1991). Treatment fidelity in outcome studies. *Clinical Psychology Review, 11*(3), 247-266. https://doi.org/10.1016/0272-7358(91)90103-2

National Assessment of Education Progress. (2022). NAEP Reading 2022 Highlights. The Nation's Report Card. https://www.nationsreportcard.gov/highlights/reading/202/.

Nelson, M. C., Cordray, D. S., Hulleman, C. S., Darrow, C. L., & Sommer, E. C. (2012). A procedure for assessing intervention fidelity in experiments testing educational and behavioral interventions. *The Journal of Behavioral Health Services & Research, 39*(4), 374-396. https://doi.org/10.1007/s11414-012-9295-x

O'Donnell, C. L. (2008). Defining, conceptualizing, and measuring fidelity of implementation and its relationship to outcomes in K–12 curriculum intervention research. *Review of Educational Research, 78*(1), 33-84. https://doi.org/10.3102/0034654307313793

Piasta, S. B., Soto Ramirez, P., Farley, K. S., Justice, L. M., & Park, S. (2020). Exploring the nature of associations between educators' knowledge and their emergent literacy

classroom practices. *Reading and Writing*, *33*(6), 1399-1422. https://doi.org/10.1007/s11145-019-10013-4

Podhajski, B., Mather, N., Nathan, J., & Sammons, J. (2009). Professional development in scientifically-based reading instruction: Teacher knowledge and reading outcomes. *Journal of Learning Disabilities*, *42*(5), 403-417. https://doi.org/10.1177/0022219409338737

Proctor, E., Silmere, H., Raghavan, R., Hovmand, P., Aarons, G., Bunger, A., Griffey, R., & Hensley, M. (2011). Outcomes for implementation research: conceptual distinctions, measurement challenges, and research agenda. *Administration and Policy in Mental Health and Mental Health Services Research*, *38*(2), 65-76. https://doi.org/10.1007/s10488-010-0319-7

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference.* Houghton Mifflin.

Snow, C. E., Griffin, P. E., & Burns, M. (2005). *Knowledge to support the teaching of reading: Preparing teachers for a changing world.* Jossey-Bass.

Solari, E. J., Denton, C. A., Petscher, Y., & Haring, C. (2018). Examining the effects and feasibility of a teacher-implemented Tier 1 and Tier 2 intervention in word reading, fluency, and comprehension. *Journal of Research on Educational Effectiveness*, *11*(2), 163-191. https://doi.org/10.1080/19345747.2017.1375582

Solari, E. J., Terry, N. P., Gaab, N., Hogan, T. P., Nelson, N. J., Pentimonti, J. M., Petscher, Y., & Sayko, S. (2020). Translational science: A road map for the science of reading. *Reading Research Quarterly*, *55*(S1), S347-S360. https://doi.org/10.1002/rrq.357

Spear-Swerling, L., & Zibulsky, J. (2014). Making time for literacy: Teacher knowledge and

time allocation in instructional planning. *Reading and Writing*, *27*(8), 1353-1378. https://doi.org/10.1007/s11145-013-9491-y

Swanson, E., Wanzek, J., Haring, C., Ciullo, S., & McCulley, L. (2011). Intervention fidelity in special and general education research journals. *Journal of Special Education, 47*(1), 3-13. https://doi.org/10.1177/0022466911419516

Trivette, C. M., Dunst, C. J., Hamby, D. W., & O'herin, C. E. (2009). Characteristics and consequences of adult learning methods and strategies. *Research Brief, 3*(1), 1-33. http://files.eric.ed.gov/fulltext/ED565253.pdf

Tschannen-Moran, M., & Johnson, D. (2011). Exploring literacy teachers' self-efficacy beliefs: Potential sources at play. *Teaching and Teacher Education, 27*(4), 751-761. https://doi.org/10.1016/j.tate.2010.12.005

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

VanDerHeyden, A. M., Witt, J. C., & Gilbertson, D. (2007). A multi-year evaluation of the effects of a Response to Intervention (RTI) model on identification of children for special education. *Journal of School Psychology*, *45*(2), 225–256. https://doi.org/10.1016/j.jsp.2006.11.004

Varghese, C., Garwood, J. D., Bratsch-Hines, M., & Vernon-Feagans, L. (2016). Exploring magnitude of change in teacher efficacy and implications for students' literacy growth. *Teaching and Teacher Education, 55*(1), 228-239.

https://doi.org/10.1016/j.tate.2016.01.011

Wanzek, J., Gatlin, B., Al Otaiba, S., & Kim, Y. S. G. (2017). The impact of transcription writing interventions for first-grade students. *Reading & Writing Quarterly, 33*(5), 484-499. https://doi.org/10.1080/10573569.2016.1250142

Wolgemuth, J. R., Abrami, P. C., Helmer, J., Savage, R., Harper, H., & Lea, T. (2014). Examining the impact of ABRACADABRA on early literacy in Northern Australia: An implementation fidelity analysis. *The Journal of Educational Research, 107*(4), 299-311. https://doi.org/10.1080/00220671.2013.823369

Appendix A

Fidelity Coding Guide

Fidelity Coding Guide

General information:

You will code for fidelity using recorded videos of teachers implementing the activities. Prior to coding, have all your materials easily accessible, including the video, coding guide, and coding sheet. As you watch the video of the teacher implementing the activity, you will code their adherence to the activity steps and four quality indicators (i.e., modeling, scaffolding, pacing, and preparation/organization) for each activity. You will also write a short summary of the activity.

Coding guidance:

***Adherence* is about whether the teacher implements the activity steps.** The teacher should adhere to the activity directions very closely. The number of steps differs by activity, but most activities have about 10 steps. Each activity step will be coded as **1 (implemented) or 0 (not implemented)**.

***Quality* indicators focus on how well the teacher is implementing the activity.** Below is a definition and the expectations for each indicator. Each indicator will be coded as **3 (high quality), 2 (moderate quality), or 1 (low quality)**. See the table below for further coding guidance.

- **Accuracy:** This is how well the teacher models pronunciations and concepts. All modeling should be accurate. Accurate modeling includes pronouncing letter sounds correctly, not adding a schwa at the end of sounds, and smoothly blending the sounds together to read words.
- **Scaffolding:** This is how well the teacher scaffolds incorrect responses. The teacher should follow the activity directions to appropriately scaffold the child to the correct response. Each child should ultimately achieve the performance objective criterion without assistance.
- **Pacing:** This is how well the teacher uses the instructional time. The teacher should keep a brisk pace, but allow sufficient time for the children to respond. There should be little time when the children are not actively involved in the activity (i.e., time devoted to behavior management or non-instructional activities should be minimal). The activity should not seem rushed.
- **Preparation/Organization:** This is how well prepared and organized the teacher appears. The teacher should appear to be familiar with the activity format and materials. The teacher should have all materials organized and accessible when needed.

	3 (High)	2 (Moderate)	1 (Low)
Accuracy	Modeling is all or nearly all (≥80%) accurate.	Modeling is accurate at least half of the time.	More than half of the modeling is incorrect.
Scaffolding	All or nearly all (≥80%) errors are scaffolded.	At least half of the errors are appropriately corrected.	More than half of the errors are not appropriately corrected.
Pacing	Good pacing, with little or no down time and not rushed.	Adequate pacing, with some down time or somewhat rushed.	The pacing was too slow or too rushed.
Preparation / Organization	Well prepared and organized, with few or no lapses.	Adequately prepared and organized, but with some lapses.	Did not appear prepared or organized.

Summary: Write 1-3 sentences about your overall impression of the activity. This summary may be more subjective than the other coding. Please include here anything that stood out about the activity.

- It is okay to include information that is already captured in the adherence and quality sections
    - "The teacher skipped some steps, which made it feel like she hadn't prepared for the activity ahead of time."
- We are *most* interested in information that is not already documented
    - "The teacher really seemed to know what she was doing, but there was one student with behavior issues and the teacher spent a lot of time redirecting that student."
    - "The teacher added a lot of information that was not part of the activity, such as providing a definition or example for each word in the word list."
- If there is anything you notice related to the group adaptation, please be sure to note it here.
    - "Instead of each student having their own worksheet, the teacher paired the students, and each pair had a worksheet."
    - "The teacher wrote each word on the board rather than each student writing each word on their own."

Appendix B

Interview Guide

Interview Guide

1. Can you tell me a little bit about your experience with the SPELL-Links intervention?
2. How confident did you feel providing the SPELL-Links intervention to your students?
    a. What factors influenced your level of confidence to provide the intervention?
3. How did you feel about the training you received?
    a. Is there anything you would have liked more guidance on or to learn more about in the training?
    b. Was the training helpful in delivering the intervention?
    c. Did you use the SPELL-Link'd website? (Tell me more about how you used it.)
    d. Did you seek out any additional resources? (Tell me more about how you used that resource.)
4. Was the SPELL-Links intervention practical for you to implement in your classroom? (Why or why not?)
    a. Did you have any problems related to providing the intervention to your students?
    b. Were you able to implement the intervention consistently (i.e., at least four days a week)? (Why or why not?)
    c. Were you able to implement the activities in the expected amount of time (i.e., 20-25 minutes)? (Why or why not?)
    d. Were there particular parts of the intervention (e.g., certain activities) that were difficult or confusing to implement?
    e. Were there any parts of the intervention that you particularly liked or disliked?
5. We know the intervention was designed for one-on-one implementation, so how did you adapt the intervention activities for your small groups?
    a. Did you use the “Group Adaptation” suggestions?
6. What are some potential problems to implementing an intervention like this in a reading intervention classroom? What kinds of support might reading intervention teachers need to implement it well?
7. How effective do you think the intervention was in helping your students learn to read and write?
    a. Were there any parts of the intervention that you thought were particularly helpful or not helpful for your students?
8. Do you plan to continue to use the intervention with your students?
    a. If so, will you use the intervention as designed or will you make changes?
    b. If not, why not?
9. Do you have any other comments, suggestions, or questions?

Table 1

Teacher Information

	Gender	Race	Highest Level of Education	Years as a Reading Interventionist	TSELI Survey Rating	TULIP Survey Score
T1	F	W	M	20	7.55	95%
T2	F	W	M	21	6.64	97%
T3	F	W	M	4	7.27	80%
T4	F	W	M	1	7.00	100%
T5	F	W	M	1	7.82	90%
T6	F	W	M	18	6.82	95%

Note. F = female; W = White; M = master's degree; TSELI = Teachers' Sense of Efficacy for Literacy Instruction (Tschannen-Moran & Johnson, 2011); TULIP = Teacher Understanding of Evidence-Based Literacy Instructional Practices (Hall et al., 2023).

Table 2

Examples of Using the CFIR to Analyze Interview Data

CFIR Construct	CFIR Definition	Interview Example	Assigned Theme
Access to Knowledge & Information	Guidance and/or training is accessible to implement and deliver the innovation	"There is a lot packed into that one day, and it seems like there were some things that we could have spent more time on" (T6)	Barrier: intervention training
Compatibility	The innovation fits with workflows, systems, and processes	"It was just too much time and energy for me to devote for one 30-minute group when I have nine groups" (T5)	Barrier: compatibility with existing practices
Mission Alignment	Implementing and delivering the innovation is in line with the overarching commitment, purpose, or goals in the inner setting (school/classroom)	"I was never going to get to where I needed to get for my second graders to be able to pass that PALS spelling" (T3)	Barrier: alignment with goals
Capability	The individual(s) has interpersonal competence, knowledge, and skills to fulfill role (innovation deliverer)	"I already had a lot of background about the skills I was teaching, so I think that helped" (T4)	Facilitator: teacher capability

Note. CFIR = Consolidated Framework for Implementation Research (Damschroder et al., 2022).

Table 3

Implementation Fidelity

	Dosage*	**Average Adherence**	**Average Quality**
T1	50 (2.5)	81%	97%
T2	48 (2.4)	88%	89%
T3	23 (2.1)	73%	92%
T4	53 (2.7)	100%	100%
T5	44 (2.2)	85%	100%
T6	65 (3.3)	80%	89%
Average	47 (2.5)	85%	95%

Note. *Dosage = total number of sessions implemented (average number of sessions implemented per week).

Table 4

Teacher Survey Ratings

	T1	T2	T3	T4	T5	T6	Average
The literacy skills targeted in the program are important.	4	5	5	4	3	5	4.33
The activities (e.g., Tap and Map, Sort It Out, Picture This) were appropriate for my students.	4	4	1	4	3	5	3.50
I was satisfied with the program.	3	3	1	2	4	3	2.67
I was confident in my ability to implement the activities.	3	3	2	4	5	3	3.33
The activities were easy to implement.	4	4	3	2	5	2	3.33
The activity materials were easy to use.	3	3	3	2	3	2	2.67
I could complete the activities in a reasonable amount of time.	4	4	3	2	4	3	3.33
I could implement the activities without extensive preparation.	3	4	4	2	4	2	3.17
The training and online platform (SPELL-Link'd) provided me with adequate support to implement the program.	2	5	1	2	4	2	2.67
The program improved my students' abilities to read words.	3	3	4	4	3	3	3.33
The program improved my students' abilities to write words.	4	3	1	2	3	3	2.67
I think continuing to use the activities with my students will make them better readers.	4	3	2	2	5	3	3.17
I think continuing to use the activities with my students will make them better writers.	4	4	1	2	5	3	3.17
I believe the learning that occurred during this program will help my students be successful in the upcoming years of education.	4	3	1	3	4	3	3.00
I have noticed that my students have generalized the skills they learned from the program to other environments and contexts.	3	3	2	3	5	2	3.00
I think this program holds promise for other students.	4	2	1	2	4	3	2.67

Note. 1 = strongly disagree; 5 = strongly agree