Language-Literacy Intervention through Telepractice for School-Age Children: A Single Case Design Study

By

Shih-Yuan Liang

Dissertation

Submitted to the Faculty of the

Graduate School of Vanderbilt University

In partial fulfillment of the requirements

For the degree of

DOCTOR OF PHILOSOPHY

in

Hearing and Speech Sciences

August 31, 2020

Nashville, Tennessee

Approved:

C. Melanie Schuele, Ph.D.

James W. Bodfish, Ph.D.

Ellen M. Kelly, Ph.D.

Christopher J. Lemons, Ph.D.

Amy M. Elleman, Ph.D.

To Mom, Dad, my sister Sheng-Yuan, and Eugene

ACKNOWLEDGEMENTS

This work was supported by the International Literacy Association (Jean S. Chall Research Fellowship) and the Vanderbilt Graduate Leadership Institute (Dissertation Enhancement Grant). The content is solely the responsibility of the author and does not necessarily represent the official views of the International Literacy Association.

Seven years of graduate study may seem long, but the years were done in a blink as I devoted myself to the work that I am passionate about. I am grateful for all the children and families I have met and worked with in the past as a clinician and as a PhD student. Each encounter made its unique contribution to my learning. I strove to improve the services that I provided to meet children’s needs in order to help them meet their potential.

I would like to thank my dissertation committee for their input throughout this project. I especially acknowledge and thank Dr. Schuele, my primary mentor, for her supervision over the years. Her meticulous thinking about research rigor, dedication to advocating for children with speech-language needs, and perspectives about life have influenced me in many ways. I am deeply grateful for all the Child Language and Literacy Lab members for their continuous support. I could never ask for better peer models and lifelong friends like you all.

Lastly, I would like to thank my family and Eugene for their love and support. Their attitudes about hard work and their care for others have fostered in me the strengths that I needed to persevere.

TABLE OF CONTENTS

Page

DEDICATION iii

ACKNOWLEDGEMENT iv

LIST OF TABLES vii

LIST OF FIGURES viii

Chapter

I. Introduction 1

Effectiveness of Telepractice Interventions for School-Age Children 3
Speech and language 3
Literacy 6
Inquiry into Quality Evidence for Telepractice Intervention Effectiveness 8
The Need for Extending Telepractice to Literacy and Language Interventions 10
Does It Make Sense: A Comprehension Monitoring Program *11*
Theory-based intervention 11
The iterative development of *Does It Make Sense* 12
Self-regulated learning in connected text reading 15
Single-strategy intervention 16
The Present Study 17

II. Method 20

Experimental Design 20
Participants 22
Recruitment 22
Descriptive assessment 26
Procedures 27
Probe assessments 27
Software and instructional materials 31
Study Conditions 32
Baseline condition 32
Intervention condition 33
Maintenance condition 37
Response Definitions and Measurement System 37
Self-correction proportion 37

Narrative comprehension 39
Interobserver Agreement (IOA) and Fidelity 41
Data Analysis 42

III. Results 45

Interobserver Agreement 46
Procedural Fidelity 46
Treatment Effects: Self-Correction Proportion 47
Self-correction proportion 47
Tau-*U* and baseline corrected Tau...55
Treatment Effects: Changes on Norm-Referenced Assessments...........................56
Treatment Effects: Story Comprehension as an Exploratory Outcome...................59

IV. Discussion..64

Evidence for Feasibility: Implementing *Does It Make Sense* via Telepractice.........64
Scoring reliability 65
Procedural fidelity...67
Treatment effectiveness: *Does It Make Sense* via telepractice 68
Does It Make Sense: What Works for Whom and Under Which Conditions 69
Automatized grapheme-phoneme knowledge and word recognition...................70
Foundational language proficiency...72
Telepractice Services as an Emerging Alternative 74
Benefits 74
Particular cautions to be exercised 81
Limitations 83
Implications and Future Directions..86
Conclusion 87

Appendix

A. Telepractice Candidacy Screening Checklist 89
B. Graphic Organizers...91
C. Sample Word Work Activities..92
D. CUBED Story Comprehension Scoring Guidelines...93
E. Procedural Fidelity Checklist...94

REFERENCES...96

LIST OF TABLES

Table Page

1. Participant Demographic Information 24

2. Descriptive Assessments..25

3. Descriptive Data on Pre-Intervention Assessment...27

4. Operationalized Definition and Examples of Reading Miscues 38

5. Operationalized Definition and Examples of Self-Correction Behaviors 40

6. Data Properties Analyzed via Visual Analysis..43

7. Interobserver Agreement Data by Conditions and Participants 47

8. Tau-*U* and Baseline Corrected Tau Analyses..56

9. Social Validity Survey 76

LIST OF FIGURES

Figure Page

1. Proportion Self-Correction by Participants..49

2. Number of Reading Miscues and Self-Corrections by Participants 50

3. Pre- and Post-Intervention Standard Scores by Reading Measure Types................58

4. CUBED Narrative Comprehension Percent Accuracy by Participants......................60

CHAPTER I

INTRODUCTION

Many struggling students in the primary grades respond adequately to a combination of Tier 1 instruction and Tier 2 intervention (e.g., McMaster et al., 2005). However, a substantial portion of these students do not benefit adequately – even when high quality interventions are implemented with fidelity. Nonresponse or inadequate response to primary prevention indicates a need for even more intensive and specialized educational services (Fuchs et al., 2014). Yet, students who need intensive support do not have routine access to adequate services. Limited access may be due to service delivery barriers such as shortage of skilled therapists, insufficient facilities in geographic locations, and transportation issues (Boswell, 2006; Theodore, 2011). Given these limitations, *telepractice* is rapidly becoming a supplement or alternative method of service delivery to traditional, on-site services in order to optimize service intensity (“ASHA’s Telepractice Practice Portal”, n.d.).

ASHA’s Telepractice Practice Portal (n.d.) defines telepractice as the application of telecommunication technology to the delivery of health-related services at a distance by linking clinician to client or clinician to clinician for assessment, intervention, and/or consultation. Although emerging evidence suggests that telepractice holds substantial promise in removing the service delivery barriers (Coufal et al., 2018; Sutherland et al., 2017; Taylor et al., 2014), knowledge gaps must be addressed before wide-scale adoption of telepractice. The present study tackled several general issues associated with telepractice services for school-age children. First, there is a paucity of research concerning the feasibility of delivering literacy and language interventions to school-age children via telepractice (Hetherton, 2013; Houge & Geier, 2009). Most telepractice intervention studies have focused on the remediation of speech sound disorders (Coufal et al., 2018; Gabel et al., 2013; Grogan-Johnson et al., 2010; Grogan-Johnson et al., 2011; Grogan-Johnson et al., 2013). Second, many telepractice-based intervention studies are considered poor quality resulting from flawed research designs and/or inadequate measurement systems (Rudolph & Rudolph, 2015; Weidner & Lowman, 2019). Very few intervention studies have reported fidelity of intervention and interrater agreement (Grogan-Johnson et al., 2013; Hetherton, 2013). Treatment outcomes of telepractice and on-site services were often compared on the basis of measurement systems that yield results that are difficult to aggregate or compare across groups (e.g., expert rating; Coufal et al., 2018; Gabe et al., 2013; Grogan-Johnson et al., 2010). Most measures lack the sensitivity to detect the within-individual incremental changes expected in intervention. Third, although empirical investigations suggest that delivering standardized assessment via telepractice can be feasible and reliable (Taylor et al., 2014), less is known about whether progress-monitoring instruments, a critical feature of response-to-intervention (RTI), can be administered with fidelity and scored reliably via telepractice.

In the following sections, we evaluate the empirical evidence of telepractice intervention research that pertains primarily to pediatric speech, language, and/or literacy disorders in school-age populations. Each of the aforementioned issues are discussed in detail.

Effectiveness of Telepractice Interventions for School-Age Children

Telepractice has achieved initial success in obtaining persuasive evidence of reliable assessment with school-age populations (Eriks-Brophy et al., 2008; Ciccia et al., 2011; Taylor et al., 2014; Waite et al., 2006; Waite, Theodoros et al., 2010). Overall, the reported total percentage agreement between telepractice and face-to-face therapists (mostly scoring simultaneously) on language assessments was high with a range between 98% – 100% and kappas (i.e., a point-by-point agreement index) between .88–.99 (criteria: .60). Results of articulation assessments revealed acceptable percentage agreement between 80% – 92% (Eriks-Brophy et al., 2008; Waite et al., 2010). A few limitations were noted, for instance, reduced agreement for voicing, high-frequency sounds (e.g., /s/), and sounds without visible articulation (e.g., /k/ and /g/; Taylor et al., 2014).

Questions remain about whether telepractice interventions produce treatment outcomes that are comparable with traditional, face-to-face service delivery. Across studies, speech sounds were the most common treatment targets (Coufal et al., 2018; Gabel et al., 2013; Grogan-Johnson et al., 2010; Grogan-Johnson et al., 2011; Grogan-Johnson et al., 2013) whereas language and literacy skills less frequently have been the treatment targets (Hetherton, 2013; Houge & Geier, 2009).

Speech and Language

Of the five identified speech-sound intervention studies, intact groups comparison was the most commonly-used research design. Across studies intervention dose frequency was 1 or 2 sessions weekly (20-30 minutes); duration of intervention ranged from 5 weeks to 8 months. Some studies assigned participants to one of two conditions: on-site or telepractice treatment. For instance, Grogan-Johnson et al. (2011) assigned 13 6- to 9-year-old children to one of two service delivery models and targeted Individualized Education Program (IEP) speech goals in one-on-one sessions. Telepractice services yielded outcomes comparable to on-site services for: (a) pre- and post-intervention scores on the Goldman-Fristoe Test of Articulation–Second Edition (GFTA-2; Goldman, 2000), (b) sound probes (increased accuracy from baseline), and (c) amount of change documented in quarterly progress reports. Grogan-Johnson et al. (2010) conducted a within-subject comparison study with 17 participants receiving telepractice services followed by on-site services; another 17 participants were assigned to the reversed treatment order. The two groups did not differ in GFTA-2 outcome scores and accomplished a comparable proportion of IEP objectives. In an exit survey, the students, parents, and educational/administration staff were generally satisfied with the telepractice services.

Several studies compared outcomes for a telepractice group with an extant outcome database for on-site services (Coufal et al., 2018; Gabel et al., 2013; Grogan-Johnson et al., 2010). The database involved the ASHA's National Outcome Measurement System (NOMS) Functional Communication Measures (FCMs), which comprise a series of disorder-specific 7-point rating scales to quantify child progress. For instance, Coufal et al. (2018) compared the median change scores on speech sound production (i.e., the difference between the initial and final NOMS FCM scores) between 1,331 ASHA NOMS cases and 428 telepractice cases, 6 to 9 years of age. These studies reported no significant difference in the outcomes (i.e., median change scores) between the telepractice group and the outcome database. Thus, the authors concluded that telepractice yielded treatment outcomes that were comparable to on-site services for children exhibiting speech sound disorders.

Regardless of the methodological variation, the available evidence seems to be overwhelmingly positive – all studies reported equivalent or greater improvement in the telepractice group relative to the on-site group. Nevertheless, a few limitations are noteworthy. First, treatments and measurement contexts in telepractice intervention research generally have been defined loosely. For instance, several studies provided speech as well as language services but none of the studies described exactly how language services were delivered. Furthermore, only Grogan-Johnson et al. (2013) documented the implementation fidelity of telepractice intervention. Less structured intervention procedures or measurement contexts increase the likelihood of measurement error and extraneous factors that could potentially weaken the inferences drawn by the investigators (Yoder et al., 2018). In a systematic review of telepractice, Rudolph and Rudolph (2015) disclosed that across studies effect sizes were negatively correlated with study quality – the higher the quality of study, the smaller the effect size (in favor of telepractice). The authors conducted a meta-analysis based on three studies of the highest quality, which yielded a small and nonsignificant aggregated effect size ($d = -0.18$) with a 95% confidence interval that included zero (-0.56 to 0.20). Thus, the true effect size could range from a moderate effect size in favor of on-site treatment to a small effect in favor of telepractice.

A second limitation pertains to the choice of analysis. Nearly all studies shared the same goal: to determine if a new mode of intervention delivery, telepractice, had equivalent or noninferior effectiveness to the traditional mode of intervention delivery. If so, equivalence/noninferiority testing, a specific type of comparison should be conducted rather than the traditional (two-sided) comparative approach (Weidner & Lowman, 2019). Walker and Nowacki (2010) pointed out that, when seeking evidence for equivalence or noninferiority, the equivalence/noninferiority test imposes a stringent equivalence margin. In contrast, in two-sided comparative studies the burden of proof rests on the research hypothesis of ‘difference’ between conditions. If the evidence is not strong enough in favor of the alternative hypothesis (i.e., difference is present), equality cannot be ruled out. In contrast, the null and research hypotheses in equivalence testing are reversed: If the evidence in favor of equivalence is not strong enough, nonequivalence cannot be ruled out. Equivalence/noninferiority cannot be established when the testing has a confidence bound outside of the equivalence margin, even if the therapy means are "nonsignificantly different". Noninferiority testing may be one of the most appropriate telepractice study designs; nevertheless, none of the aforementioned studies took this approach (Weidner & Lowman, 2019).

Literacy

There have been very few studies on delivering literacy interventions via telepractice. In an unpublished doctoral dissertation, Hetherton (2013) reported a single-subject research study, which combined a multiple-baseline design across behaviors with an alternating treatment design with eight 5- to 7-year-old children. All intervention was provided in a school-based setting but instruction was delivered alternately between the face-to-face and telepractice platforms, sequentially targeting letter naming, letter sounds, and decoding accuracy (i.e., multiple baseline) within each participant. Various analyses including improvement rates, percentage of non-overlapping data, and visual analysis indicated comparable outcomes between face-to-face and telepractice settings across all skills and across all participants. The reported average interobserver agreement on letter naming, letter sound naming and decoding probes was 85% with a range of 83% – 100%.

Houge and Geier (2009) reported that tutors at a university reading clinic delivered telepractice literacy interventions to 9- to 18-year-old adolescents who had deficits in one or more literacy domains (i.e., word recognition, decoding, reading fluency, comprehension). Of the 61 students, 44 received the telepractice intervention while at school and 17 while at home. A comparison of pre- and post-test scores on norm-referenced reading and spelling assessments indicated statistically significant change, with small to large effect sizes, in all four areas: reading accuracy (Cohen's d = 1.78), fluency (d = 0.22), comprehension (d = 0.30), and spelling (d = 0.32). This study did not have a control group and/or randomization. Further, it was unclear about the duration of intervention needed for the observed changes in reading outcomes to occur.

An unpublished pilot study conducted by Waite et al. (2010) evaluated the feasibility of delivering a telepractice literacy intervention to eight 8- to 9-year-old children who struggled with word decoding. Intervention was provided in a laboratory setting between two rooms for two 1-hour sessions per week for 10 weeks. Group pre-post means revealed statistically significant gains in nonword spelling and reading accuracy. However, beyond intervention feasibility, limited conclusions can be drawn because study details were limited, for example with respect to treatment implementation fidelity, assessment administration fidelity, and scoring reliability.

Despite emerging evidence in favor of telepractice as a viable means of delivering literacy interventions, we note a few limitations in so far as a lack of sound research design and insufficient details on implementation fidelity. For instance, Hetherton (2013) used a combination single-subject design – embedding an alternating treatment component within a multiple-baseline design. It is advisable to use two different but functionally similar sets of target letters and sounds in the telepractice and face-to-face setting. The author, nevertheless, violated this expectation. Houge and Geier (2009) would best be considered a case report on a group as there was no control group. None of the literacy intervention studies evaluated fidelity of treatment implementation. Because none of these studies are of reasonable methodological quality, we cannot firmly conclude that telepractice works effectively in providing literacy services as intended.

Whereas most prior telepractice intervention studies sought to answer the question — whether or not telepractice is as effective as face-to-face services (i.e., *noninferior outcomes*) – the present study focused on a fundamental question: *Is telepractice a feasible form for providing literacy services?* This question would be best answered by a treatment-control comparison or a single case design study.

Inquiry into Quality Evidence for Telepractice Intervention Effectiveness

Rudolph and Rudolph's review (2005) necessitates the need for additional studies with quality measurement systems and sound research methodology to provide support for the effectiveness of telepractice.

Nearly all group-design studies reported that on-site and telepractice services yielded comparable treatment outcomes (Coufal et al., 2018; Grogan-Johnson et al., 2010; Grogan-Johnson et al., 2011; Grogan-Johnson et al., 2013). Telepractice treatment outcomes were frequently measured by expert ratings (e.g., ASHA's NOMS), norm-referenced assessments, and qualitative quarterly reports. Although these measurements provide valuable information about a child's functional outcomes, a few limitations must be noted. First, Likert-like rating such as NOMS requires the rater to have substantial knowledge concerning the construct of interest and may involve somewhat arbitrary decisions regarding the amount of evidence needed to increase the score on the rating scale. Adequate knowledge to ensure reliable ratings among nationwide users of NOMS cannot be assumed. Second, expert ratings provide a rather restricted range of potential scores and thus, the amount of change may not have been adequately captured obscuring differences between conditions. Third, norm-referenced assessments may provide only gross estimates of improvement and like expert ratings, may not capture within-individual incremental growth. These measurement limitations provide an impetus for introducing other measurement types (e.g., count-coding progress monitoring tools) that allow greater sensitivity to change and alignment between treatment and instruments.

In addition to a rising concern about measurement, lack of documentation of treatment fidelity is an important methodological issue that reduces the credibility of reported effects associated with telepractice interventions. Fidelity refers to the extent to which the intervention is delivered as it was intended. Fidelity for an intervention program includes two components – *fidelity of intervention* and *procedural fidelity of test administration*. In fact, treatment procedures and measurement contexts have often been defined loosely in telepractice intervention studies. Very few telepractice intervention studies documented adequate information about fidelity and interobserver reliability (Grogan-Johnson et al., 2013). Although exploring treatment effects is a major goal of a feasibility study on telepractice, we argue that treatment fidelity and scoring reliability are also vital aspects of feasibility that must be considered prior to interpreting treatment effects.

The Need for Extending Telepractice to Literacy and Language Interventions

Limited research on telepractice literacy intervention may reflect a general attitude of speech-language pathologists (SLPs; e.g., confidence) towards providing literacy services to school-age populations. Although the ASHA position statement (n.d.) asserts the roles and responsibilities of SLPs in supporting literacy, a national survey of over 600 school-based SLPs (Katz et al., 2010) reported that only 50% felt prepared to help struggling readers. Furthermore, 50% believed that literacy should not be in their scope of practice. Thus, intentional effort is required for extending the scope of practice/telepractice from conventional speech therapy to instructional programs targeting the language underpinnings that are foundational to curricular content related to literacy development (Scarborough, 2005; Nelson, 2010; Schuele, 2017). The present study is a realization of such intentional effort with a focus on child comprehension monitoring contextualized in oral text reading. We examined the effectiveness of a comprehension monitoring intervention *Does It Make Sense* (Liang et al., 2017; Liang & Schuele, 2019), which has undergone several iterative design changes. We considered comprehension monitoring of oral reading as a venue where language and literacy meet – when employing the comprehension monitoring strategy in oral text reading, one must draw on not only literacy (e.g., decoding) but also spoken language knowledge to evaluate the adequacy of the message produced orally. Therefore, we characterized our intervention as a language-literacy intervention.

Does It Make Sense: A Comprehension Monitoring Program

Theory-Based Intervention

Design of the *Does It Make Sense* program was theoretically driven by Dollaghan's dual-stage theory of comprehension monitoring (Dollaghan, 1987). Successful comprehension monitoring requires (a) *detecting* occasions when one does not understand and (b) *reacting* effectively when such difficulties are recognized. We noted that proficient readers display effective comprehension monitoring. They make few errors due to strong word reading skills, but more importantly, their reading suggests that they build and monitor comprehension as they read. That is, they take actions to problem-solve any encountered challenges (words, comprehension, etc.), and have many successful self-corrections resulting in efficient processing of information in print. By contrast, struggling readers have high error rates in word reading; they produce few attempts, of any kind, at self-initiated repairs, and hence, rarely self-correct. Likewise, they show few outward behaviors that suggest attempts to construct meaning from printed text (Clay, 2001). We dubbed these struggling readers "conveyor-belt readers" as when performing oral reading, they often proceed on autopilot, oblivious to the comprehension breakdowns that result from their reading miscues. Across studies, proficient readers self-correct 20 to 35% of their reading miscues when provided with instructional- or independent-level text, whereas struggling readers self-correct only about 5 to 15% of reading miscues (Chinn et al., 1993; Clay, 1969; Pflaum, 1979). These observations became an impetus for us to develop a strategy that can effectively break this "conveyor belt" reading pattern.

The Iterative Development of Does It Make Sense

Developing effective interventions that improve outcomes for young at-risk children is an iterative process. *Does It Make Sense* has undergone three iterative cycles of program development to improve its effectiveness, feasibility, and usability. In 2016 we piloted our first iteration of *Does It Does Sense* with three teacher-identified struggling readers in the primary grades. In this iteration, we simply asked the students to intentionally pose the question, "Does it make sense?" at the end of each sentence. That is, when a child made a reading miscue within a sentence, we did not stop the child at the point of miscue. We expected that intentionally asking this question would facilitate attention to the meaning of what was read. Thus, explicit prompting would urge the struggling reader to focus on meaning and make an initial effort to understand whether his/her orally-rendering of the text made sense, that is, was meaningful. Non-meaningful messages, those that did not make sense, provided opportunities to detect occurrences of misunderstanding. Our hypothesis was that prior to intervention the child approached the passage as a word-by-word reading task, not a task of creating meaning. Our theory of change was that practice with the *Does It Make Sense* strategy would facilitate the child's movement toward reading for understanding and this would be evidenced by increased self-corrections in independent oral reading.

We soon realized that simply having students recognize a comprehension breakdown or a word challenge did not always ensure comprehension monitoring would proceed further to evaluation and successful self-initiated repairs. Struggling readers may know many grapheme-phoneme correspondences yet inefficiently exploit this knowledge to sound out words unless prompted. Instead, they may rely on more familiar yet less effective word-solving strategies such as guessing from the context or the visual appearance of the unidentified word. Thus, in our second iteration we added *'fix-up tools'* to the *Does It Make Sense* strategy. The fix-up tools include reading again, sounding out, and parsing the word into smaller parts to facilitate the decoding process, all of which are evidence-based word-solving strategies highly recommended in a What Work Clearinghouse practice guide (Shanahan et al., 2010).

For the current telepractice study, a few additional adaptations were introduced: (a) vocabulary preview, (b) unobtrusive intervention when encountering words with low decodability and/or miscues that do not cause significant changes to meaning, and (c) opportunities for repeated reading. We noted that in our last iteration, children were more likely to apply the *Does It Make Sense* strategy in a cursory manner (i.e., uncritically accepting non-meaningful sentences) when word decoding and lexical inferencing placed heavy demands on their reading systems. In the present study, prior to practicing the strategy with connected text within a session, the interventionist introduced a set of general and domain-specific vocabulary selected from the text, drew child attention to effective word-solving strategies and discussed lexical meanings. The rationale was to increase the likelihood that later, when students encounter these words in the text, the phonological and meaning representations are already activated and readily retrievable. More cognitive resources may be freed up from word decoding and lexical/semantic retrieval and be reallocated to in-depth evaluation of the constructed meaning representation (Perfetti, 1992). Likewise, opportunities for repeated reading benefit slow and disfluent readers. When struggling readers first encounter an unfamiliar text, laborious word reading nearly always depletes a substantial amount of mental resources, leaving little for comprehension and comprehension monitoring (LaBerge & Samuels, 1974). Nevertheless, such challenges may be overcome in repeated reading (Therrien, 2004) as it builds accuracy and speed to a point where comprehension is possible. Lastly, to promote swift instruction, the interventionist imposed stronger scaffolding if the recognized reading miscue carried a low-frequency spelling-sound pattern that was unlikely to be solved independently by the students (e.g., *enough*). The interventionist also adopted a less obtrusive intervention when miscues resulted in minimal change to meaning, for instance, omission of grammatical markers (e.g., past-tense -*ed*, third-person singular -*s*). Instead of prompting the child to re-evaluate his/her understanding, the interventionist simply recast the child’s rendered sentence (i.e., restating the sentence in a grammatically complete and accurate sentence; Cleave et al., 2015) while visually highlighting where the miscue occurred. This modification reduces potential confusion with respect to ungrammatical sentences that nevertheless convey the essential meaning (e.g., I miss/*ed the school bus this morning).

Self-Regulated Learning in Connected Text Reading

Connected text reading provides the best venue where beginning readers can practice and integrate component skills learned in isolation (e.g., grapheme-phoneme correspondences, orthographic patterns) and develop self-monitoring skills. Nevertheless, in a study of 107 first- and second-grade classrooms, Foorman et al. (2006) reported that teachers on average allocate limited instructional time (approximately 5 to 8%) to connected text reading within a 90-minute language arts block. Moreover, Allington (2013) argued that the design of reading lessons differs for good and poor readers in that good readers get assigned more reading activities whereas poor readers get more work on tasks that require little reading (e.g., worksheets). In addition to limiting authentic text reading, teachers and parents interrupt proportionally more often when listening to poor readers read than with proficient readers. They nearly always supply the word immediately at points of difficulty; even a child hesitation can prompt an interruption (Allington, 1980). Such corrective feedback (i.e., immediate word/phrase supply), referred to as terminal feedback (Evans et al., 1998; Evans et al., 2003; Mansell et al., 2005), invites no opportunities for independent word study, self-monitoring, and self-initiated repairs. Thus, it is perhaps not surprising that some struggling readers become conveyer belt readers.

The *Does It Make Sense* strategy addresses these issues in the following ways. First, this comprehension monitoring strategy is operated in the context of oral text reading. Second, instead of interrupting the child at each point of miscue and immediately supplying the word, the interventionist draws child attention to the meaning of the sentences by using sentential boundaries as checkpoints. The latter encourages comprehension monitoring, which in turn, leads to more self-corrections in the course of oral reading.

Single-Strategy Intervention

Prior research on comprehension monitoring in the context of oral reading nearly always included comprehension monitoring strategies (e.g., clarifying, fix-up) as part of a multi-strategy intervention package (e.g., Brown, 2008; Denton et al., 2013; Paris, Cross, & Lipson, 1984). One limitation of these multi-strategy programs remains that isolating the effect associated with a single strategy is not possible. The effectiveness of the embedded comprehension monitoring strategies must be vetted along with other reading strategies.

Unlike prior research studies, we characterize *Does It Make Sense* as a single-strategy intervention, which comprises two essential steps – *detection* and *reaction*. For struggling readers, all the remediation efforts would eventually lead to increased flexibility in selecting appropriate tactics from a variety that best fit the text, purpose, and occasion. Currently there is not enough evidence to advocate the use of a multi-strategy instruction over single-strategy instruction or vice versa (Shanahan et al., 2010). However, implementing a theoretically-driven, single-strategy intervention allows us to draw stronger inference about the effectiveness of our comprehension monitoring program implemented through telepractice.

The Present Study

In the current study, initiated in the summer of 2019, we scaled up the *Does It Make Sense* project further by launching it through telepractice. Summer reading loss has been documented extensively for low-income and minority students (Alexander et al., 2007; Cooper et al., 1996; Chin & Phillips, 2004). Prior research indicates that simply providing access to books is not sufficient to prevent reading loss (Byrnes, 2000; Carver & Leibert, 1995; Kim & White, 2008; Stahl, 2004). Specialized instruction by reading professionals or directed scaffolding by parents is a necessary ingredient for closing the achievement gap (Kim & White, 2008). In the present study, the *Does It Make Sense* program provided intensive, one-on-one home-based reading tutoring via telepractice, with a dosage of 2 to 3 weekly 45- to 60-minute sessions. Such intensified instruction, perhaps, is only achievable through telepractice coupled with close professional-family collaboration. Most teachers and school-based interventionists struggle on a daily basis to manage their workload and maximize productive time that actually promotes learning for struggling readers. We certainly empathize with these daily struggles; hence, we explored telepractice as an innovative tool for optimizing services in qualitative (e.g., quality of instruction, material management, etc.) and quantitative (e.g., maximizing productive time and optimizing intensity of treatment) manners rather than a direct replacement of any of the regular and special educational services that the participants concurrently received from their schools.

We addressed the limitations noted in prior telepractice research by conducting a single-subject design study to establish a functional relation between a reading comprehension monitoring intervention delivered through telepractice and child self-correction proportion in independent oral reading (a proxy indicator of improved reading comprehension). The current study advanced the evidence base in several ways. First, using a single-subject, multiple-probe design across participants enabled a rigorous evaluation of three critical dimensions that define feasibility: (a) treatment effectiveness (b) fidelity for intervention implementation, and (c) scoring reliability. These aspects are vital to corroborating the postulation that telepractice is equivalent to the quality of services provided face-to-face. Second, we introduced a weekly-administered progress-monitoring system using a count-coding measurement as part of the single-case design. As compared with expert rating scales or third-party reports, count coding provides a larger range of potential scores. These properties potentially provide a more sensitive measure of individuals' incremental growth over time and allow for more rigorous definition of interobserver agreement. Careful progress monitoring aids in and drives instructional decisions making (Fuchs, 2017). To the best of our knowledge, none of the extant research on telepractice included progress monitoring as an integral part of the study design.

We addressed two falsifiable research questions.

(1) Can an online interventionist deliver a language-literacy intervention via telepractice with adequate procedural fidelity ($\geq$ 80%) and administer and evaluate child behavior via telepractice with a progress monitoring assessment?

We hypothesized that: (a) an interventionist can implement the instructional program delivered via telepractice with adequate procedural fidelity and (b) reading progress-monitoring assessments delivered via telepractice can be administered with fidelity and scored reliably.

(2) Is there a functional relation between language-literacy intervention delivered via telepractice and child self-corrections in independent oral reading?

We hypothesized that the language and literacy intervention delivered via telepractice would increase a child's reading self-correction proportion measured in independent oral reading probes.

Additionally, we conducted a pre-post intervention comparison of other literacy outcomes (e.g., word/non-word reading, reading comprehension) and intermittently assessed child story comprehension using the CUBED Narrative Language Measures (Petersen & Spencer, 2016). These data supplemented results derived from the single-case design to answer an exploratory research question: Does a treatment effect demonstrated in the primary outcome (i.e., self-correction proportion) transfer to a distal outcome measure of reading comprehension? Because this question was exploratory, we did not make an a priori hypothesis of change.

CHAPTER II

METHOD

The conduct of this study was approved by the Institutional Review Board at Vanderbilt University. All procedures were administered by the primary study investigator, who is a doctoral student specializing in speech-language pathology with multiple years of psychoeducational assessment and intervention training.

Experimental Design

The study employed a single-case, multiple probe design across four primary-school children identified with a reading disability with/out a language impairment (per school records). We examined the functional relation between a telepractice-based literacy intervention (i.e., independent variable) and child reading self-correction proportion (i.e., dependent variable). A multiple-probe design was preferred over other single-case methods (e.g., adapted alternating treatment design) because the objects of measurement are considered generalized person characteristics rather than context-bound behaviors: (a) target outcome, reading self-correction behavior, and (b exploratory secondary outcome, reading comprehension. Thus, changes in the target outcome due to treatment is not readily reversible to baseline level when intervention activities are withdrawn.

The multiple-probe design comprised three conditions: (a) a baseline condition, wherein participants received business-as-usual practice, (b) a 7- to 10-week intervention condition, and (a) a maintenance condition. Prior to the baseline condition and in the maintenance condition, each participant completed a literacy and language battery with norm-referenced measures. We captured growth in spontaneous self-corrections via a progress-monitoring measurement system as part of the single-case design. Additionally, we included two exploratory analyses: (a) a pre-post comparison of reading outcomes assessed by static, norm-referenced measures and (b) change in story comprehension assessed intermittently by the CUBED NLM.

Initiation of intervention was intentionally staggered such that the length of baseline varied across participants. The order whereby each participant entered intervention was selected randomly by a faculty member familiar with single-case research design but blind to the study purpose. Participant 1 entered the intervention after demonstrating stable performance on the primary dependent variable (i.e., self-correction proportion). In the meantime, other participants remained in the baseline condition. Participant 2 entered the intervention after maintaining stable performance on the primary dependent variable concurrent with the first participant demonstrating observable response to intervention (i.e., increased self-correction proportion). The same criteria applied to the remaining participants.

The dependent variable was self-correction proportion, measured as the ratio of self-correction behaviors in response to reading miscues over total reading miscues in a four-minute reading probe. The four-minute probe assessment was collected twice weekly when there were three sessions scheduled per week or once weekly when there were two sessions scheduled per week throughout the project. Child self-correction proportion drove the decision making in intervention initiation. We monitored change in this dependent variable across participants to determine if the desirable change in the targeted skill was attributable to the intervention and if the intervention effect replicated across participants.

The CUBED story comprehension composite, which did not drive any intervention decisions, was monitored alongside self-correction proportion to evaluate transferred effects on reading comprehension. In the CUBED NLM, each participant answers six narrative comprehension questions following his/her reading of a grade-level passage. The CUBED NLM was collected intermittently (typically every two weeks) in the baseline and intervention conditions and once weekly in the maintenance condition. We monitored change in this dependent variable across participants to provide a preliminary consideration of how the intervention might influence reading comprehension as a more distal evaluation of treatment effects.

Participants

Recruitment

We employed a screening procedure to identify children in Grades 1 to 4 who demonstrated intensive literacy needs with/out a confirmed diagnosis of language impairment (per school reports). In the first stage, we administered a grade-specific reading probe from the Dynamic Indicators of Basic Early Literacy Skills 8th Edition (DIBELS 8th; University of Oregon, 2018) and a reading narrative probe form the

CUBED NLM. We audio-recorded child reading and narration and transcribed online any instances of reading miscues and self-corrections. We verified these notes later while replaying the video-recordings and made corrections as needed. We then divided the number of self-corrections by the number of reading miscues to form a self-correction proportion. At the screening stage we sought to eliminate children who produced insufficient number of reading miscues (i.e., < 10; limited opportunity for observing self-corrections) and/or self-corrected more than 30% of their reading miscues and/or produced reading miscues that mostly resulted in low meaning change to the sentence.[1] Children who produced mainly low meaning-change miscues, for instance omitting inflectional morphology (e.g., past-tense *ed*), were excluded because these errors typically do not disrupt the meaning of the sentence significantly. These grammatical errors are less likely to be picked up by the children when using the *Does It Make Sense* strategy. Additional exclusionary criteria included: (a) nonverbal IQ standard score < 80 and (b) no reported history of autism, hearing loss, visual impairment, or other neurological disorders. The screening process proceeded until five participants who met the study criteria were identified jointly by the first author and the dissertation committee chair. Table 1 provides an overview of participants' demographic information.

With the five participants identified at the first stage, we administered a more thorough descriptive assessment battery (Table 2). The descriptive assessment was

[1] Prior research on child oral reading miscues has shown that typical readers on average self-correct about a third of their reading miscues whereas struggling readers self-correct less than 20% of their reading miscues (Chinn, Waggoner, Anderson, Schommer, & Wilkinson, 1993; Clay, 1969; Pflaum, 1979).

aimed at confirming reading and/or language problems and characterizing the nature of such problems.

Table 1

Participant Demographic Information

Variable	Participant			
	Kylee	Michael	Adam	Kevin
Age (years;months)	7;8	8;1	7;0	9;0
Race	African-American	African-American	African-American	African-American
Sex	Female	Male	Male	Male
Bilingual status	No	No	No	No
Hearing loss	No	No	No	No
Visual impairment	No	No	Astigmatism	No
Parent-reported learning or speech-language concerns	Reading comprehension	Learning disability, struggles with word decoding	Speech/language	No
Type of school child attended	Religion-Affiliated School	Public Charter School	Public Charter School	Public Charter School
History of Special Education and Speech/Language Services	No	Yes (IEP) No services in preschool	Yes (IEP) Services began in preschool	No
Grade Retention	No	Yes (1st grade)	No	Yes (2nd grade)
Family history of speech, language, hearing, learning	Paternal cousin	Not reported	Great grandmother-stuttering; father-dyslexia	Not reported
Other medical history	No	Not reported	Preterm Birth; initially had trouble breathing	Not reported
Years of maternal education	16	12	17	15

Note: Pseudonyms were assigned to participants

Table 2

Descriptive Assessments

Measure	Battery	Task Description	Score
Woodcock Reading Mastery Test (WRMT–III)			
Word Identification	Word reading	Students read printed real English words (untimed); words increase in complexity	grade-referenced standard score (*M* = 100; *SD* = 15)
Word Attack	Word decoding	Students read pronounceable nonsense words (untimed);); words increase in complexity	grade-referenced standard score (*M* = 100; *SD* = 15)
Passage Comprehension	Reading comprehension	Students read a short passage and identify a key word missing from the passage	grade-referenced standard score (*M* = 100; *SD* = 15)
Wechsler Individual Achievement Test (WIAT–III)			
Reading Comprehension	Reading comprehension	Students read each passage (fictional or informational text) silently or aloud and then answer literal and inferential comprehension questions read aloud by the examiner	grade-referenced standard score (*M* = 100; *SD* = 15)
Clinical Evaluation of Language Fundamentals (CELF–5)			
Sentence Comprehension	Language	Students select the pictures that illustrates the referential meaning of the sentences read aloud by the examiner; sentences increase in length and complexity	Age-referenced standard score (*M* = 100; *SD* = 15)
Word Structure	Language	Students are asked to complete unfinished spoken sentences by applying their morphological knowledge or using appropriate pronouns to refer to people, objects and possessive relationships	Age-referenced standard score (*M* = 100; *SD* = 15)
Formulated Sentences	Language	Students formulate complete, semantically and grammatically correct, spoken sentences, using given words (e.g., car, because) and contextual constraints imposed by illustrations	Age-referenced standard score (*M* = 100; *SD* = 15)
Recalling Sentences	Language	Students imitate English sentences with increasing length and grammatical complexity	Age-referenced standard score (*M* = 100; *SD* = 15)
Test of Nonverbal Intelligence (TONI–4)	Nonverbal intelligence	Matrix reasoning: students respond with only simple but meaningful gestures such as pointing	Age-referenced standard score (*M* = 100; *SD* = 15)

Upon completion of the descriptive assessments, we began technology training with the five participants by delivering mock tutoring lesson(s) via ZOOM® (i.e., the videoconference platform used in this study) to explore participant telepractice candidacy. Candidacy was determined on the basis of a behavior checklist filled out by the examiner during descriptive assessment and technology training (see Appendix A). Note that one family withdrew their child from the study shortly after he entered the intervention phase. Though this child passed the candidacy screening including the mock training, the incompatibility of his complicated learning needs with the telepractice format become apparent once the telepractice sessions began. Thus, we report the data collected from the remaining four participants who completed all the study conditions. In the discussion section, we consider how multiple areas of developmental needs (e.g., language, gross/fine motor, attention) can influence a child's success in telepractice services.

Descriptive Assessment

Several assessment measures were administered pre-intervention and post-intervention to describe participants and for exploratory description of changes on norm-referenced measures. Prior to the baseline condition, participants completed a language and literacy battery (see Table 2) administered face-to-face, with norm-referenced measures capturing three skill areas: (a) word-level reading, (b) reading comprehension, and (c) oral language abilities. During the maintenance condition, we re-administered the norm-referenced measures that pertained primarily to word-level reading and reading comprehension. Preintervention descriptive assessment results are presented in Table 3.

Table 3

Descriptive Data on Pre-Intervention Assessment

	Standard Scores and Scaled Scores			
Assessment	Kylee (Age: 7;8)	Michael (Age: 8;1)	Adam (Age: 7;0)	Kevin (Age: 9;0)
WRMT-III				
Word Identification	75	83	83	89
Word Attack	79	83	98	75
Passage Comprehension	86	83	86	92
WIAT-III				
Reading Comprehension	87	90	90	91
TOWRE-2				
Word Reading Efficiency	75	71	77	82
Sight Word Efficiency	87	81	81	91
Phonemic Decoding Efficiency	65	63	75	75
CELF-5				
Core Language	75	90	81	74
Sentence Comprehension[a]	7	13	6	N/A
Word Structure[a]	4	5	3	N/A
Formulated Sentence[a]	7	7	6	6
Recall Sentences[a]	4	4	6	4
Word Classes[a]	N/A	N/A	N/A	9
Semantic Relationships[a]	N/A	N/A	N/A	8
TONI-4	105	95	102	98

Note. a = Scaled score (M = 10, SD = 3); the rest of the test scores are presented as standard scores (M = 100, SD = 15). WRMT-III = Woodcock Reading Mastery Test, 3rd Edition (Woodcock, 2011); WIAT-III = Wechsler Individual Achievement Test, 3rd Edition (Wechsler, 2009); TOWRE-2 = Test of Word Reading Efficiency, 2nd Edition (Torgesen et al., 2012); CELF-5 = Clinical Evaluation of Language Fundamentals, 5th Edition (Wiig et al., 2003); TONI-4 = Test of Nonverbal Intelligence, 4th Edition (Brown et al., 2010).

Procedures

Probe Assessments

Two progress-monitoring instruments were administered via telepractice to provide repeated measurement of the skills of interest across time and conditions. The primary outcome variable, self-correction proportion, was calculated from a child's performance on a four-minute reading probe administered once or twice weekly, depending on the number of weekly sessions scheduled across conditions. The secondary outcome variable, story comprehension composite (a proxy indicator of reading comprehension), was derived from the CUBED NLM administered intermittently (typically every two weeks) in the baseline and intervention condition and weekly in the maintenance condition.

Four-minute reading probe (primary). The four-minute reading probe utilized the passages from the DIBELS 8th Progress Monitoring Oral Reading subtest (University of Oregon, 2018). This reading curriculum-based measurement comprises 20 alternate forms engineered with comparable difficulty. Instead of using the conventional 1-minute protocol of the DIBELS 8th, the examiner asked the child to read for four minutes and provided no corrective feedback during child reading. The examiner presented an unnumbered copy of a passage by sharing the computer screen, and said: *I want you to read this story out loud. Read it the best you can. I can't help you but you can take as much time as you need to read it. If you're not sure how to say a word, give it your best guess and continue reading. Remember, a best guess is the best idea you have for what the word says. After a few minutes, I will ask you to stop. I might ask you what this passage is about.* If, while reading the passage, the child sought support from the examiner or a paused for 3 seconds, the examiner reminded the child that no help could be provided and encouraged the child to continue reading. After the child read for 4 minutes or finished reading the entire passage within 4 minutes, the examiner said, *Thank you for reading. Now you tell me what you just read.* This request along with the last sentence in the directions provided prior to the probe was an implicit prompt for the child to read for meaning. No other support was provided that might have suggested to the child that he or she pay attention to meaning while reading the DIBELS passage. In the course of the probe and the child talking about what he or she read, the examiner provided neutral prompts as she deemed appropriate: *It's OK. Just do your best. I can't help, but you can just tell the parts you remember. Tell me more about it. Is there anything else you'd like to add to what you just told me?* The four-minute reading probe was administered minimally in the first session of each week to monitor a child's progress in self-correction proportion over time. It was always administered at the outset of the session.

CUBED Narrative Language Measurement (exploratory). We administered probes from the CUBED NLM: Reading intermittently during the baseline condition and the intervention condition (typically every two weeks) and on a weekly basis in the maintenance condition to monitor growth in reading comprehension over time. The CUBED NLM comprises 32 alternate forms. Each NLM probe consists of three sections: (a) narrative retell, (b) story comprehension questions, and (c) inferential word learning. We collected narrative retells and responses to story comprehension questions as exploratory data but analyzed only responses to story comprehension questions as a proxy indicator of reading comprehension. The examiner asked the child to read a brief story that contained basic story grammar elements (e.g., character, setting, problem, attempt, consequence, ending), *Please read this out loud. Do your best reading. I'll help you if you need it. When you're done, I might ask you to tell me the story.* Unlike the four-minute reading probes, if the child requested help reading a word or paused for 3 seconds without making any attempt to read the word, the examiner supplied the word to the child. After the child finished reading, the examiner said, *Thanks for reading. Now you tell me that story.* Acceptable prompts included: *It's OK. Just do your best. I can't help but you can just tell the parts you remember. Is there anything else you'd like to add to your story?* Following the child's story retell, the examiner asked six story questions that were designed to probe child understanding of the characters, setting, problem, attempt, consequence, and lesson learned. Sample questions included: (a) character: *Who was this story about?* (b) setting: *Where was Tara in the beginning of the story?* (c) problem: *Why was Tara scared?* (d) action: *How did she first try to fix her problem?* (e) consequence/unresolved problem: *Why did she talk to her friend?* (f) ending: *How did the story end?* (g) *What will Tara do the next time she can't find her mom?*

Participants' grade level was the main factor that drove selection of the CUBED NLM level administered. We administered the second-grade CUBED NLM Reading probes to Kevin and Kylee during the summer sessions and the third-grade probes as they entered third grade. However, we continued using the first-grade passages with Adam even after he entered second grade after summer. We noted that the episodic construct and narrative macrostructure differ substantially between the first-grade (1 episode) and second-grade probes (2 episodes), whereas such differences are not evident between the second- and third-grade probes. With Michael, we used the kindergarten reading narrative probes but did not collect the CUBED NLM data during the intervention condition to reduce task-induced anxiety. We noted that in the baseline condition reading the entire passage before answering the comprehension questions was a laborious and time-consuming task for Michael. To prevent excessive data collection from depleting the time spent on instruction that actually drove learning, we decided to tolerate this study protocol deviation. We provide only preliminary insights into the treatment effects on the CUBED performance and note that no conclusion for a functional relation should be drawn from the CUBED NLM data.

Software and Instructional Materials

The interventionist used ZOOM®, a secure videoconferencing software, as the platform to deliver instruction via a laptop to each participant who used an iPad. ZOOM® allowed the interventionist and the child to share application windows and to annotate directly on the shared screen so that notations were visible on both devices.

Lesson materials were adapted or drawn from multiple resources including basal reading series such as the Houghton Mifflin Harcourt's Journeys reading series and the Open Court basal reading series (Bereiter, 2004), Flyleaf emergent reader series, popular children's books, educational websites (e.g., Wonderopolis.org, Newsla.com), and Story Champs 2.0 (a narrative-based intervention program). Word cards and grade-level passages were remade into DOCX or PPTX files to be presented via screen sharing. The instructional books included primarily text with limited pictures. Our rationale for minimal pictures was to focus the child's attention on constructing meaning from reading the text (i.e., linguistic input) rather constructing meaning from text as well as visual support (i.e., pictures). However, the inclusion of some pictures was done to maintain a child's interest, similar to what the child might experience in classroom reading materials.

We introduced three graphic organizers (see Appendix B) on the first day of intervention when introducing the Does it Make Sense strategy. The graphic organizers guided participants to reflect on the "*What, When, Why"* of using the strategies, the types of mistakes in reading that can cause misunderstandings, and the fix-up tools (i.e., effective word-solving strategies) that can be applied when encountering unfamiliar written words. These graphic organizers were made available to the participants in the subsequent intervention sessions and they were prompted at the outset of the session to recall the strategy. The graphic organizers were not available in the probe assessments.

Study Conditions

This single-case intervention study comprised three conditions.

Baseline Condition

In the baseline condition, the interventionist met with each child online two to three times a week; baseline sessions involved a business-as-usual adult-child reading interaction (30-40 minutes). Additionally, the child completed a 4-minute probe assessment once or twice weekly, depending on whether two or three weekly sessions, respectively, were scheduled with the individual. Participants also completed the supplemental assessment, the CUBED NLM probe, every two weeks. This schedule maintained a balance between assessment and business-as-usual instruction.

The business-as-usual instruction provided by the interventionist was designed to mirror typical feedback children receive in oral reading practice in class or at home.

When the interventionist listened to a child read, she stopped the child immediately at each point of miscue, supplied the word, and told the child to continue reading. The interventionist provided the instruction as follows: *Today we will learn how to become a better reader by correcting reading mistakes. When you read a word incorrectly, I will tell you the right word.* In these business-as-usual interactions, the child acted as a passive receiver of terminal feedback (i.e., word/phrase supply). The child was not required to re-read the misread word. The interventionist posed a few generic comprehension questions in each session to keep the participant engaged in the reading activities, for instance, *Who was this story about? What do you think about this part*? *What would happen next*? *How does he feel*? *What does it mean*? *How did the story end?* These questions added to the ecological validity of the study as these are the types of questions teachers would ask in children's school reading instruction. The reading passages used in the business-as-usual instruction were those described in the paragraph above. As appropriate, a book or passage was continued across sessions until it was read in full.

Intervention Condition

With each participant, instruction in the intervention condition was delivered in two to three one-hour weekly sessions across 7 to 10 weeks. We minimally collected a 4-minute reading probe at the beginning of the first session of the week. We collected the CUBED NLM probe in the last session every other week. Probe assessments were always collected at the outset of the session to prevent possible practice effects after intervention activities, thereby, best representing a scenario wherein a child reads independently in an educational setting.

The intervention program consisted of two components: (a) comprehension monitoring intervention via *Does It Make Sense* (Liang & Schuele, 2017) and (b) word work (optional).

Applying the comprehension monitoring strategy in oral text reading. The interventionist implemented the *Does It Make Sense* comprehension monitoring intervention (Liang & Schuele, 2017), which involves 5 steps within 35- to 45-minute lessons: (a) introduction/strategy review/vocabulary preview, (b) I DO (explicit modeling), (c) WE DO (guided practice), (d) YOU DO (repeated reading and independent practice), and (e) daily checks/vocabulary review/wrap-up. In the beginning sessions, the interventionist gave the student a reason for learning the skill and specific objectives for each session. She explained WHAT question to ask (*Does it make sense?*), WHEN to ask (each clausal/sentence boundary, marked by an ending punctuation and occasionally by a comma), and WHY to ask the question. She discussed the types of reading mistakes that interfere with text meaning construction. The interventionist also modeled the use of effective word reading strategies, referred to as “fix-up tools,” after recognizing the miscues that disrupted reading comprehension. The fix-up tools included – *read again, sound out,* and *break the word apart into smaller pieces*.

The interventionist introduced the comprehension monitoring strategy through explicitly modeling in the I DO step; as she read to the end of each sentence/clause in a passage, regardless of accuracy, she posed the question *Does it make sense?* If her oral rendering of the text made sense, she verbally acknowledged this (e.g., *Yes, that make sense to me. What I read makes sense. I'll put a check here.*), put a check mark by the ending punctuation, and proceeded reading. If it did not make sense, she paused, acted surprised, and said *Wait! It does not quite make sense to me. Does it make sense to you?* She then scanned the sentence quickly, underlined where the sentences seemed to have not made sense (i.e., the location of word(s) she misread), and applied the fix-up tools to solve unknown or misread word(s) so she could repair her understanding, all while using a think-aloud strategy to guide the child to recognize what she was doing.

The interventionist and the child worked through guided practice in the WE DO step; the interventionist guided the participant to apply the strategy when she and the participant took turns reading. In WE DO the interventionist gradually released control (across days, within sessions) to allow the participant to practice the strategy on his or her own – as the participant demonstrated independence in using the strategy. To encourage the child to develop autonomy in utilizing the fix-up tools, generic feedback was provided sometimes, but only after the participant recognized the miscue by applying the *Does It Make Sense* strategy. Feedback included *What sound does this letter make? These two vowels together make the 'ai' sound.* or *Can you break this word apart into small pieces so you can better sound out the word?*

In the YOU DO step, the student was encouraged to practice the strategy independently while the interventionist provided positive reinforcement and feedback to acknowledge accurate employment of the strategy. The interventionist provided scaffolds only as needed, for instance, when the participant used *Does It Make Sense* to identify unfamiliar written words but struggled to effectively decode the word(s) effectively due to inadequate phoneme-grapheme knowledge. Opportunity for repeated reading was provided if the participant demonstrated laborious reading during the I DO and WE DO steps; otherwise, the participant could choose to read a new portion of the text in the YOU DO step, that is, while practicing *Does It Make Sense* independently.

During the practice of *Does It Make Sense*, the interventionist also asked ad hoc questions intermittently to encourage the participant to track comprehension not just at a sentence level but also at a text level. Ad hoc questions included questions focused on story grammar elements (e.g., *Who was the story about? What was his/her problem? How did he/she feel about the problem?*) and 're-evaluating' questions (e.g., *Now quickly scan the page. Is there anything on this page that does not make sense to you?*).

See *Does It Make Sense Intervention Protocol* (Liang & Schuele, 2017) for additional details on the lesson structure applied in the baseline and intervention conditions.

Word work (optional). The purpose of the word work was to strengthen the participant's use of phoneme-grapheme knowledge in decoding written words so as to use the fix-up tools within *Does It Make Sense* with relative ease. The interventionist determined if the participant needed additional word work instruction based on information collected from standardized measures (e.g., pre-intervention assessment) and error analysis of progress monitoring probes. In the word work activities, the interventionist engaged the participant in 10- to 15-minute word-reading-related activities, wherein the participant was asked to integrate phonemic awareness (i.e., the analysis of sounds in words) with graphemes and/or apply orthographic/morphological knowledge to read or spell words. Appendix C presents sample activities.

Maintenance Condition

In the maintenance condition, across three weeks of no intervention, we collected one 4-minute reading probe and one CUBED NLM probe on different days each week. We also administered the post-intervention reading battery after the probes.

Response Definitions and Measurement System

The dependent variables that were measured repeatedly in this study included: (a) self-correction proportion, measured in the 4-minute reading probe and (b) narrative comprehension, measured by the CUBED NLM probe.

Self-Correction Proportion (Primary; Count Coding)

In the 4-minute reading probe, the author transcribed online child miscues and self-corrections on a hard copy of the reading passage. All probes were video-recorded so that the author could verify her transcribed reading miscues and self-corrections and make corrections (if needed) while replaying the files, at a later time.

The primary dependent variable – self-correction proportion – was operationalized as the number of self-correction behaviors divided by the number of reading miscues produced in a 4-minute reading probe. Reading miscues were errors in oral reading (e.g., Laing, 2002, Woodcock, 2011). A miscue creates an opportunity for the reader to self-correct. Table 4 provides a listing of the miscue types that we recorded, as well as a definition and example that characterizes each miscue type.

Table 4

Operationalized Definition and Examples of Reading Miscues

Miscue Type	Definition	Example
Addition of Word	Reader inserts an extra word into a sentence.	= T The sheep… R The sleeping sheep…
Omission of Word or Word Part(s)	Reader omits a word or part of a printed word.	= T At times… R Times… *OR* R At time…
Mispronunciation of Word	Reader mispronounces a printed word.	= T …combed… R …com-bed…
Word Substitution	Reader substitutes a word for a printed word.	= T He cried… R He circle…
Non-Word Production	Reader substitutes a nonword or makes unintelligible sound sequence for a printed word.	= T…protect them... R …poick them…
Word Reversal	Reader reverses the order of 2 printed words.	= T Something was tapping… R Was something tapping…
Pause 4 sec.	Reader ceases reading for more than 3 seconds.	= T …from wolves. R …from (long pause).

Note. This table was adapted from information in the WRMT-III manual and Oral Reading Fluency protocol form (Woodcock, 2011). T = text, R = child oral reading.

Self-correction behavior was defined as a spontaneous child behavior (i.e., not prompted or elicited by the interventionist) that was exhibited after a reading miscue and that involved an attempt to revise what had been read aloud. Such behaviors reflect child's attempts to read the text with better accuracy that we hypothesize are a result of the child's attempts to derive meaning from the read text. We counted successful corrections as well as quality attempts of self-correction (though not successful) in our tally of self-correction behaviors (see Table 5). In a successful self-correction, the participant accurately reads the word that was previously misread (i.e., corrects miscues). In a quality attempt of self-correction, the participant engages himself or herself in re-reading the text and attempts to self-correct, sometimes with application of fix-up tools. However, although the child's attempted self-correction does not yield the correct word, the rendered word shares either (a) part of the pronunciation ($\geq$ 75%), (b) the stem (e.g., *dancing*, the student says *dank* but later corrects himself by saying *dances*), or (c) the meaning (e.g., *no*, the student says *on* but later corrects herself by saying *not*), or (d) generally conforms to the graphophonemic rules (e.g., *confident*, the student says *contain* but later corrects himself by saying /kɑnfaɪdənt/ *con-fie-dent*).

Narrative Comprehension (Secondary; Rating Scale)

In the CUBED NLM probe assessment, the child retells a story that he or she has just read and responds to the six (first-grade level of CUBED NLM) or seven (second-grade level of CUBED NLM) comprehension questions. Each question probes a specific story grammar element. We video recorded the CUBED NLM to allow for later scoring. The child's response to each of the story comprehension questions was scored using a three-point rating scale (see the Appendix D for scoring guidelines excerpted from the CUBED manual). We tallied the individual question scores to form a total story question composite score (first-grade maximum = 12; second-grade maximum = 14).

Table 5

Operationalized Definition and Examples of Self-Correction Behaviors

Self-Correction Type	Definition	Examples
Self-correction (successful attempt) **Code: SC**	Following a miscue, child reads the word correctly or corrects himself/herself a few words after. Child may make one or several attempts, including saying an incomplete word, a different word, or a nonsense word before successfully rendering the word(s).	= T They went horse riding. C They went force riding…no horse riding. = T His grandpa lives on the farm. C His grandpa lives on the arm…f-arm. = T It was delicious. C It was delicions delicious. *Non-example:* If the child shows a consistent pattern of sounding out each letter sound prior to reading the word in full, particularly in young slow decoder with limited automatic word recognition. Do not code as self-correction. = T They are having a party. C They are having a p-æ(ah)-r-t-ee party. (p-æ(ah)-r-t-ee is a sounding out attempt rather than a miscue, so do not code *party* as SC)
Quality self-correction attempt (unsuccessful) **Code: UA** Word(s) that were produced in child's unsuccessful attempts share:	Following a miscue, child attempts to correct him/herself by using more effective word-solving strategies (e.g., sounding out) rather than primitive strategies such as guessing; however, such attempts does not lead to successful word identification.	= T At times… R Times… *OR* R At time…
(a) Part of the pronunciation (over 75%)		= We were inside a space shuttle. C We were inside a space ship, shut, shut-tle-ly.
(b) Base word (stem)		= T They danced all night. C They dank all night…dancing.
(c) Meaning		= T…no… C …on…not
(d) Word with pronunciation that generally conforms to the graphophonemic rules		= T Tray and Mary were confident. C Tray and Mary were contain…con-f-igh-dent…confiedent. *Non-example: eyes*, the child says *less* and then changes it to *else*.

Interobserver Agreement (IOA) and Fidelity

Prior to initiating the project, the first author met with the reliability coder (a student in the master's program in speech-language pathology) in an hour-long training to review the training materials, including the coding manual, the lesson manual, and the rubrics for scoring story comprehension questions. In the next meeting, the author and the reliability coder went through a compilation of edited audio-files that had been recorded from previous projects. These audio-recordings provided various examples of reading miscues and self-correction behaviors. Clarification in coding was addressed by the author if questions were raised by the reliability coder. The reliability coder then scored four training 4-minute reading probes and four CUBED probes that had been recorded from pilot or prior work and compared her scoring with a coding key that had been generated by the author. The reliability coder only began scoring study data after she had completed the four training probes and attained 90% or higher interobserver agreement on each probe.

To record data collection, all sessions were video-recorded using ZOOM®'s recording feature. The author used the video-recordings to verify scoring taken online and made corrections as needed. The reliability coder used the video-recordings to score 30% of randomly selected probes.[2] The reliability coder was blind to the experimental condition of each randomly selected probe; files used for reliability were edited to remove dates and included only the probe (i.e., not other research activities). The coding of the primary investigator and the reliability coder were compared using a point-by-point procedure to calculate interobserver agreement.

[2] Probe reliability checking was performed throughout the study period rather than after all data had been collected. For every three probes collected, one probe was randomly selected by the reliability coder.

For the procedural fidelity, the reliability coder collected agreement data for approximately 30% (randomly selected) of the intervention sessions for each participant. Using a researcher-created checklist of intervention implementation steps, the observer noted those steps that indeed occurred in each of the intervention sessions (see Appendix E). The number of steps completed were divided by the total number of steps on the checklist and multiplied by 100 to note procedural fidelity for each session.

Data Analysis

Visual analysis was the primary means to ascertain the presence of a functional relation (see Table 6). Specifically, we attended to (a) number of data points plotted within a condition, (b) level stability and changes in level within and between conditions, (c) trend direction, trend stability, and changes in trend within and between conditions, and (d) percentage of data points in one condition that fall within the range of data plotted in an adjacent condition (e.g., percentage of non-overlapping data; PND). PND is calculated by (a) evaluating the range of baseline values, (b) counting the number of intervention data points outside of the range of baseline values, and (c) dividing these data points by the total number of data points in intervention and multiplying by 100 (Gast, 2009).

Table 6

Data Properties Analyzed via Visual Analysis

Visual analysis features assessed within conditions	
Level	The magnitude of data on the dependent measure as indicated by the ordinate scale value in a data series. Two basic aspects of level that are important: level stability and level change within the same condition
Trend	The steepness or direction of the data path across time (e.g., increasing, decreasing, combination, or remaining the same)
Variability/stability	Fluctuations of data point value observed in a data series
Visual analysis features assessed between conditions	
Immediacy	The change in level or trend between the last three data points in one condition and the first three data points of the next. The more immediate the effect, the more convincing the inference that change in the outcome measure was due to manipulation of the independent variable. However, predicted delayed effects or gradual effects of the intervention may sometimes be built into the design of the experiment
Overlap	The proportion of data from one condition that overlaps with data from the previous condition. The smaller the proportion of overlapping data points, the more compelling demonstration of an effect
Consistency	The extent to which data patterns are the same within like conditions (e.g., in baseline conditions for all participants in a multiple probe design across participants)

In addition to visual analysis, we quantified the impact of intervention by calculating the Tau-*U*/Tau$_{\text{nonoverlap}}$ and Baseline Corrected Tau across conditions (Parker et al., 2011; Tarlow, 2017). Tau-*U* is a family of rank correlation indices, among which the most commonly used ones are simple non-overlap (Tau-$U_{\text{baseline vs. treatment}}$) and simple non-overlap controlling for positive baseline trend (Tau-$U_{\text{baseline vs. treatment – baseline trend}}$). Parker et al. (2011) present Tau-*U* as a desirable effect size for single-case research. The online Tau-*U* calculator developed by Vannes et al. (2011) permits researchers to analyze data for several phase contrasts and properly aggregate these effect sizes for an overall or omnibus effect size (www.singlecaseresearch.org/calculators/tau-u). We also included Baseline Corrected Tau analysis, an improved effect size statistic developed by Tarlow (2017) to address potential limitations associated with Tau-*U*. These limitations include (a) inflated values, (b) results may not be bounded between -1 and 1, (c) Tau-*U* baseline control cannot be visualized, and (d) a relative weak method for trend control leads to inflated Type I error. An online calculator for Baseline Corrected Tau (Tarlow, 2016) is available at www.ktarlow.com/stats/tau.

CHAPTER III

RESULTS

The present study examined three critical components that defined feasibility of a telepractice intervention including: (a) treatment effects – the functional relation between the comprehension monitoring intervention and proportion of child self-corrections in oral text reading, (b) scoring reliability, and (c) fidelity of intervention implementation. Participants received an average of 16.5 intervention sessions (*SD* = 1.91, range = 15 – 19) delivered across an average of 8 weeks (*SD* = 1.22, range = 7 – 10). Each intervention session lasted between 45 to 60 minutes.

Results are presented in four parts. First, the percentage of agreement between independent raters' coding of reading miscues and self-correction behaviors across observed 4-minute reading probes are reported as means and ranges. Second, the calculated means and ranges of implementation fidelity of treatment are presented for each participant. The fidelity of probe administration was reported for all participants in all study conditions. Third, participant baseline, intervention, and maintenance assessment data are described in text and graphed with days on the abscissa and self-correction proportion on the ordinate. Lastly, two sets of exploratory data are presented including a pre-post intervention comparison of participants' performance on norm-referenced reading assessments as well as CUBED's story comprehension that was repeatedly assessed across conditions.

Interobserver Agreement

Interobserver agreement (IOA) data for participants' coded reading behaviors gathered from the 4-minute reading probes were computed using point-by-point agreement method, dividing total agreements on each occurrence of reading miscues and self-correction behaviors by total agreements and disagreements (Gast, 2010). IOA was calculated for a minimum of 30% of the probes collected in the baseline, intervention, and maintenance conditions (range = 30.43% – 31.25% within condition). The probes to be checked for agreement were selected at random by the reliability coder across study conditions and scored independently between the examiner and the reliability coder. Table 7 displays the IOA data within each condition across participants. Average IOA across all four participants was 87.03% (range = 84.59% – 88.56%). Across conditions, the mean IOA on scoring Adam's self-correction proportion was 84.59% (range = 66.67% – 92.86%). The mean IOA on scoring Kevin's self-correction proportion across observed sessions was 88.56% (range = 84.62% – 92.31%). The mean IOA on scoring Michael's self-correction proportion across sessions was 88.57% (range = 84.85% – 90.91%). Kylee's mean IOA for self-correction proportion across observed sessions was 86.40% (range = 76.47% – 95%).

Procedural Fidelity

Fidelity for the intervention program included: (a) fidelity of intervention implementation and (b) procedural fidelity of probe administration. We collected the procedural fidelity data using a researcher-created checklist (see Appendix E). The mean procedural fidelity of probe administration across participants was 100% (*SD* = 0).

The mean fidelity of intervention implementation across participants was 96.13% (range = 93.88% – 98.33%). Fidelity data on intervention implementation were collected on a minimum of 30% of the intervention sessions (range = 29.41% – 33.33%). The mean fidelity of intervention implementation for Adam was 97.62% (range = 90.48% – 100%). The mean fidelity of intervention implementation for Kevin was 98.33% (range = 95% – 100%). The mean fidelity of intervention implementation for Michael was 94.69% (range = 90% – 100%). The mean fidelity of intervention implementation for Kylee was 93.88% (range = 89.47% – 95.24%).

Table 7

Interobserver Agreement Data by Conditions and Participants

	Participants			
Condition	Kylee	Michael	Adam	Kevin
Baseline				
Mean (SD)	87.07 (6.97)	88.26 (2.37)	86.51 (5.99)	87.46 (2.58)
Range	82.14 – 92	86.36 – 90.91	80.95 – 92.86	84.62 – 89.66
Intervention				
Mean (SD)	85.74 (13.10)	90.91 (0)	80.30 (12.40)	90.38 (2.72)
Range	76.47 – 95	90.91	66.67 – 90.91	88.46 – 92.31
Maintenance	86.36	84.85	91.67	88.24

Treatment Effects: Reading Self Correction Proportion

Participants in the present study were assessed repeatedly with the 4-minute reading probes in the baseline, intervention, and maintenance conditions to answer the research question: Does the comprehension monitoring intervention delivered via telepractice result in increased use of self-correction behaviors in child unsupported oral text reading?

Self-Correction Proportion

The primary dependent variable, self-correction proportion, was operationalized as the number of self-correction behaviors divided by the number of reading miscues observed in a 4-minute reading probe. Figure 1 presents all four participants' data on self-correction proportion across observed sessions in all study conditions. Visual analysis was our primary means to determine the presence of a functional relation between our intervention and the dependent variable. Estimates of level, trend, and variability were assessed within and across conditions, with special attention to the degree of overlap (e.g., percentage of non-overlapping data, PND), immediacy of effect, and similarity of data patterns in similar conditions. Figure 2 displays change in the number of reading miscues and the number of self-corrections observed in the 4-minute probe assessment.

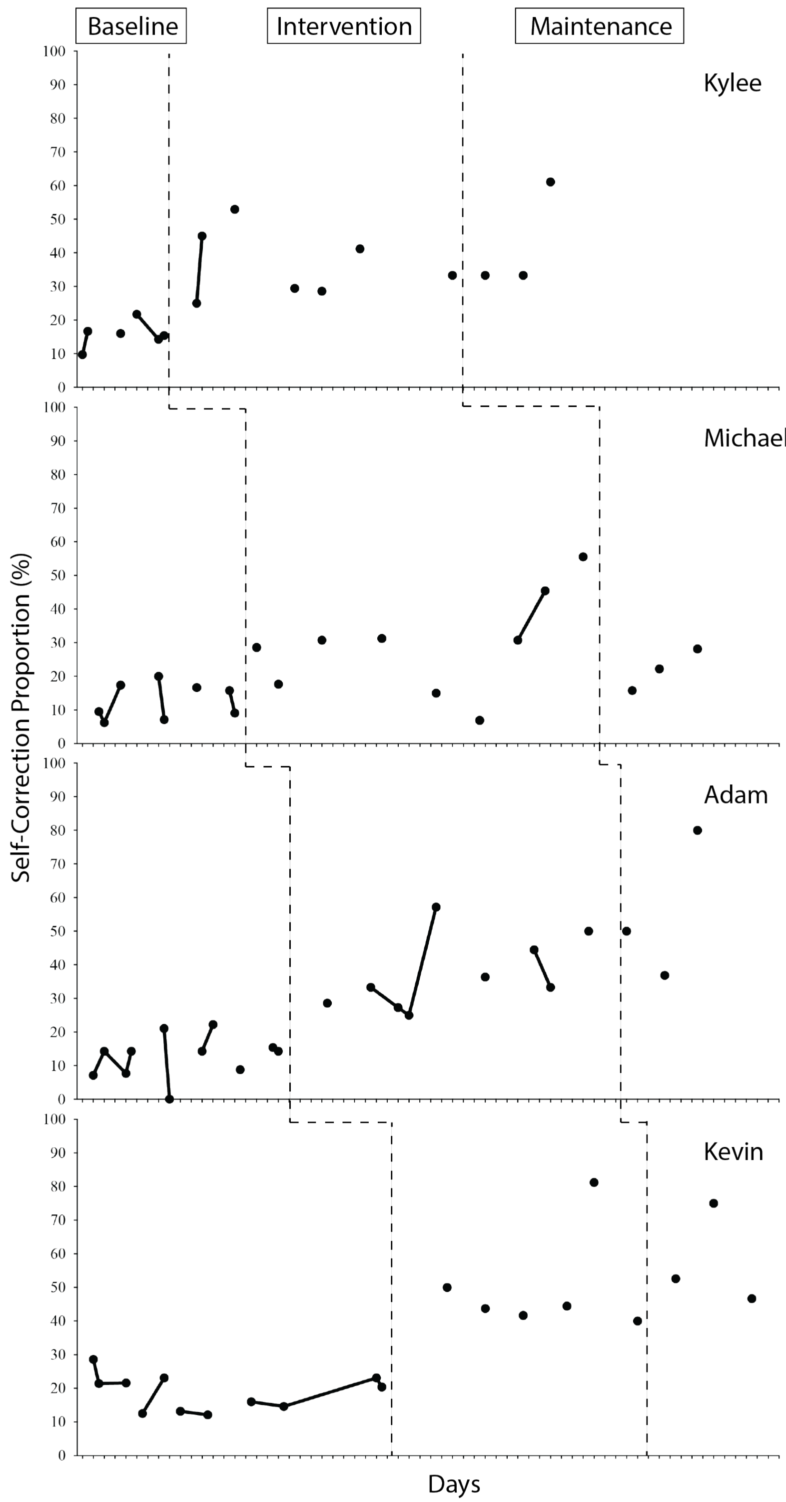


Figure 1. Proportion self-correction (primary dependent variable) by each participant derived from the 4-minute reading probes during the study period. The major tick interval on the ordinate is set to 2.

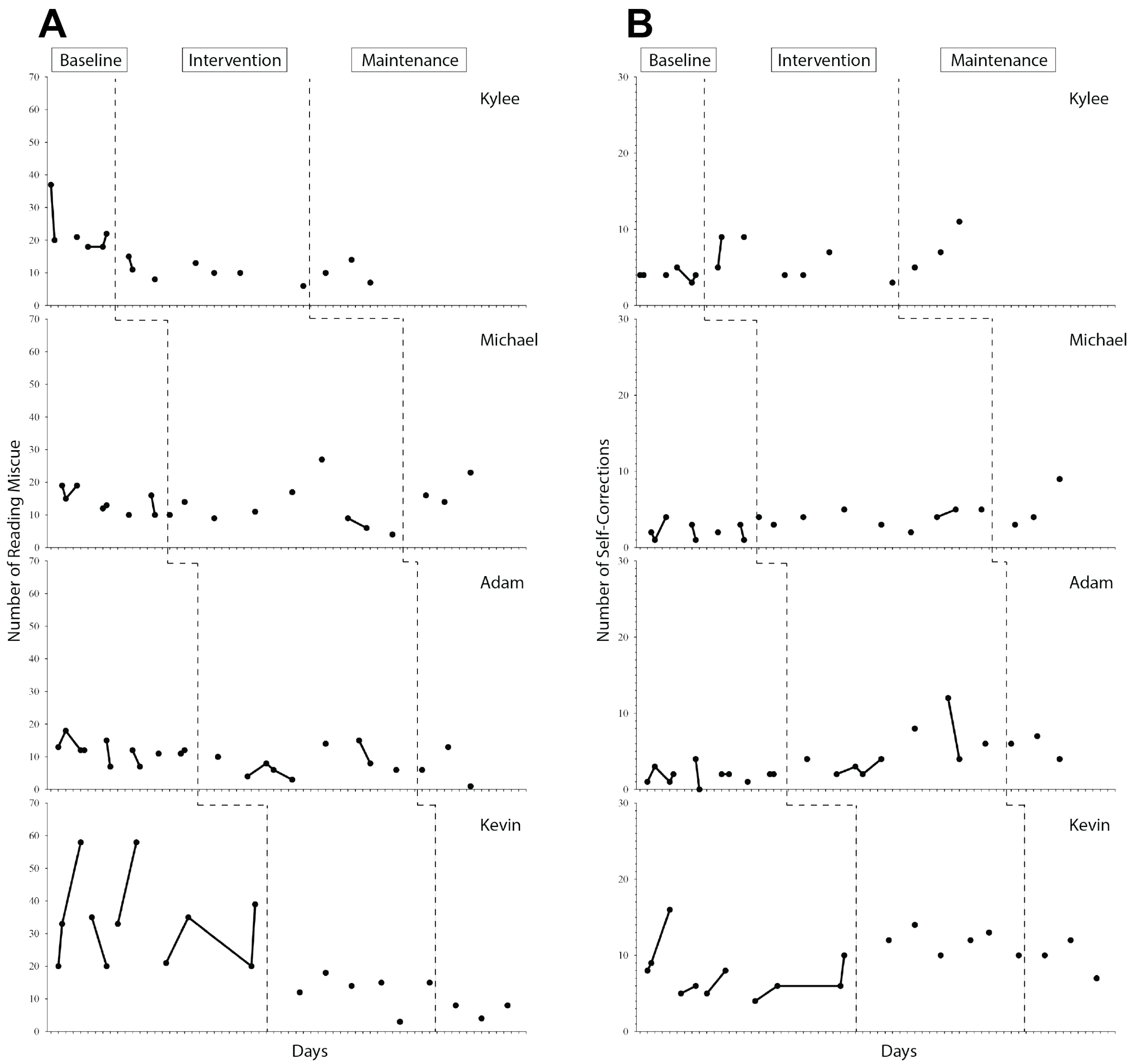


Figure 2. Number of reading miscues (A) and number of self-corrections (B) by each participant derived from the 4-minute reading probes during the study period. The major tick interval on the ordinate is set to 2.

Kylee. Kylee was selected randomly to enter the intervention condition first after all participants demonstrated level stability or a countra-therapeutic trend (i.e., a deteriorating trend). The first tier of Figure 1 depicts results of Kylee's self-correction proportion across all observed sessions. Kylee participated in a total of 7 baseline sessions across 2 weeks with 6 probes administered and 15 intervention sessions across 7 weeks with 7 probes administered. She demonstrated mean self-correction of 36.49% (*SD* = 10.16%; Range = 25% – 52.94%) in the intervention condition and 42.59% (*SD* = 16.04%; Range = 33.33% – 61.11%) in the maintenance condition as compared with her mean baseline performance of 15.64% (*SD* = 3.87%; Range = 9.76% – 16.67%).

With regards to immediacy of effect, Kylee showed evident change in her reading self-correction behaviors after four intervention sessions with notable slowed reading and more frequent pauses to attempt self-corrections. These changes resulted in significantly decreased number of reading miscues and increased use of self-correction behaviors. We noted that throughout the intervention condition Kylee continued to struggle with phonologically and morphologically complex words such as *encourage* and *biography*, which required her to develop more advanced decoding skills and/or morphological knowledge to support successful word solving. However, she evidenced more self-monitoring behaviors including sounding-out attempts and/or appeals for help in the intervention condition relative to her baseline data.

Kylee's vulnerability in the grammatical aspects of spoken language also permeated through written language. She persistently produced a substantial number of omission and/or commission errors when reading words that were inflected to mark, for example, plural -*s* and past-tense -*ed* (e.g., omission: *bat* for *bats*; commission: *babies* for *baby*). She evidenced irregular past-tense confusion such as *come* for *came* and *tell* for *told*. These types of errors, though considered as low meaning-change miscues, resulted in a less drastic change and some variability in Kylee's self-correction proportion when she exited the intervention. Despite the variability, her PND across sessions was 100%.

Michael. As Kylee began to show evident change in her self-monitoring behaviors in oral text reading while other participants maintained stable baselines, Michael was selected randomly to enter the intervention condition. Despite some degree of variability observed in his baseline data, there was no gradual accelerating trend in the direction of improvement. The ordinate data points were no greater than 20%. As a result, we decided to introduce the intervention.

Michael participated in 11 baseline sessions across 4 weeks with 8 probes administered and 19 intervention sessions across 10 weeks with 9 probes administered. He demonstrated variable performance of his self-correction proportion throughout the intervention course. This variable performance was characterized by a moderate level change in the first half of the intervention condition and a combination trend (decelerating-accelerating) in the second half. He scored mean self-correction of 27.06% (*SD* = 14.64%; Range = 6.9% – 55.56%) in the intervention condition as compared with his mean baseline performance of 12.73% (*SD* = 5.29%; Range = 6.25% – 17.39%). He entered the maintenance condition after completing 10 weeks of intervention. His mean self-correction observed in the maintenance condition was

22.05% (*SD* = 6.17%; Range = 15.79% – 28.13%), overlapping substantially with where he scored early in the intervention phase. Michael's PND across sessions was 62.5%. Because of the variability noted in the intervention condition and substantial overlap between the adjacent conditions, a demonstration of treatment effects was weakened.

It is noteworthy that, among all the participants, Michael presented the most restricted knowledge of grapheme-phoneme correspondences and words that he could readily recognize (i.e., automatic word recognition) prior to intervention. His reading experiences during the probe assessment (unsupported reading) required tremendous mental effort on his part and often were imbued with an accumulating sense of frustration, which quickly resulted in mental fatigue.

Adam. Adam was randomly chosen to enter the intervention condition two weeks after we began the intervention with Michael. Based on our prior experiences implementing *Does It Make Sense*, most participants would at least demonstrate some preliminary changes in their use of self-correction behaviors two weeks into intervention. Nevertheless, we speculated that Michael might show somewhat delayed or less drastic response to intervention based on his educational history per teacher report (e.g., inadequate responsiveness to intensive and targeted intervention in first grade, grade retention, etc.). Additionally, information drawn from standardized measures, error analysis of progress monitoring probes, and observations collected during the baseline condition suggested that Michael had rather restricted automatized phoneme-grapheme knowledge as well as less developed word recognition relative to other participants. Thus, we began the intervention with Adam, as he had maintained stable baseline, once minimal evidence of change was observed in Michael's self-correction percentage. Similar to Michael, some variability was detected in Adam's baseline data without any indication of an improving trend; consequently, we decided to proceed.

Adam participated in 13 baseline sessions across 6 weeks with 11 probes administered and 17 intervention sessions across 8 weeks with 9 probes administered. With respect to immediacy of effect, he displayed a relatively slow response yet a moderate level-change in his self-correction percentage during the first two weeks of intervention, primarily driven by a decreased number of reading miscues. In the second half of the intervention, Adam began to show more marked change in self-correction proportion due likely to an increased use of self-correction behaviors and/or decreased instances of reading miscues. Hence, the visual analysis supports level change within the intervention condition.

Overall, Adam scored a mean self-correction of 37.27% (*SD* = 10.99%; Range = 25% – 57.14%) in the intervention condition and 55.61% (*SD* = 22.12%; Range = 36.84% – 80%) in the maintenance condition relative to his mean baseline performance of 12.63% (*SD* = 6.43; Range = 0% – 22.22%). His PND across sessions was 100%.

Kevin. Kevin entered the intervention condition as he demonstrated level stability in the baseline condition, concurrent with Adam showing at least three data points above his baseline range. Kevin participated in 14 baseline sessions across 8 weeks with 11 probes administered and 15 intervention sessions across 7 weeks with 6 probes administered. Results of the visual analysis indicate abrupt level change between the baseline and intervention conditions, indicating that the intervention had an immediate effect on his self-monitoring behaviors. Kevin scored a mean self-correction of 50.19% (*SD* = 15.59%; Range = 40% – 81.25%) in the intervention condition and 58.1% (*SD* = 14.94; Range = 46.67% – 75%) in the maintenance condition relative to his mean baseline performance of 19.2% (*SD* = 5.5; Range = 12.12% – 28.57%). His PND across sessions was 100%.

In sum, Kevin evidenced a clear demonstration of a treatment effect with an immediate response to intervention shortly after he entered intervention. The marked increase in his self-correction proportion was driven by reduced number of reading miscues as well as increased use of self-correction across observed sessions.

Tau-U and Baseline Corrected Tau

Both Tau-*U* and Baseline Corrected Tau use a two-step process to estimate an effect size across two adjacent conditions (baseline vs. treatment). In Step 1, baseline trend was estimated and (if necessary) corrected. In Step 2, an effect size was calculated based on either the original or corrected data. Results derived from both statistical methods appear comparable (see Table 8) and also in agreement with visual analysis. Data showed 46% to 100% overall improvement between the baseline and treatment conditions. This amount of improvement is significant at *p* values that ranged from .04 to <.001. The averaged effect size .906.

Table 8

Tau-U and Baseline Corrected Tau Analyses

	Tau-*U*					Baseline Corrected Tau			
	Kylee	Michael	Adam	Kevin	All	Kylee	Michael	Adam	Kevin
Step 1: Determine if statistically significant baseline trend is present and if baseline correction is needed									
Tau	.067	0	-.273	-.178	-	.0670	0	.289	-.180
p value	.850	1.00	.240	.470	-	1.000	1.100	.260	.53
Baseline correction (Yes/No)	No	No	No	No	-	No	No	No	No
Step 2: Compute effect size									
Tau value	1	.625	1	1	.906	.734	.458	.736	.710
p value	<.01	<.05	<.001	<.01	<.001	<.01	<.05	<.001	<.01

Treatment Effects: Changes on Norm-Referenced Assessments

We re-administered part of the pre-intervention battery to evaluate if any of the incremental gains observed in the progress monitoring assessment were captured by the norm-referenced assessment that pertained to word-level reading and reading comprehension. Figure 3 depicts participants' pre-intervention and post-intervention scores on the WRMT-III Word Identification (WI), Word Attack (WA), and Passage Comprehension (PC), the WIAT-III Reading Comprehension (RC), and the TOWRE-2 Sight Word Efficiency (SWE) and Phonemic Decoding Efficiency (PDE).

On the WRMT-III WI, an untimed word recognition measure, most participants demonstrated some gains that ranged from 5 to 8 standard score points, except for Michael who showed a slight decline from 83 to 79. Greater score gains were observed in the WRMT-III WA, an untimed measure that assesses nonword decoding. This test likely mirrors situations wherein a child must utilize his grapheme-phoneme knowledge to decode unfamiliar written words encountered in text. Three of the participants showed gains in standard score that ranged from 5 to 20 standard score points. In contrast, although Adam demonstrated a pre-post gain of 5 raw score points, there was a slight decline in standard score from 98 to 95.

For reading comprehension, most participants demonstrated gains in standard score that ranged from 5 to 8 points on the WRMT-III PC, an untimed reading comprehension measure that uses a cloze procedure, except for Michael whose received the same standard score. We also administered the WIAT-III RC, an untimed measure that assesses reading comprehension with open-ended questions. Participants' data patterns of the pre-post comparison of the WIAT-III RC was variable. Adam maintained the same standard score whereas Michael showed a moderate decline of 5 points. Kevin and Kylee demonstrated a pre-post intervention gain in standard score that ranged from 4 to 10 points.

The least transferred intervention effects were found in timed word-reading and decoding measures. On the TOWRE-2 SWE, all the participants experienced a decline in standard score that ranged from 3 to 8 points. On the TOWRE-2 PDE, three participants maintained the same level of performance but Michael evidenced a 9-point score gain in the pre-post comparison.

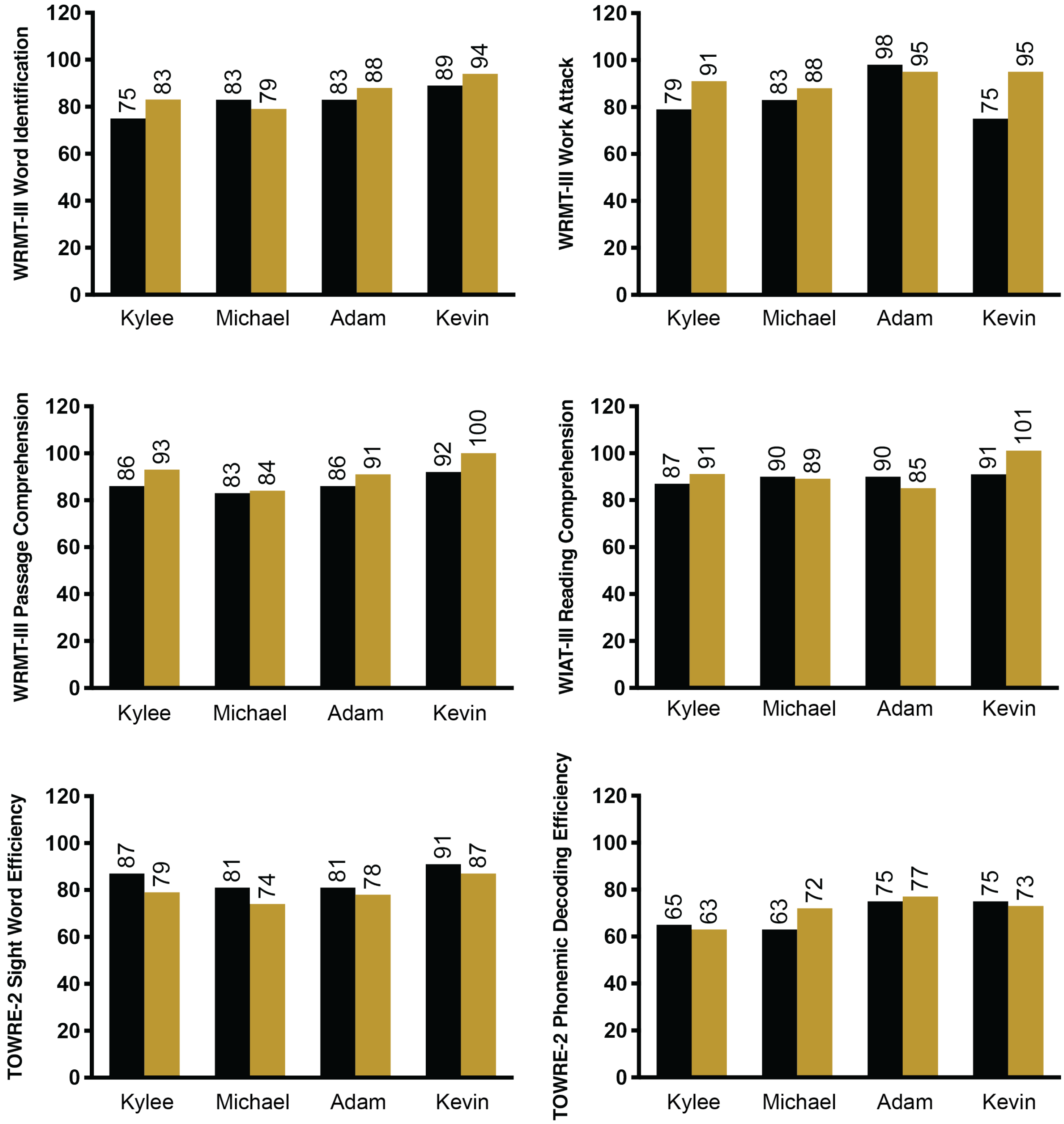


Figure 3. Pre- and post-intervention standard scores by reading measure types.

Treatment Effects: Story Comprehension as an Exploratory Outcome

We anticipated the *Does It Make Sense* comprehension monitoring program would have a direct impact on child reading accuracy as evidenced by increased self-correction proportion (i.e., primary dependent variable). In addition to reading accuracy, we were also interested in whether or not this comprehension monitoring program would exert any impact on child reading comprehension, a distal, far-transfer outcome that requires multiple subcomponent skills in the domains of language and literacy to be orchestrated harmoniously. We assessed child reading comprehension with the CUBED NLM Reading progress-monitoring probes that were administered intermittently. Note that we assessed Michael's written narrative comprehension with the kindergarten probes because the first-grade passages were considered to be at his frustration level. Nevertheless, even with these shorter and easier passages, we found the presented texts continued to provoke limiting emotions over time and could potentially interfere with instruction (e.g., taking substantial time away from delivering instruction). Thus, we only collected Michael's data on CUBED NLM during the baseline and maintenance conditions. For the rest of the participants, the CUBED NLM probes were administered intermittently (typically every two weeks) in all study conditions. Figure 4 displays all four participants' data on CUBED story comprehension composite scores with days on the abscissa and CUBED percentage scores on the ordinate. Note that we divided each composite score by the maximum score and then multiplied by 100 to form the percentage score.

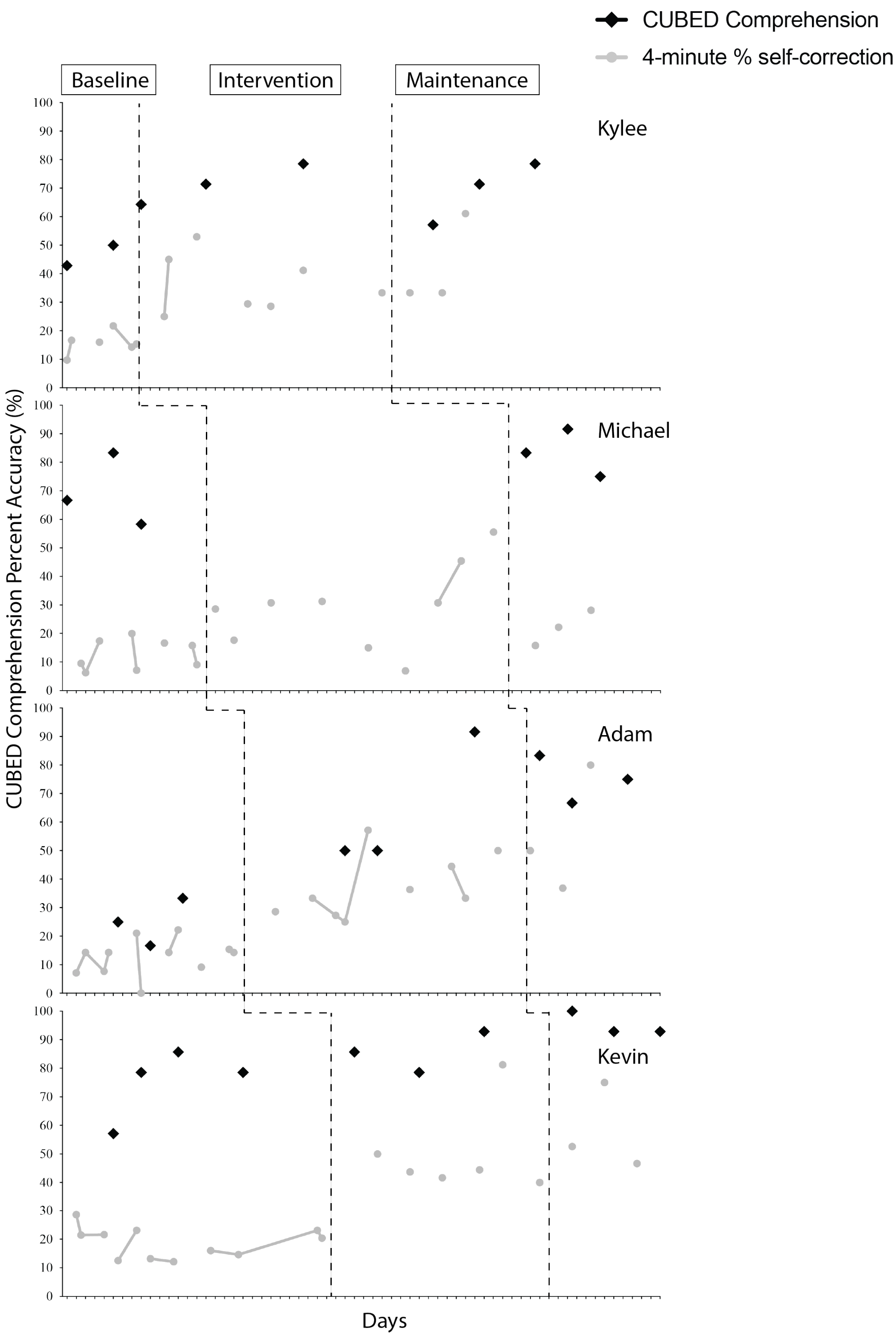


Figure 4. CUBED written narrative comprehension percent accuracy derived from dividing the total CUBED score by the maximum score and multiplying by 100. These scores are overlaid on the self-correction scores for comparison.

Kylee

Kylee scored 6 and 7 out of 14 on her baseline narrative probes. Her baseline responses indicated a lack of understanding of the story grammar and episodic construct. In the intervention condition, Kylee showed level change in her narrative comprehension scores, ranging from 9 to 11 out of 14. She evidenced a better grasp of the basic story grammar. Nevertheless, her expressive language deficits, such as difficulties in finding words and putting words into sensible and grammatical sentences sometimes resulted in responses that were somewhat related to the story but did not meet the full-credit description. In addition, she continued to struggle with comprehending the story as an episodically complex narrative. Kylee's narrative comprehension scores for the maintenance probes ranged from 8 to 11, higher than her baseline scores.

Michael

Michael's baseline narrative comprehension scores ranged from 7 to 10 out of 12. He scored between 9 and 11 out of 12 on his maintenance probes. His baseline and maintenance performances overlapped substantially with limited room left for demonstrating noticeable improvement. For Michael, the CUBED narrative comprehension scores should be interpreted with particular caution because the test instruction differed substantially from that of the 4-minute reading task. When administering the CUBED NLM probes, the child received help in reading unknown words if he or she requested or if the child failed to read a word over 3 seconds. Thus, Michael's comprehension of the CUBED written narratives appeared to be less affected by his word reading deficits in contrast to the 4-minute probe assessment, a measurement context wherein he received no support when reading.

Adam

Adam's baseline story comprehension scores ranged from 2 to 4 out of 12. During baseline, he responded minimally to most of the story questions with answers such as *I don't know*, *I forgot*, or no response (NR) due likely to his expressive language deficits and effortful word reading. These linguistic weaknesses depleted a substantial amount of limited cognitive resources available for constructing meaning from text. He struggled with understanding basic story grammar elements except for naming character(s) of the story. In the intervention condition, Adam showed remarkable progress in his responses to story comprehension questions with scores ranging from 6 to 11. His NR answers reduced substantially with him nearly always attempting each question even if difficulties in word finding and formulating sentences continued to be a hindrance that prevented him from elaborating further. In the maintenance condition, Adam appeared to maintain the gains he made during intervention with his CUBED scores ranging from 8 to 10.

Kevin

For CUBED baseline probes, Kevin scored between 8 and 12 out of 14. He evidenced a good grasp of basic story grammar but often did not recognize that the presented story entailed an episodically complex structure (i.e., two episodes). In the intervention condition, he moved more toward the top of the scale with his CUBED story comprehension scores ranging from 11 to 13 out of 14. In the maintenance condition, he scored between 13 and 14 out of 14. Given the limited room for further improvement, Kevin gained a better insight albeit numerically limited into the episodic construct throughout intervention. His maintenance responses suggested that he retrieved information on characters' first attempt in the first episode, whereas he consistently left out this information in his baseline responses.

CHAPTER IV

DISCUSSION

The present study examined the feasibility of delivering language-literacy intervention via telepractice as an innovative approach to enhance access to intensive services. Although the evidence on telepractice intervention has been overwhelmingly positive, confidence in the evidence is reduced due to multiple study limitations. These limitations include flawed research design, inadequate measurement systems to detect incremental gains due to intervention, and not documenting information on implementation fidelity (Rudolph & Rudolph, 2015; Weidner & Lowman, 2019). We addressed these limitations by conducting a single-case research study, which allowed for rigorous documentation of treatment effects (i.e., a functional relation), interobserver agreement (IOA), and implementation fidelity. When these three components are considered together, we are able to draw firm conclusions about feasibility. A discussion of the present study is organized as follows: (a) the feasibility of delivering *Does It Make Sense* via telepractice, (b) revisiting telepractice: benefits and cautions to be exercised, (c) study limitations, and (d) implications and future directions.

Evidence for Feasibility: Implementing *Does It Make Sense* via Telepractice

Results of this study corroborate that telepractice is a viable means for delivering effective language-literacy intervention with replicated treatment effects and adequate procedural fidelity and probe assessment reliability. We demonstrated a causal relation between the independent variable, the *Does It Make Sense* instruction, and the dependent variable, child reading self-correction proportion. Three out of the four participants demonstrated a level change in self-correction proportion concurrent with the introduction of the intervention and/or a level change within the intervention condition. These three participants maintained about the same level of performance in a three-week follow-up phase without intervention (i.e., maintenance phase). Additionally, the reported IOA and fidelity of treatment implementation were deemed to meet single-case design standards (i.e., $\geq$ 80%). Participants' mean IOAs collected from 4-minute independent oral text readings across all conditions ranged from 84.59% to 88.56%. The mean fidelity of intervention implementation averaged across all intervention sessions for each participant ranged from 93.88% to 98.33% using a researcher-created checklist.

Scoring Reliability

To the best of our knowledge, this study is the first to examine reliability of scoring progress monitoring probe assessment contextualized within oral text reading via telepractice. Prior research addressed scoring reliability in telepractice primarily within the context of norm-referenced speech and language assessments (Ciccia et al., 2011; Eriks-Brophy et al., 2008; Sutherland et al., 2017; Waite et al., 2006; Waite et al., 2010; Waite et al., 2012). Across those studies, scoring reliability between telepractice and face-to-face therapists appeared to be higher for language assessments than for speech assessments, with a range of percent agreement between 98% – 100% and kappas between .88 – .99. Percent agreement in scoring speech assessments was acceptable with a range between 80% – 92%; yet, several limitations have been noted (Taylor et al., 2014). These limitations include difficulties in voicing judgement and reduced agreement for sounds without visible articulation (e.g., /k/ and /g/), high-frequency sounds (e.g., /s/), and consonant clusters.

Because identifying self-corrections and quality self-correction attempts requires phonetic transcription, we expected that limitations noted in speech sound judgement via telepractice would somewhat but not greatly affect IOA. The reported ranges of IOA across participants confirmed this expectation. In the present study, the mean IOA on scoring self-correction proportion across all four participants and conditions was 87.03%. Across conditions, the mean IOAs for each participant were: Adam (*M* = 84.59%; Range = 66.67% – 92.86%), Kevin (*M* = 88.56%; Range = 84.62% – 92.31%), Michael (*M* = 88.57%; Range = 84.85% – 90.91%), and Kylee (*M* = 86.40%; Range = 76.47% – 95%). Of the 24 probes examined for IOA, 22 probes achieved point-by-point agreement above 80%, a professionally-suggested criterion level for single-case studies (Yoder et al., 2018). We noted that two individual probes collected from Adam and Kylee respectively presented relatively low IOA (66.67% and 76.47%). A close inspection of the discrepancies revealed that low IOA was due primarily to the low frequency of reading miscues and self-corrections observed in these two particular probes. When the frequency rates are low, point-by-point agreement of 80% is quite difficult to achieve (Yoder et al., 2018). In sum, the IOA data substantiate that child oral text readings observed in telepractice can be scored reliably between a telepractice interventionist and a reliability coder who listens to the recorded files asynchronously.

Procedural Fidelity

The present study improves on the extant literature by providing rigorous documentation of fidelity of probe administration and intervention implementation via telepractice. The evidence for adequate fidelity of implementing *Does It Make Sense* via telepractice is considered to be very good. Across participants and conditions, the mean fidelity of probe administration was 100%. The overall procedural fidelity of intervention averaged across participants was 96.13%. The mean fidelity of intervention implementation for each participant was: Adam (*M* = 97.62%; Range = 90.48% – 100%), Kevin (*M* = 98.33%; Range = 95% – 100%), Michael (*M* = 94.69%; Range = 90% – 100%), and Kylee (*M* = 93.88%; Range = 89.47% – 95.24%). Multiple strategies were integrated with the *Does It Make Sense* program to facilitate fidelity of implementation, including a behavior management system (see below for further details) and close collaboration with an E-helper. Additionally, instead of having the participants activate the annotation tool each time after posing *Does it make sense?*, the telepractice interventionist sometimes allowed children to use gestures (e.g., thumbs up/down) to indicate whether or not they understood the sentence while the interventionist placed check marks for them. This modification allowed for a swift pace of instruction delivery.

Across all study sessions, technology glitches that resulted in significant disruption occurred 2 to 3 times per participant. Unstable wireless connection was due mostly to weather (e.g., storm), unpredictable regional service outage, or proximity to the router. Additionally, we noted that children sometimes may experience elevated mental and physical fatigue after school. Fatigue can manifest in several ways including fidgeting and inability to focus. When disconnection occurred three times within a session (range = 1 – 2 times during the study period across participants) or when serious fidgeting was noted, we consulted the family to reschedule the session.

Treatment Effectiveness: Does It Make Sense via Telepractice

Given the use of a novel service delivery platform (i.e., telepractice), we were able to document three demonstrations of the treatment effect of *Does It Make Sense* with Kylee, Adam, and Kevin. We consulted evidence derived from the visual analysis, the PNDs, and the omnibus effect size (Tau) and confirmed the existence of a functional relation. Prior research suggests that typical readers on average self-correct about a third of their reading miscues whereas struggling readers self-correct less than 20% of their reading miscues (Chinn et al., 1993; Clay, 1969; Pflaum, 1979). Consistent with these findings, our participants self-corrected only 10% – 20% of their reading miscues in the baseline condition. By the time they exited the intervention, their mean self-correction proportions had reached 30% – 50%.

Of the three participants who demonstrated a PND of 100%, Kevin had the most impressive growth in self-correction proportion, which appeared to be driven simultaneously by a reduced number of reading miscues and an increased number of self-corrections throughout the intervention (Figure 2). Adam showed a between-condition change in self-correction proportion primarily driven by a decreased number of reading miscues. In the second half of the intervention, he demonstrated a within-condition change in data despite some variability. This change was driven by a marked increase in his use of self-corrections. We noted that, during the 4-minute probe assessment, Kevin and Adam evidenced several instances wherein without examiner's prompting they spontaneously drew lines on phonologically or morphologically complex words in order to solve the words by breaking them into smaller parts and sounding out. Kylee persistently produced a lower number of reading miscues throughout the intervention relative to her baseline data. She showed a significant increase in self-corrections during the early phase of intervention. Despite notable change in her reading behaviors, her self-correction proportions observed in the intervention were sometimes depressed by a substantial number of omission and commission errors of grammatical markers (e.g., past-tense -*ed*, plural -*s*, third-person singular -*s*), which did not typically cause drastic change in the meaning of the sentence. Her persistent struggle with decoding phonologically and morphologically complex words (e.g., *encourage*, *biography*) also made her data susceptible to the fluctuation introduced by passages that had slightly more instances of phonologically and morphologically complex words.

Confidence in the treatment effect demonstrated by Michael is limited because of a high percentage overlap between data points of the adjacent conditions and variable performance noted in his intervention data. Additionally, his maintenance data indicated that he did not sustain the use of self-monitoring behaviors observed in the last two intervention data points after the intervention was withdrawn.

***Does It Make Sense*: What Works for Whom and Under Which Conditions**

The iterative development of *Does It Make Sense* reflects emerging understanding in the field about the importance of using a successive set of small-scale studies to develop and refine a promising intervention (Diamond & Powell, 2011; Warren & Yoder, 1997). *Does It Make Sense* underwent two iterations that guided the program modification to improve the quality of intervention. This iterative approach also permits the evidence to converge and inform regarding the candidates who would benefit most from this program under which conditions. Analysis of child profiles and treatment outcomes suggests that two factors are important to consider when hypothesizing treatment effects: (a) automatized grapheme-phoneme knowledge and word recognition and (b) foundational language proficiency.

Automatized Grapheme-Phoneme Knowledge and Word Recognition

Pre-intervention code-related skills, specifically, the size of the pool of words that a child can instantly and effortlessly access as well as the automaticity of grapheme-phoneme correspondences appear to moderate the intervention effects. More developed automatic word recognition allows the child to allocate more cognitive resources to reading comprehension, that in turn, promotes better comprehension monitoring (Laberge & Samuels, 1974; Perfetti, 1985; Skarakis-Doyle & Dempsey, 2008). For instance, prior to intervention Michael showed the most restricted grapheme-phoneme knowledge and the least developed automatic word recognition of the four participants. Whether reading text within the intervention activities or within the assessment probes, Michael struggled to read most words in the passage due likely to his core deficit in phonological processing. When prompted to check his understanding, Michael typically attempted multiple rereads of the sentence. He showed great difficulty retaining the phonological representations of the words he had just decoded. Because of this severe word reading deficit, he often drew upon his spoken language ability to predict words that came next in text as he read (as a compensatory strategy). He sometimes appeared to be oblivious to the semantically/syntactically inappropriate sentences produced in his own reading. However, when the interventionist repeated what he just read, he had higher success in detecting inappropriate sentences. As a result, his comprehension monitoring faltered not because he was not able to comprehend or identify sentences that were semantically or syntactically inappropriate, but because laborious word reading consumed most of his limited cognitive resources. A sufficient amount of information processing effort is a prerequisite to successfully evaluating comprehension of a message (Dollaghan, 1987). Additionally, laborious word reading may induce task-related negative effects such as anxiety, frustration, or tantrum, which also compete for “common-pool resources” (e.g., attention and effortful control), leaving even fewer resources available for higher-order cognitive processes such as self-monitoring (Calkins & Bell, 2010; Pessoa, 2009).

In addition, readily accessible knowledge on grapheme-phoneme correspondences seems to be a critical precursor that fosters autonomy in applying fix-up tools. For instance, on the WRMT-III Word Identification, Adam and Michael shared the same grade-referenced standard score at the outset of this study. However, Adam had a significantly higher standard score on the WRMT-III Word Attack (i.e., nonword reading) than did Michael. Error analysis of participants’ nonword reading and reading miscues produced in oral reading suggested that Adam might experience fewer obstacles in accessing his knowledge of grapheme-phoneme correspondences and appeared to draw upon such knowledge more efficiently and consistently when he attempted to solve unknown words. Indeed, Adam demonstrated more impressive intervention outcomes and more instances of spontaneous use of the fix-up tools during guided and independent practice as compared with Michael.

Foundational Language Proficiency

Strong foundational linguistic skills also promote better comprehension monitoring. Language comprehension and reading comprehension are two highly-interrelated constructs (Scarborough, 2001). Poor language comprehension sets the limit for reading comprehension skills to advance further particularly when timely interventions are not in place. Because comprehension and comprehension monitoring share a symbiotic relation (Skarakis-Doyle & Dempsey, 2008), reading comprehension that is diminished by language deficits is often a roadblock that hinders a child with language impairment from developing and employing comprehension monitoring strategies effectively and efficiently.

In this study, the participants evidenced language vulnerabilities across multiple expressive and/or receptive language domains as informed by their language assessment results. Adam and Kylee had the lowest CELF-5 core language scores, 74 and 75 respectively. Both participants obtained a sentence comprehension score and an expressive language index score that were ranked at or below the 16^{th} percentile. Their baseline CUBED performance also confirmed language deficits in receptive and expressive language domains. Unlike Adam and Kylee, Kevin and Michael had relatively intact language comprehension as suggested by their CUBED baseline scores. Michael's CELF-5 core language score was 90. He had the highest sentence comprehension score (13) and an expressive language index score of 74. Consistent with his CUBED story comprehension performance, Kevin demonstrated relative strengths in his receptive language but notable weaknesses in his expressive language skills. His CELF-5 core language score was 81.

In fact, we noted that when students posed the question *Does it make sense?* at each sentence boundary, Kylee and Adam more often than the other participants uncritically accepted the rendered message even if it did not make sense. They displayed a delayed change and greater variability in their data patterns in contrast to Kevin, who possessed stronger foundational language skills and appeared to perform more critical evaluation of his oral rendering. Not surprisingly, children who possess poor language skills appear to have higher tolerance to violations that result in syntactically or semantically inappropriate sentences, due likely to their restricted spoken vocabulary and confusion about how sentences work in English (Leonard, 2014). They perhaps construct some sort of meaning representations when reading text but evaluate these representations in a hasty or superficial manner. If the meaning representations are not evaluated adequately, comprehension monitoring will not proceed to the second stage – *reacting effectively* unless the interventionist imposes a stronger level of scaffolding, for instance, acting surprised upon hearing the message and repeating the message to encourage re-evaluation. Furthermore, children with more severe forms of language impairment like Kylee often produce ungrammatical sentences that nevertheless convey the core meaning. These grammatical errors are unlikely to draw child attention even if the *Does It Make Sense* strategy is operated accurately.

Telepractice Services as an Emerging Alternative: What Do We Know and Where Do We Go?

The literature supporting the use of telepractice across settings and disorders is emerging (e.g., Boisvert et al., 2012; Lowe et al., 2013; Wales et al., 2017). Our study improves on earlier studies by addressing several study limitations. Hence, the question that now matters most is not *if* – but *how* telepractice can be used to support the everyday activities of clinic-based and school-based service providers (Boisvert & Hall, 2019). Based on this study, we summarize a number of benefits and limitations of telepractice as follows.

Benefits

Reducing traveling and maximizing instructional time. Converging evidence has shown that telepractice reduces time and costs associated with traveling for multisite SLPs, therefore, maximizing the usable time for instruction delivery (Hall et al., 2013; Towey, 2012). The interventionist of the present study was a doctoral student whose daily activities occurred primarily within an academic medical center. Participants' homes and schools were located about 20 to 40 minutes away from the medical center. Although the shortest geographical distance (7 miles away) may seem small, the estimated loss of productive time is about 40-60 minutes per travel session (e.g., time walking to the car, driving, parking, checking in, unpacking, getting students or sending them back to the classroom). Additionally, even within the same school, students' availability may disperse throughout the day. With telepractice, the online interventionist can reallocate her time more efficiently despite the scattered appointments. As soon as she ends the online session, she can seamlessly transition to other tasks such as lesson planning, documentation, paperwork, consultation and collaboration with other professionals.

Telepractice not only maximizes the productive time for multisite SLPs but also benefits families that are overburdened on a daily basis. With telepractice, busy parents can access clinical services for their children without worrying about traffic, weather, work schedules, or juggling the needs of siblings (Ben-Aharon, 2019). See Table 9 for parents' responses for satisfaction with telepractice intervention.

Table 9

Social Validity Survey

Survey questions		Kylee	Michael	Adam	Kevin
Delivery of reading-language services using the telepractice technology	Very good	v	v	v	v
	Above average				
	Average				
	Below average				
	Not very good				
	Do not know				
Your child's overall therapy progress in the past months	Very good	v	v	v	v
	Above average				
	Average				
	Below average				
	Not very good				
	Do not know				
The telepractice therapist's availability for communication with you regarding your child	Very good	v		v	v
	Above average		v		
	Average				
	Below average				
	Not very good				
	Do not know				
Attitude of your child about receiving reading-language therapy services via telepractice	Very good	v		v	v
	Above average				
	Average		v		
	Below average				
	Not very good				
	Do not know				
Your attitude about reading-language therapy services via telepractice	Very good	v	v	v	v
	Above average				
	Average				
	Below average				
	Not very good				
	Do not know				
How do you think telepractice compares to face-to-face therapy	Very good		v	v	v
	Above average				
	Average				
	Below average				
	Not very good				
	Do not know	v			
What is the likelihood that you would recommend telepractice to other parents	Very good	v	v	v	v
	Above average				
	Average				
	Below average				
	Not very good				
	Do not know				

Optimizing intensity of intervention. Despite recent advances in intervention research, 25% to 50% students with significant learning problems do not benefit sufficiently from effective Tier II programs (e.g., Gilbert et al., 2013; O'Connor & Fuchs, 2013; Vaughn et al., 2010). A lack of implementation features that optimize the intensity of instruction is one of the limitations that may explain students' inadequate responsiveness (Fuchs & Fuchs, 2015). Furthermore, a nation-wide survey (Brandel & Loeb, 2011) based on nearly 2000 school-based SLPs' self-reports revealed limited variability across disability and severity level with respect to program intensity and service delivery model (e.g., small-group vs. individual). Students with moderate to severe disabilities on average participated in group (not individual) intervention 2 to 3 times a week for 20-30 minutes. Students with the least severe disabilities received group intervention 1 time a week for 20-30 minutes. Decisions about service delivery was primarily driven by caseload size rather than research-substantiated evidence or students' characteristics.

Providing early access to intensive educational services can be costly; however, intensive intervention may be critical for some at-risk students to prevent their achievement gap from widening further (Fuchs et al., 2014). Two variables associated with intensifying services are, but not limited to, group size and instructional time, for instance, "dose" (i.e., the number of teaching episodes in a single session), "dose form" (e.g., drill-and-practice vs. activity-based), the time spent in each session, "dose frequency" (i.e., the number of sessions over a set period of time), or the length of treatment received (Vaughn et al., 2012; Warren et al., 2007). In this study, we utilized telepractice as an innovative tool for promoting greater treatment intensity (40-60 minutes, 2-3 times per week) than the intensity we achieved in our last iteration (15-30 minutes, 2-3 times per week; Liang & Schuele, 2019). Participants' PNDs in this study ranged between 62.5% – 100% as compared to 0% – 81.25% in the last iteration. In light of this finding, one may conclude that increased intensity is associated with better treatment outcomes. Nevertheless, this comparative analysis should be interpreted with caution as these two intervention studies differ from each other on many dimensions. Additionally, there is limited support for the hypothesis that greater dose frequency can be more beneficial than small dose frequency (Yoder, Fey, & Warren, 2012; Denton, Cirino, Barth, Romain, Vaughn, & Wexler, 2011; Ukrainetz, Ross, & Harm, 2009). There is much work needed for disentangling the effect of varying intensity levels of *Does It Make Sense*.

Notwithstanding the complexity of studying intensity, telepractice has shown promise in promoting flexible scheduling and minimizing interruption to the core instructional times in school. For instance, some of the early school-age children with intensive learning needs receive most of their services through school while maintaining their clinic-based services (e.g., once a week). One of the challenges faced by these children is missing a significant portion of classroom instruction (e.g., language arts, math, etc.) in order to access the desired services (e.g., aural habilitation) that are only available in a medical center located several miles away. With telepractice, the clinic-based SLPs can meet with the child and family remotely, for instance, within or outside regular school hours (e.g., after school). Longer time spent in each session and increased weekly sessions may be achieved, with better quality of instruction and minimal interruption to essential time and parents' work schedule.

Likewise, multisite school-based SLPs can utilize telepractice as a workload management tool in conjunction with onsite services (Boisvert & Hall, 2019). Workload refers to all activities that are required and performed by a school-based SLP (Edgar & Rosa-Lugo, 2007), including but not limited to providing direct services and evaluations, supervising, attending meetings, and conducting indirect services (e.g., consultation). Consider this scenario – a school-based SLP works primarily at an elementary school but also serves a few high school students at another site. Grouping high school students is not possible because of scheduling conflicts and the different academic classes the students attend. Traveling between two sites several times a week seems inefficient. Telepractice can help to create flexible scheduling by introducing a web-based model to the high school students with highly relevant and motivating materials. Reduced time spent traveling permits more time to be reallocated to direct and indirect services, thereby intensifying treatment by promoting better quality (e.g., lesson preparation) and quantity (e.g., instructional time).

Overcoming barriers to reciprocal SLP–parent communication. The importance of engaging families in planning, decision-making, and intervention implementation has gained much attention over the past decade (Harry, 2008; Epstein et al., 2002). Parent involvement is associated with a range of positive academic outcomes including higher class grades, test scores, aspiration, and other achievement outcomes (Grolnick & Slowiaczek, 1994; Hill et al., 2004; Hill & Tyson, 2009; Simons-Morton & Crump, 2003). Despite the intentional effort, professionals working in multicultural contexts often reported having limited contacts with parents of minority students (Young & Westernoff, 1996; Harry, 1992). Communication gaps result from factors including professionals' lack of competence and experience working with diverse populations, stereotypical assumptions, parents' limited awareness of individual rights and cultural assumptions regarding the high status of school authorities, and mismatches in cultural views of disability (Harry, 2008). In addition, typical IEP meetings, wherein goals for the next year are the primary focus of discussion, are often not conducive to family involvement. Surrounded by various professionals, families tend to feel bombarded with information and disconnected to the team process.

In the present study, we provided home-based, education-related services outside typical school hours via telepractice. Because an E-helper (a parent or grandparent) must be present to facilitate engagement and effective behavior management, the interventionist was able to establish regular communication with the caregivers to discuss progress and objectives. Regular interactions between parents and professionals may increase mutual respect and increase professionals' perceptions about families' needs and expectations and how they value education (Epstein, 2018). We shared subjective and objective observations with the families using multimedia support (e.g., video demonstration) and cheered accomplishments together. When encountering problems, we worked collaboratively with the families to brainstorm ideas and solutions. We ended many sessions with a parent debriefing time, wherein the interventionist reviewed what activities had been done, invited the child to demonstrate what strategies or ideas she or he had learned, and the rationale for teaching these skills and strategies. Even for children who were frequently imbued with reading-related frustration, we found these times were extremely precious and motivating when these young readers had the opportunities to show their parents and celebrate together the little wins they had achieved each time we worked together.

Overall the families that participated in this study reported very positive feedback on the telepractice interventionist's availability for communication with regards to their child (Table 9). In light of these preliminary findings, we postulate that telepractice may promote home-based involvement as it affirms the knowledge and instruction received at school and enhances and encourages motivations.

Particular Cautions to be Exercised in Telepractice

Although several tutorials underscore the need for an initial screening to determine if a child is a good fit for telepractice (Ben-Aharon, 2019; Boisvert & Hall, 2019; Grogan-Johnson, 2012; Grogan-Johnson, 2018), there is limited information on the core features that define telepractice eligibility and guide student selection. We learned the hard way through trial and error. One candidate who passed our initial screening in the mock treatment sessions withdrew from our study during his first week of intervention. We attribute this unsuccessful attempt to a combination of factors. First, this family reported a history of preterm birth, which is commonly associated with complex developmental and educational needs (Aarnoudse-Moens et al., 2009). This participant received not only Tier II reading instruction but also occupational therapy in school. As time went by in each session, he showed increasing difficulty maintaining the upright position in front of the screen due likely to generalized hypotonia. In addition, the child had a prolonged history of developing disruptive behaviors in response to learning-related frustration. We noted that developing physical and mental fatigue in sessions gradually resulted in a surge of disruptive behaviors. Such behaviors were not evident initially during the mock sessions due perhaps to the novelty of technology. In fact, we recommend that task-related limiting emotions and externalizing behaviors should be tackled directly through mindful designs of lessons and a behavior management system rather than digital appeal. However, functional assessments of challenging behaviors and establishing reinforcement contingencies for alternative behaviors are easier to achieve in face-to-face contexts.

Reading disabilities and attentional deficits tend to co-occur (Swanson et al., 1999). Though behavioral or attentional concerns should not be the primary factor that precludes a child from receiving telepractice services, particular caution and intentional effort are required for managing behaviors to promote better engagement. When working with young readers, a full implementation of behavior management system is strongly recommended even for those who do not have a confirmed diagnosis related to attention and/or behavior. We recommend using a combination of strategies including setting clear behavior expectations, visual schedule, material management, opportunities for choice-making, token system and long-term incentives, and behavior-specific praise (see SPARK; Lemons et al., 2018). These strategies need to be fully integrated into the daily lesson routines with E-helpers' assistance. Thus, it is crucial to have an adult as an E-helper in presence and to clearly specify E-helper's role (e.g., intermittently checking in).

Lastly, while telecommunication technology has facilitated long-distance interaction in the auditory and visual realms, reinforcement and scaffolding achieved through other sensory modality such as tactile cues is nearly not possible (Cascio et al., 2018). Social touch bears the capacity for evoking a sense of proximity and establishing human connections (Haans & IJsselsteijn, 2006). Not surprisingly, tactile cues play a critical role in facilitating positive reinforcement in education through social reinforcers like providing encouragements or emotional support (e.g., high-fives, a pat on the back). In addition, human beings are imbued with multisensory presentations from birth. Multisensory integration becomes a fundamental element of our everyday behavior (Dionne-Dostie et al., 2015). A combination of two or multiple sensory modalities may shorten response time, increase performance accuracy, and sensory acuity, therefore facilitating and optimizing learning. Not surprisingly, the presence of somatosensory or tactile information can facilitate visual and auditory processing. For instance, therapists sometimes provide hand-on-hand guidance when the most intrusive form of scaffolding is needed for learning to occur. Nevertheless, telepractice is heavily visual- and auditory-driven. Future study is needed to elucidate the impact of absent tactile cues or physical proximity on child learning contextualized within telepractice.

Limitations

Several study limitations should be acknowledged. First, the *Does It Make Sense* intervention targets a very specific skill set, that is, reading self-corrections. Findings from this study may not directly transfer to other outcome variables such as oral reading fluency and reading comprehension as these variables are indicators of a more complex construct. In fact, shortly after intervention was introduced we typically observed children slowed their reading and attempted more phrasal-level repetitions to self-correct, resulting in decelerating fluency. Through the middle phase of intervention, oral reading fluency might gradually rebound as some children gained greater accuracy. For some effortful readers, fluency might maintain at the same level despite increased use of self-repair strategies.

Additionally, we acknowledge that the *Does It Make Sense* strategy engages children in monitoring their comprehension primarily at the level of sentences. The extent to which *Does It Make Sense* may have a direct or indirect impact on text comprehension and how long it may take for the effect to occur awaits elucidation by future research. Nevertheless, we postulate that for young readers to become active participants in the process of creating meaning for text, they must begin with paying attention to sentences. Sentences are considered a fundamental element of text that carries meaning and propositions and is essential to updating situation models (i.e., text representations; van Dijk & Kintsch, 1983). The *Does It Make Sense* intervention was designed to serve this purpose, which complements other reading comprehension interventions that place greater emphasis on the macrostructure of a text.

Estimating transferred effects on reading comprehension is difficult. Far-transfer effects on commercially-available comprehension tests are typically hard to achieve despite of years of remedial teaching (Kilpatrick, 2015). However, there are limited comprehension progress-monitoring tools that are research-validated (Muijselaar et al., 2017; Petersen & Spencer, 2016; Shin et al., 2000). Despite these limitations, exploratory data from this study suggest that *Does It Make Sense* may have some beneficial impacts on reading comprehension. Children with stronger treatment effects appeared to show improvement across multiple comprehension measures (e.g., Kevin) regardless of the forms of testing. In contrast, children with complicated literacy and language needs showed more pronounced transferred effects when comprehension questions were highly structured (e.g., CUBED) or when questions required simple responses (e.g., a cloze procedure). However, caution needs to be exercised when interpreting the CUBED results because of limited testing sessions.

Second, we noted that a completely flat baseline or a clean level/trend are hard to achieve in real-world data because multiple factors influence self-corrections, for instance, nature of the reading passages, which is impossible to absolutely equate difficulty across passages. The third limitation pertains to the characteristics of participants. Because a small number of students participated, it is unclear to what extent results of this study might generalize to other students who have similar characteristics to the individuals in this study. Although all of our participants demonstrated below-average performance on multiple literacy and language measures at baseline, their text reading behaviors were notably different. No one treatment suits all struggling learners. Future research is warranted to better understand the group of students that benefit most from *Does It Make Sense*.

Lastly, the average duration of our intervention was shorter relative to other large-scale intervention studies (e.g., Denton et al., 2013). It seems promising that comprehension monitoring is a malleable skill given a short duration of time. Nevertheless, it remains unknown whether the learned strategies are sustainable over a long period of time. It is also unclear to what extent these children would spontaneously transfer these strategies to new contexts outside our tutoring sessions. Much work is needed for promoting generalization and retention through spaced teaching rather than massed-trial training (Riches, Tomasello, & Conti-Ramsden, 2005; Yoder et al., 2012).

Implications and Future Directions

We anticipate results of this study would evoke reflective thinking on how to promote better early reading comprehension instruction for children with linguistic vulnerabilities. First, we caution against overreliance on terminal feedback (e.g., immediate word/phrase supply) particularly when working with struggling readers. We acknowledge that adult terminal feedback is generally effective for average readers and allows swift reader-text interaction. Nevertheless, when the purpose of reading is remediation, we encourage parents and teachers who read together with struggling readers to shape text reading into an opportunity to exercise one's autonomy in using comprehension monitoring and word-solving strategies rather than an accuracy task. *Does It Make Sense* provides a simple strategy that is aimed to foster such autonomy. Because our strategy requires few technical skills and low-stakes training, *Does It Make Sense* may be implemented by clinicians, teachers, and parents at relative ease. Whether teacher or parent implementation of *Does It Make Sense* yield comparable effects and treatment fidelity to researchers awaits further investigation.

In brief, our intervention seems to be more effective with early school-age children who have not consistently "broken the code" but have shown emerging, readily accessible knowledge of grapheme-phoneme correspondences that continue to grow over time. In contrast, *Does It Make Sense* appears to be less effective when it is implemented alone with children who have a more intractable form of reading disability, for instance, readers who struggle with multiple aspects of reading (e.g., naming speed, phonological memory, phonological awareness, letter-sound knowledge, language). If so, we recommend teaching *Does It Make Sense* in conjunction with systematic word decoding instruction and phonological awareness training.

Lastly, we designed this study with scientific rigor in mind to address the limitations noted in prior research. Overall, results of our study, in line with the extant literature, support the use of telepractice as an appropriate service delivery model. Despite the aforementioned benefits, telepractice cannot fully replace traditional, face-to-face services in some circumstances. For instance, telepractice services mostly occur in the form of one-on-one meeting (similar to clinic-based services) whereas school-based SLPs may alternate small-group pullout sessions with classroom-based services every other week (e.g., McGinty & Justice, 2006). Moreover, telepractice entails more than simply converting the lesson contents into digital forms. Being a telepractice practitioner requires a unique set of skills that can only be acquired by doing. Mindful and careful planning before and during the implementation is critical (e.g., organization of online materials, onsite support personnel, scheduling and communication). A few resources provide an overview of the logistics and principles to facilitate the process of setting up telepractice (Ben-Aharon, 2019; Boisvert & Hall, 2019; Grogan-Johnson, 2012; Grogan-Johnson, 2018).

Conclusion

In brief, the *Does It Make Sense* comprehension monitoring strategy has been demonstrated effective to increase struggling readers' self-corrections of reading miscues, thereby improving overall reading accuracy. Preliminary evidence suggests that there may be transferred effects on reading comprehension. Long-term retention of the acquired skills and strategies awaits further investigation. Additionally, we provided quality evidence in support of the feasibility and effectiveness of delivering language-literacy intervention via telepractice using rigorous research design and a measurement system that allowed regular documentation of incremental growth during intervention. To advance intervention research further, we underscore the importance of using a successive set of small-scale studies to validate a promising intervention and mindfully aligning the goals of intervention and methods of evaluation. We also see single-strategy interventions (or reduced forms of complex instructional programs) as tantamount to multiple-strategy interventions. Lastly, we provide an overview of the facilitating factors and barriers to telepractice through our field experiences to inform future research and practices.

APPENDIX A

Telepractice Candidacy Screening Checklist

DATE OF EXAMINATION: ________________
CHILD'S CODE: ________________ **EXAMINER'S INITIALS:** ________________

(To be filled out by the research team after the training session is completed)

1. The child was:	Compliant Compliant with support Noncompliant
2. The child's activity level was:	Appropriate Too active Not active enough
3. The child was fidgety/restless	Never Some of the time Most of the time
4. The child required breaks	Never At appropriate times Too frequently (interfered with planned activities)
5. The child maintained attention:	The whole time Most of the time Some of the time Never
6. The child was distracted by task-related factors or equipment (e.g., webcam, playing w/ annotation functions, etc.)	Never Occasionally Often
7. The child engaged in off-task behaviors	Never Occasionally Often
8. The child demonstrated fatigue, boredom and/or frustration during the tasks	Never Rarely Sometimes Often
9. The child recovered from fatigue, boredom, and/or frustration:	After 15 minutes After 10 minutes After 5 minutes
10. The child demonstrated anxiety/nervousness during the test:	Most of the time Some of the time Never
11. The child requested help, support or clarification:	Never Occasionally Often

Check the box when the child/E-helper demonstrates sufficient knowledge to complete each step independently (C = child, E = E-helper)

Steps to be completed in the training session	**C**	**E**
The child/E-helper knows how to turn on and sign in to the iPad		
The child/E-helper knows where to find the most recent Zoom link in Mail		
The child/E-helper knows how to accept the link and sign in to the Zoom		
The child/E-helper knows how to adjust the volume or unmute the sound (if it is turned off unexpectedly)		
The child/E-helper knows how to activate the annotation function		
The child/E-helper knows how to write and erase		
The child/E-helper knows how to clear the page		
The child/E-helper knows how to cease the annotation function		
The child/E-helper knows how to zoom in/out		
When switching to MURAL activities, the child/E-helper knows where to find the home button that allows him/her to leave Zoom (still running in the background)		
The child/E-helper knows how to accept the link to enter MURAL		
The child/E-helper knows how to sign in to the MURAL if password is required		
The child/E-helper knows how to zoom in/out in MURAL		
The child/E-helper knows how to move cards and pictures around in MURAL		
The child/E-helper knows how to exit MURAL by pressing the home button		
The child/E-helper knows how to re-enter Zoom		

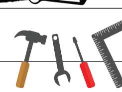
Notes:

APPENDIX B

Graphic Organizers

WHAT	Does it make sense?
WHEN	At the <u>end</u> of each sentence  
WHY	Good readers always think carefully about what they read and make self-corrections or ask questions

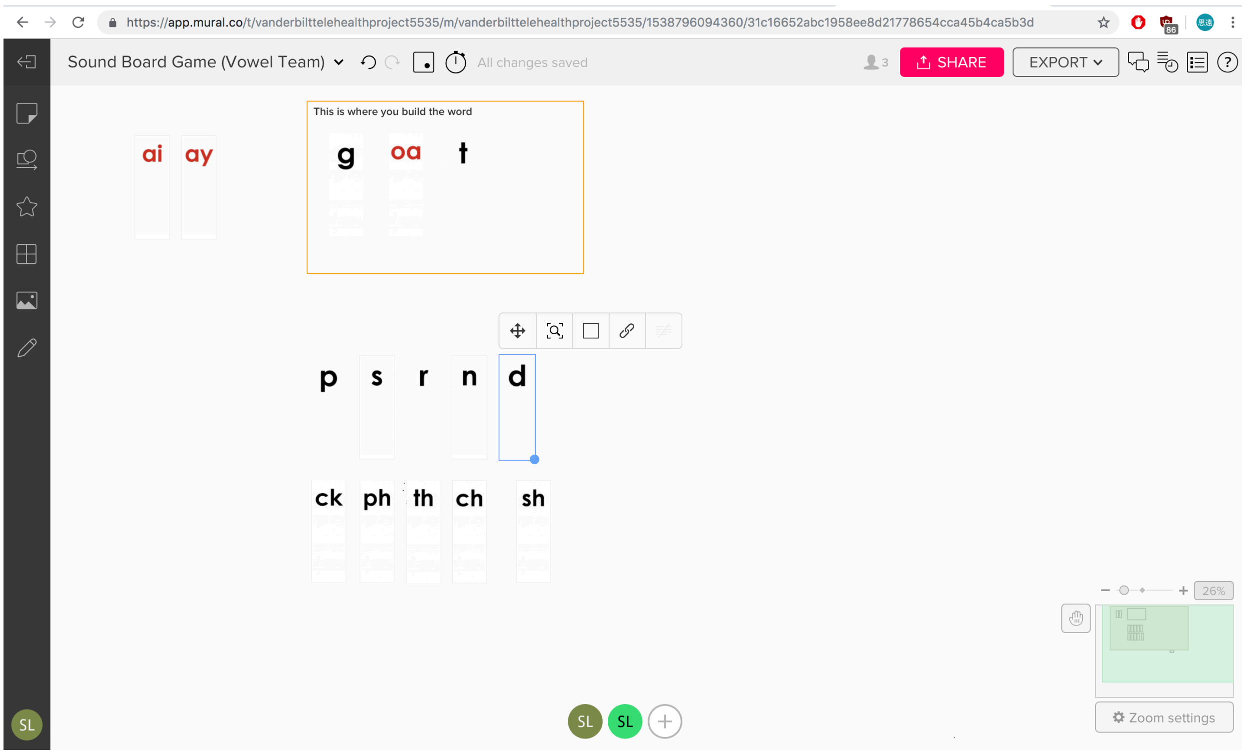
 Saying the wrong word or word ending	 Leaving out a word or word ending
 Adding a word or word ending	**I have no idea!**  Waiting longer than a few seconds  

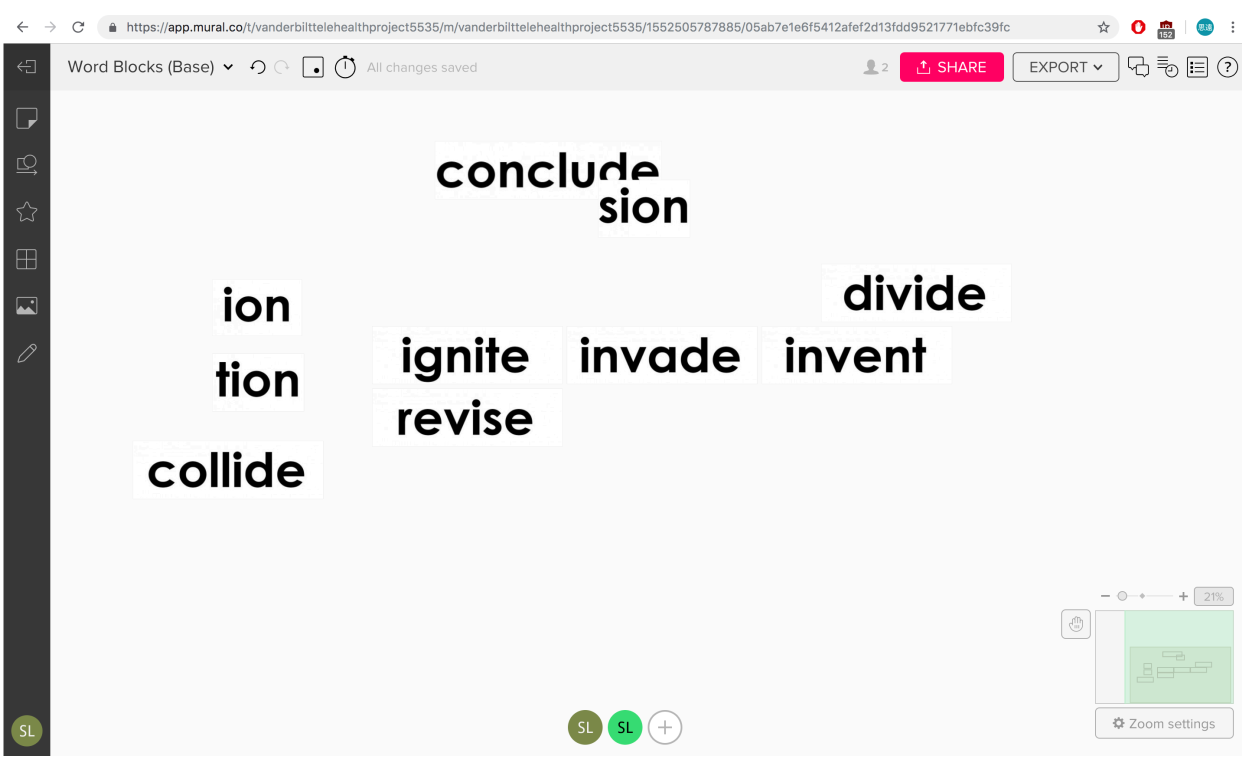
FIX-UP TOOLS 

*Read again *Sound it out *Break the word into smaller parts	Ask your teacher or friends: What is that word? What does it mean?

APPENDIX C

Sample Word Work Activities





APPENDIX D

CUBED Story Comprehension Scoring Guidelines

	2 Points	1 Point	0 Points
Question 1: Character	Main character's name/any proper name used to identify the main character	Generic character description (boy, sister) NOT pronouns	Only pronouns or generic <u>secondary</u> character (e.g., family, mom, sister, friend)
Question 2: Setting	If a student accurately provides the specific location that matches or approximates where the character was in the beginning of the story	If a student provides a general or *vague* location (e.g., outside)	Incorrect answers or when there is no answer
Question 3: Problem	Complete AND clear problem	Incomplete OR unclear problem	Incorrect answers or when there is no answer
Question 4: Attempt	Specific attempt by the main or secondary character to fix the problem using dialogue or brief description of their action (asking someone for help)	General attempt to fix the problem without dialogue or a description of action	No attempt to fix the problem
Question 5: Ending	Complete OR clear events after solving problem	Incomplete OR unclear events after solving problem	No description events after solving problem
Question 6: What will the character do the next time when he/she encounters the same problem?	If a student provides a logical, clear answer that is related to the story or clearly inferred from the story	If a student provides an answer that is somewhat likely, or possibly related to the story	If a student provides an answer that is improbable, does not answer the question, or is unrelated to the story

Note: Adapted from Petersen, D. B., & Spencer, T. D. (2016). *CUBED*. Language Dynamics Group.

APPENDIX E

Procedural Fidelity

Fidelity of Intervention Implementation

Steps	Yes (enter 1) No (enter 0)	Notes
Step 1: Intro/Review/Vocabulary Preview, Time:		
First day of intervention: Does the interventionist provide the intro statement?		
Subsequent days: Does the interventionist review the does-it-make-sense strategy?		
Subsequent days: Does the interventionist review the fix-up tools?		
Subsequent days: Are both visual organizers presented?		
Subsequent days: 3-5 written words that might be unfamiliar to the child were previewed		
Step 2: I DO Demonstrate the strategy via explicit modeling, Time:		
First day: Does the interventionist explicitly explain the comprehension monitoring strategy (does-it-make-sense)?		
First day: Does the interventionist check student's understanding of the WHAT/WHEN/WHY?		
First week: Does the interventionist introduce a new type of mistake (if applicable) and also review those introduced in previous lessons?		
Does the interventionist show appropriate use of sentences or passages to demonstrate the does-it-make-sense strategy?		
Does the interventionist explicitly model how to use the fix-up tools or encourage the student to reflect on the fix-up tools? (On the first day, the interventionist may help the student type them up)		
Step 3: WE DO Practice the skill with student, Time:		
Do the interventionist and the student take turns reading?		
Does the interventionist provide corrective feedback if the does-it-make-sense strategy is not implemented correctly by the student? For instance, prompt the student to use the fix-up tools if they do not use them spontaneously after detecting an error, or guide the student to re-evaluate his/her own comprehension if does-it-make-sense is applied in a cursory manner by the student		
Does the interventionist make mistakes on purpose in her reading/or act out getting stuck on a word when she's reading?		
Does the interventionist demonstrate how to fix the mistakes?		
Does the interventionist intermittently check student's local comprehension beyond the sentence level?		
Does the clinician correctly respond to those non-teachable moments (e.g., omitting grammatical markers, mistaking on irregular past tense; check the manual for examples)?		

Step 4: YOU DO Student practices the strategy independently, Time:		
Does the interventionist only provide scaffolds if the student was not successful in applying the strategy?		
Does the interventionist correctly respond to those non-teachable moments?		
Does the interventionist prompt the student to use the fix-up tools if they do not seem to use them after detecting a reading mistake?		
Step 5: Daily checks/Wrap-up/Graphing, Time:		
Does the interventionist go through the vocabulary at the end of the lesson?		
Does the interventionist briefly review what has been taught in today's lesson?		
Does the interventionist show the progress monitoring graph to the student? (optional)		
General observation		
Are the materials ready for each activity?		
Expectations were clear (e.g., lesson plan posted, stated, or referred to)		
Does the interventionist provide specific feedback and/or verbal/nonverbal encouragement and praise on an ongoing basis?		
Does the interventionist provide appropriate 'think time'?		
Students are on-task and/or off-task behavior is addressed		

Probe Assessment

General Administration	Yes (enter 1) / No (enter 0)	Comments
Does the examiner read the directions prior to child reading? (Simplified directions may be used once the child is used to the routine)		
Did the examiner turn on the video recording feature of ZOOM?		
The passage is presented appropriately or viewed by the student		
Does student read no less than four minutes or complete the passage?		
Does the interventionist provide no instructional feedback during testing?		
The interventionist only provided neutral prompts such as "'just keep reading" "mhm" or "you may skip the word" when student was stuck on a word.		

REFERENCES

Aarnoudse-Moens, C. S. H., Weisglas-Kuperus, N., van Goudoever, J. B., & Oosterlaan, J. (2009). Meta-analysis of neurobehavioral outcomes in very preterm and/or very low birth weight children. *Pediatrics*, *124*, 717-728.

Alexander, K. L., Entwisle, D. R., & Olson, L. S. (2007). Lasting consequences of the summer learning gap. *American Sociological Review*, *72*, 167-180.

Allington, R. L. (1980). Poor readers don't get to read much in reading groups. *Language Arts*, *57*, 872-876.

Allington, R. L. (2013). What really matters when working with struggling readers. *The Reading Teacher*, *66*, 520-530.

American Speech-Language-Hearing Association (n.d.). *Telepractice Practice Portal*. http://www.asha.org/Practice-Portal/Professional-Issues/Telepractice.

American Speech-Language-Hearing Association. (n.d.). *Roles and responsibilities of speech-language pathologists in school*. http://www.asha.org/policy/PI2010-00317.

Blachman, B. A., & Tangel, D. M. (2008). *Road to reading: A program for preventing and remediating reading difficulties*. Paul H. Brookes Publishing.

Calkins, S. D., & Bell, M. A. (2010). *Child development at the intersection of emotion and cognition*. American Psychological Association.

Bereiter, C., Brown, A., Campione, J., Carruthers, I., Case, R., Hirshberg, J., & Treadway, G. H. (2002). *Open court reading*. McGraw-Hill Education.

Ben-Aharon, A. (2019). A practical guide to establishing an online speech therapy private practice. *Perspectives of the ASHA Special Interest Groups*, *4*, 712-718.

Boisvert, M. K., & Hall, N. (2019). Telepractice for school-based speech and language services: A workload management strategy. *Perspectives of the ASHA Special Interest Groups*, *4*, 211-216.

Boisvert, M., Hall, N., Andrianopoulos, M., & Chaclas, J. (2012). The multi-faceted implementation of telepractice to service individuals with autism. *International Journal of Telerehabilitation*, *4*, 11.

Boswell, S. (2006, August 1). *Seven states begin work on personnel shortages, qualifications*. The ASHA Leader. http://www.leader.pubs.asha.org/doi/10.1044/leader.AN1.11102006.1

Brandel, J., & Loeb, D. F. (2011). Program intensity and service delivery models in the schools: SLP survey results. *Language, Speech, and Hearing Services in Schools, 42,* 461-490.

Brown, R. (2008). The road not yet taken: A transactional strategies approach to comprehension instruction. *The Reading Teacher, 61*, 538-547.

Brown, L., Sherbenou, R., & Johnsen, S. (2010). *Test of Nonverbal Intelligence-Fourth Edition*. Pro-Ed.

Cain, K., & Oakhill, J. (2011). Matthew effects in young readers: Reading comprehension and reading experience aid vocabulary development. *Journal of Learning Disabilities, 44*, 431-443.

Carver, R. P., & Leibert, R. E. (1995). The effect of reading library books at different levels of difficulty upon gain in reading ability. *Reading Research Quarterly*, *30,* 26-48.

Cascio, C. J., Moore, D., & McGlone, F. (2019). Social touch and human development. *Developmental Cognitive Neuroscience*, *35*, 5-11.

Catts, H. W., Adlof, S. M., & Weismer, S. E. (2006). Language deficits in poor comprehenders: A case for the simple view of reading. *Journal of Speech, Language, and Hearing Research, 49*, 278-293.

Scarborough, H. S. (2005). Developmental relationships between language and reading: Reconciling a beautiful hypothesis with some ugly facts. In H. W. Catts & A. G. Kamhi (Eds.), *The connection between language and reading disabilities* (pp.3-24). Lawrence Erlbaum Associate Publishers.

Chin, T., & Phillips, M. (2004). Social reproduction and child-rearing practices: Social class, children's agency, and the summer activity gap. *Sociology of Education*, *77*, 185-210.

Chinn, C. A., Waggoner, M. A., Anderson, R. C., Schommer, M., & Wilkinson, I. A. (1993). Situated actions during reading lessons: A microanalysis of oral reading error episodes. *American Educational Research Journal, 30*, 361-392.

Ciccia, A. H., Whitford, B., Krumm, M., & McNeal, K. (2011). Improving the access of young urban children to speech, language and hearing screening via telehealth. *Journal of Telemedicine and Telecare, 17*, 240-244.

Clay, M. M. (1969). Reading errors and self-correction behaviour. *British Journal of Educational Psychology, 39*, 47-56.

Clay, M. M. (2001). *Change over time in children's literacy development*. Heinemann Educational Books.

Cleave, P. L., Becker, S. D., Curran, M. K., Van Horne, A. J. O., & Fey, M. E. (2015). The efficacy of recasts in language intervention: A systematic review and meta-analysis. *American Journal of Speech-Language Pathology, 24,* 237-255.

Cooper, J. D., Chard, D. J., & Kiger, N. D. (2006). *The struggling reader: Interventions that work.* Scholastic.

Cooper, H., Nye, B., Charlton, K., Lindsay, J., & Greathouse, S. (1996). The effects of summer vacation on achievement test scores: A narrative and meta-analytic review. *Review of Educational Research*, *66*, 227-268.

Coufal, K., Parham, D., Jakubowitz, M., Howell, C., & Reyes, J. (2018). Comparing traditional service delivery and telepractice for speech sound production using a functional outcome measure. *American Journal of Speech-Language Pathology, 27*, 82-90.

Diamond, K. E., & Powell, D. R. (2011). An iterative approach to the development of a professional development intervention for Head Start teachers. *Journal of Early Intervention*, *33*, 75-93.

Dionne-Dostie, E., Paquette, N., Lassonde, M., & Gallagher, A. (2015). Multisensory integration and child neurodevelopment. *Brain Sciences*, *5*, 32-57.

Dollaghan, C. A. (1987). Comprehension monitoring in normal and language-impaired children. *Topics in Language Disorders*, *7,* 45-60.

Dollaghan, C. A., & Kaston, N. (1986). A comprehension monitoring program for language-impaired children. *Journal of Speech and Hearing Disorders, 51*, 264-271.

Denton, C. A., Barth, A. E., Fletcher, J. M., Wexler, J., Vaughn, S., Cirino, P. T., Romain, M., & Francis, D. J. (2011). The relations among oral and silent reading fluency and comprehension in middle school: Implications for identification and instruction of students with reading difficulties. *Scientific Studies of Reading*, *15*, 109-135.

Denton, C. A., Tolar, T. D., Fletcher, J. M., Barth, A. E., Vaughn, S., & Francis, D. J. (2013). Effects of tier 3 intervention for students with persistent reading difficulties and characteristics of inadequate responders. *Journal of Educational Psychology, 105*, 633-648.

Edgar, D. L., & Rosa-Lugo, L. I. (2007). The critical shortage of speech-language pathologists in the public school setting: Features of the work environment that affect recruitment and retention. *Language, Speech, and Hearing Services in Schools, 38,* 31-46.

Epstein, J. L. (2018). *School, family, and community partnerships: Preparing educators and improving schools*. Routledge.

Epstein, J. L., Sanders, M. G., Sheldon, S. B., Simon, B. S., Salinas, K. C., Jansorn, N. R., Van Voorhis, F. L., Martin, C. S., Thomas, B. G., Greenfeld, M. D., Hutchins, D. J., Williams, K. J. (2002). *Family, school, and community partnerships: Your handbook for action.* SAGE Publications.

Eriks-Brophy, A., Quittenbaum, J., Anderson, D., & Nelson, T. (2008). Part of the problem or part of the solution? Communication assessments of Aboriginal children residing in remote communities using videoconferencing. *Clinical Linguistics & Phonetics, 22*, 589-609.

Evans, M. A., Barraball, L., & Eberle, T. (1998). Parental responses to miscues during child-to-parent book reading. *Journal of Applied Developmental Psychology, 19*, 67-84.

Evans, M. A., Moretti, S., Shaw, D., & Fox, M. (2003). Parent scaffolding in children's oral reading. *Early Education & Development, 14*, 363-388.

Foorman, B. R., Schatschneider, C., Eakin, M. N., Fletcher, J. M., Moats, L. C., & Francis, D. J. (2006). The impact of instructional practices in grades 1 and 2 on reading and spelling achievement in high poverty schools. *Contemporary Educational Psychology, 31*, 1-29.

Fuchs, D., Fuchs, L. S., & Vaughn, S. (2014). What is intensive instruction and why is it important? *Teaching Exceptional Children, 46*, 13-18.

Fuchs, L. S. (2017). Curriculum-based measurement as the emerging alternative: Three decades later. *Learning Disabilities Research & Practice, 32*, 5-7.

Gabel, R., Grogan-Johnson, S., Alvares, R., Bechstein, L., & Taylor, J. (2013). A field study of telepractice for school intervention using the ASHA NOMS K-12 database. *Communication Disorders Quarterly, 35*, 44-53.

Gast, D. L. (2009). *Single subject research methodology in behavioral sciences*. Routledge.

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

Gillam, R. B., & Pearson, N. A. (2004). *TNL: Test of Narrative Language*. Pro-Ed.

Grolnick, W. S., & Slowiaczek, M. L. (1994). Parents' involvement in children's schooling: A multidimensional conceptualization and motivational model. *Child Development*, *65*, 237-252.

Grogan-Johnson, S. (2012). Providing school-based speech-language therapy services by telepractice: A brief tutorial. *Perspectives on Telepractice*, *2*, 42-48.

Grogan-Johnson, S. (2018). Getting started in school-based speech-language pathology telepractice. *Perspectives of the ASHA Special Interest Groups*, *3*, 21-31.

Grogan-Johnson, S., Alvares, R., Rowan, L., & Creaghead, N. (2010). A pilot study comparing the effectiveness of speech language therapy provided by telemedicine with conventional on-site therapy. *Journal of Telemedicine and Telecare, 16*, 134-139.

Grogan-Johnson, S., Gabel, R. M., Taylor, J., Rowan, L. E., Alvares, R., & Schenker, J. (2011). A pilot exploration of speech sound disorder intervention delivered by telehealth to school-age children. *International Journal of Telerehabilitation, 3*, 31.

Grogan-Johnson, S., Schmidt, A. M., Schenker, J., Alvares, R., Rowan, L. E., & Taylor, J. (2013). A comparison of speech sound intervention delivered by telepractice and side-by-side service delivery models. *Communication Disorders Quarterly, 34*, 210-220.

Haans, A., & IJsselsteijn, W. (2006). Mediated social touch: A review of current research and future directions. *Virtual Reality*, *9*, 149-159.

Hall, N., Boisvert, M., & Steele, R. (2013). Telepractice in the assessment and treatment of individuals with aphasia: A systematic review. *International Journal of Telerehabilitation*, *5*, 27.

Harry, B. (1992). Restructuring the participation of African-American parents in special education. *Exceptional children*, *59*, 123-131.

Harry, B. (2008). Collaboration with culturally and linguistically diverse families: Ideal versus reality. *Exceptional children*, *74*, 372-388.

Hetherton, M. B. (2013). *Treatment of foundational reading skills through telepractice and face-to-face environments: Single subject design.* (Publication No. 802) [Doctoral dissertation, University of Massachusetts Amherst]. Open Access Dissertations.

Hill, N. E., & Taylor, L. C. (2004). Parental school involvement and children's academic achievement: Pragmatics and issues. *Current Directions in Psychological Science*, *13*, 161-164.

Hill, N. E., & Tyson, D. F. (2009). Parental involvement in middle school: a meta-analytic assessment of the strategies that promote achievement. *Developmental Psychology*, *45*, 740.

Houge, T. T., & Geier, C. (2009). Delivering one-to-one tutoring in literacy via videoconferencing. *Journal of Adolescent & Adult Literacy, 53*, 154-163.

Katz, L. A., Maag, A., Fallon, K. A., Blenkarn, K., & Smith, M. K. (2010). What makes a caseload (un) manageable? School-based speech-language pathologists speak. *Language, Speech, and Hearing Services in Schools*, *41,* 139-151.

Kilpatrick, D. A. (2015). *Essentials of assessing, preventing, and overcoming reading difficulties*. John Wiley & Sons.

Kim, J. S., & White, T. G. (2008). Scaffolding voluntary summer reading for children in grades 3 to 5: An experimental study. *Scientific Studies of Reading*, *12*, 1-23.

LaBerge, D., & Samuels, S. J. (1974). Toward a theory of automatic information processing in reading. *Cognitive Psychology*, *6*, 293-323.

Lemons, C.J., Lambert, J.M., & Preacher, K.J. (2018, July 1). *An efficacy trial to evaluate supporting paraprofessionals by advancing reading intervention knowledge and skill (SPARK)*. Institute of Education Sciences, U.S. Department of Education. https://ies.ed.gov/funding/grantsearch/details.asp?ID=2209.

Leonard, L. B. (2011). The primacy of priming in grammatical learning and intervention: A tutorial. *Journal of Speech, Language, and Hearing Research, 54*, 608-621.

Laing, S. P. (2002). Miscue analysis in school-age children. *American Journal of Speech-Language Pathology*, *11,* 407-416.

Ledford, J. R., & Gast, D. L. (Eds.). (2009). *Single subject research methodology in behavioral sciences: Applications in special education and behavioral sciences*. Routledge.

Leonard, L. B. (2014). *Children with specific language impairment*. MIT Press.

Liang, S. Y., & Schuele, C. M. (2017). Does it make sense? Evaluation of a comprehension monitoring strategy to improve reading accuracy [Unpublished doctoral pre-dissertation study]. Vanderbilt University.

Liang, S. Y., Lemons, C. J., & Schuele, C. M. (2017, February 2-4). *Does it make sense? Reading for understanding.* Pacific Coast Research Conference, San Diego, CA, United States.

Liang, S. Y. & Schuele, C. M. (2019, November 19-21). *Does it make sense? A comprehension monitoring strategy to improve reading accuracy.* American Speech-Language-Hearing Association Convention, Orlando, FL, United States.

Lowe, R., O'Brian, S., & Onslow, M. (2013). Review of telehealth stuttering management. *Folia Phoniatrica et Logopaedica*, *65*, 223-238.

Mansell, J., Evans, M. A., & Hamilton-Hulak, L. (2005). Developmental changes in parents' use of miscue feedback during shared book reading. *Reading Research Quarterly, 40*, 294-317.

McGinty, A. S., & Justice, L. (2006). Classroom-based versus pull-out interventions: A review of the experimental evidence. *EBP Briefs, 1,* 1-25.

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

Muijselaar, M. M., Kendeou, P., de Jong, P. F., & van den Broek, P. W. (2017). What does the CBM-Maze test measure? *Scientific Studies of Reading*, *21*, 120-132.

Nation, K., Clarke, P., Marshall, C. M., & Durand, M. (2004). Hidden language impairments in children: Parallels between poor reading comprehension and specific language impairment? *Journal of Speech, Language, and Hearing Research, 47*, 199-211.

Nelson, N. W. (2010). *Language and literacy disorders: Infancy through adolescence*. Allyn & Bacon.

Newcomer, P. L., & Hammill, D. D. (2008). *TOLD-P: 4: Test of Language Development-Primary*: *Fourth Edition.* Pro-Ed.

O'Connor, R. E., & Fuchs, L. S. (2013). Responsiveness to intervention in the elementary grades: Implications for early childhood education. In V. Buysse & E. S. Peisner-Feinberg (Eds.), *Handbook of response to intervention in early childhood* (pp. 41-55). Paul H. Brookes Publishing.

Paris, S. G., Cross, D. R., & Lipson, M. Y. (1984). Informed strategies for learning: A program to improve children's reading awareness and comprehension. *Journal of Educational Psychology, 76*, 1239.

Parker, R. I., Vannest, K. J., Davis, J. L., & Sauber, S. B. (2011). Combining nonoverlap and trend for single-case research: Tau-U. *Behavior Therapy*, *42*, 284-299.

Perfetti, C. A. (1985). *Reading ability*. Oxford University Press.

Perfetti, C. A. (1992). The representation problem in reading acquisition. In P. B. Gough, L. C. Ehri, & R. Treiman (Eds.)*, Reading acquisition* (pp. 145-174). Lawrence Erlbaum Associates, Inc.

Pessoa, L. (2009). How do emotion and motivation direct executive control? *Trends in Cognitive Sciences, 13*, 160-166.

Petersen, D. B., Brown, C. L., Ukrainetz, T. A., Wise, C., Spencer, T. D., & Zebre, J. (2014). Systematic individualized narrative language intervention on the personal narratives of children with autism. *Language, Speech, and Hearing Services in Schools, 45*, 67-86.

Petersen, D., & Spencer, T. (2016). *CUBED*. Language Dynamics Group.

Pflaum, S. W. (1979). Diagnosis of oral reading. *The Reading Teacher, 33*, 278-284.

Pressley, M., & Hilden, K. (2004). Toward more ambitious comprehension instruction. In E. Silliman & L. Wilkinson (Eds.), *Language and lliteracy learning in schools* (pp. 151-174). Guildford.

Riches, N. G., Tomasello, M., & Conti-Ramsden, G. (2005). Verb learning in children with SLI. *Journal of Speech, Language, and Hearing Research, 48,* 1397-411.

Rudolph, J. M., & Rudolph, S. (2015). Telepractice vs. on-site treatment: Are outcomes equivalent for school-age children? *EBP Briefs,* 1-15.

Scarborough, H. S., (2001). Connecting early language and literacy to later reading (dis)abilities: Evidence, theory, and practice. In S. Neuman & Dickinson (Eds.), *Handbook for research in early literacy* (pp. 23-38). Guilford Press.

Schuele, C. M. (2017). The speech-language pathologist's role in reading and writing: In theory meets in reality. *Perspectives of the ASHA Special Interest Groups*, *2*, 115-116.

Shanahan, T., Callison, K., Carriere, C., Duke, N. K., Pearson, P. D., Schatschneider, C., & Torgesen, J. (2010). Improving reading comprehension in kindergarten through 3rd grade: A practice guide (NCEE 2010-4038). National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. http://whatworks.ed.gov/publications/practiceguides

Shin, J., Deno, S. L., & Espin, C. (2000). Technical adequacy of the maze task for curriculum-based measurement of reading growth. *The Journal of Special Education, 34,* 164-172.

Simons-Morton, B. G., & Crump, A. D. (2003). Association of parental involvement and social competence with school adjustment and engagement among sixth graders. *Journal of School Health*, *73*, 121-126.

Skarakis-Doyle, E., & Dempsey, L. (2008). The detection and monitoring of comprehension errors by preschool children with and without language impairment. *Journal of Speech, Language, and Hearing Research, 51*, 1227-1243.

Snow, C. E. (2014). Input to interaction to instruction: three key shifts in the history of child language research. *Journal of Child Language, 41*, 117-123.

Stahl, K. A. D. (2004). Proof, practice, and promise: Comprehension strategy instruction in the primary grades. *The Reading Teacher*, *57*, 598-609.

Stecker, P. M., Fuchs, D., & Fuchs, L. S. (2008). Progress monitoring as essential practice within response to intervention. *Rural Special Education Quarterly, 27*, 10-17.

Sutherland, R., Trembath, D., Hodge, A., Drevensek, S., Lee, S., Silove, N., & Roberts, J. (2017). Telehealth language assessments using consumer grade equipment in rural and urban settings: Feasible, reliable and well tolerated. *Journal of Telemedicine and Telecare, 23*, 106-115.

Swanson, H. L., Mink, J., & Bocian, K. M. (1999). Cognitive processing deficits in poor readers with symptoms of reading disabilities and ADHD: More alike than different? *Journal of Educational Psychology*, *91*, 321.

Tarlow, K. R. (2016). Baseline Corrected Tau Calculator. Available from www.ktarlow.com/stats/tau

Tarlow, K. R. (2017). An improved rank correlation effect size statistic for single-case designs: Baseline corrected Tau. *Behavior Modification*, *41*, 427-467.

Taylor, O. D., Armfield, N. R., Dodrill, P., & Smith, A. C. (2014). A review of the efficacy and effectiveness of using telehealth for paediatric speech and language assessment. *Journal of Telemedicine and Telecare, 20*, 405-412.

Theodoros, D. (2011). Telepractice in speech-language pathology: The evidence, the challenges, and the future. *Perspectives on Telepractice, 1*, 10-21.

Therrien, W. J. (2004). Fluency and comprehension gains as a result of repeated reading: A meta-analysis. *Remedial and Special Education*, *25*, 252-261.

Tomblin, J. B., & Nippold, M. A. (2014). *Understanding individual differences in language development across the school years*: Psychology Press.

Torgesen, J. K., Wagner, R. K., & Rashotte, C. A. (2012). *TOWRE-2 Examiner's manual.* Pro-Ed.

Towey, M. P. (2012). Speech telepractice: Installing a speech therapy upgrade for the 21st century. *International Journal of Telerehabilitation*, *4*, 73.

Ukrainetz, T. A., Ross, C. L., & Harm, H. M. (2009). An investigation of treatment scheduling for phonemic awareness with kindergartners who are at risk for reading difficulties. *Language, Speech, and Hearing Services in Schools, 40,* 86-100.

Van Dijk, T. A., & Kintsch, W. (1983). *Strategies of discourse comprehension*. Academic Press.

Vaughn, S., Cirino, P. T., Wanzek, J., Wexler, J., Fletcher, J. M., Denton, C. D., Barth, A., Romain, M., & Francis, D. J. (2010). Response to intervention for middle school students with reading difficulties: Effects of a primary and secondary intervention. *School Psychology Review*, *39*, 3-21.

Vaughn, S., Wanzek, J., Murray, C. S., & Roberts, G. (2012). *Intensive interventions for students struggling in reading and mathematics: A practice guide.* RMC Research Corporation, Center on Instruction.

von Koss Torkildsen, J., Dailey, N. S., Aguilar, J. M., Gómez, R., & Plante, E. (2013). Exemplar variability facilitates rapid learning of an otherwise unlearnable grammar by individuals with language-based learning disability. *Journal of Speech, Language, and Hearing Research, 56*, 618-629.

Waite, M. C., Cahill, L. M., Theodoras, D. G., Busuttin, S., & Russell, T. G. (2006). A pilot study of online assessment of childhood speech disorders. *Journal of Telemedicine and Telecare, 12*, 92-94.

Waite, M. C., Theodoros, D. G., Russell, T. G., & Cahill, L. M. (2010). Internet-based telehealth assessment of language using the CELF–4. *Language, Speech, and Hearing Services in Schools, 41*, 445-458.

Waite, M. C., Theodoros, D. G., Russell, T. G., & Cahill, L. M. (2012). Assessing children's speech intelligibility and oral structures, and functions via an Internet-based telehealth system. *Journal of Telemedicine and Telecare, 18*, 198-203.

Wales, D., Skinner, L., & Hayman, M. (2017). The efficacy of telehealth-delivered speech and language intervention for primary school-age children: a systematic review. *International Journal of Telerehabilitation*, *9*, 55.

Walker, E., & Nowacki, A. S. (2011). Understanding equivalence and noninferiority testing. *Journal of General Internal Medicine, 26*, 192-196.

Warren, S. F., Fey, M. E., & Yoder, P. J. (2007). Differential treatment intensity research: A missing link to creating optimally effective communication interventions. *Mental Retardation and Developmental Disabilities Research Reviews*, *13*, 70-77.

Wechsler, D. (2009). *Wechsler Individual Achievement Test-Third Edition.* Pearson.

Weidner, K., & Lowman, J. (2020). Telepractice for adult speech-language pathology services: A systematic review. *Perspectives of the ASHA Special Interest Groups*, *5*, 326-338.

Wiig, E. H., Semel, E., & Secord, W. A. (2013). *Clinical Evaluation of Language Fundamentals-Fifth Edition.* Pearson.

Woodcock, R. W. (2011). *Woodcock Reading Mastery Tests-Third Edition.* Pearson.

Yoder, P., Fey, M. E., & Warren, S. F. (2012). Studying the impact of intensity is important but complicated. *International Journal of Speech-Language Pathology*, *14*, 410-413.

Yoder, P. J., Lloyd, B. P., & Symons, F. J. (2018). *Observational measurement of behavior*. Paul H. Brookes Publishing.

Young, T., & Westernoff, F. (1996). Overcoming barriers to effective parental partnerships: Implications for professionals in an educational setting. *Journal of Educational Issue of Language Minority Students*, 16, 1-9.