Abstract

This article reports on a study designed to determine if the lowest achieving first grade students who were identified by their school districts as at-risk for dyslexia can be distinguished from students who have initial reading and writing difficulties but did not present dyslexia characteristics. Thirty-six first grade students from two different school districts participated in this quantitative study. As part of the study, students were additionally screened with the *Observation Survey of Early Literacy Achievement,* the *Feifer Assessment of Reading,* and the *Slosson Oral Reading Test-Revised* pre- and post-intervention. Characteristics, effectiveness, validity, and reliability of the assessment tools are included in the discussion. Upon receiving Reading Recovery® as a first intervention, tests indicated very large effect sizes on all measures for all children. Students whose initial screening indicated no dyslexia characteristics made greater literacy gains.

An Early Intervention Comparison Study of Children With and Without Dyslexia Characteristics

When students are experiencing difficulty learning to read and write, early intervention is critical. If interventions are implemented in early schooling, expert teaching often can dramatically change the trajectory of students' learning. In recent years, many states have passed "dyslexia laws" to create uniform guidance and requirements among schools in that state regarding screening, testing, intervention, teacher training, and funding in order to support students' reading progress. Dyslexia is designated as a type of Specific Learning Disability (SLD) that falls under the umbrella of special education in the Individuals with Disabilities in Education Act (IDEA).

In the U.S., 47 states have passed dyslexia legislation (Davis Dyslexia Association International, 2021). Many require that all kindergarten and first grade students are screened for dyslexia characteristics; however, IDEA does not provide a definition of dyslexia. The state education agency in the southern state where this study takes place recognizes any reading difference as a characteristic of dyslexia, and at present there are no screening devices available that can identify differences between a general reading difficulty and dyslexia. In fact, vanDerHeyden & Burns (2017) report that "current efforts at dyslexia screening are misleading about 50% of the time" (p. 25). Although many state education codes require identification and intervention for students at risk for dyslexia, recommended interventions are often long-term, highly systematic, sequenced, one-size-fits-all phonics approaches (e.g., Arkansas, Colorado, and Texas). There is a need for effective, short-term interventions that provide differentiated instruction for readers who initially struggle with literacy learning. Such interventions would

support many students' literacy acquisition and help identify students who truly need long-term instructional support.

With more than 30 years of research and evaluation, Reading Recovery® is one of the world's most widely studied and successful short-term early literacy interventions (May et al., 2016; Pinnell et al., 1988; Schwartz, 2005; US Department of Education, 2013). Highly trained teachers work with first grade students daily in one-to-one, individually designed lessons for a period of 12 to 20 weeks (or 30-50 hours of instruction). Reading Recovery teachers serve the lowest achieving students in the first grade cohort, regardless of the nature of the reading difficulty. Annually in the United States, 71% of children who receive a full series of Reading Recovery lessons make accelerated progress and catch up to their peers (International Data Evaluation Center, 2019). In schools that implement Reading Recovery, first grade students who are identified as exhibiting dyslexia characteristics, may receive 12 to 20 weeks of Reading Recovery as their first intervention. If reading difficulties persist after Reading Recovery, students may be referred for further evaluation and placement in a longer term intervention, one of which might be a dyslexia intervention.

Educators need highly-successful, short-term intervention approaches that provide differentiated instruction and can separate those students who can learn to read within the average of their first grade cohort with short-term supplemental teaching from those who need to be referred for a lengthier, purely synthetic phonics approach. Additionally, children and schools would benefit from assessments that are sensitive enough to help teachers know which students might benefit most from a phonics-based dyslexia intervention.

The purpose of the pilot study presented here is to determine if first grade students who have been identified as at-risk for dyslexia can be distinguished from students who have initial

reading and writing difficulties but do not have dyslexia characteristics. This research is guided by four questions. The first two questions relate to understanding dyslexia identification, and the last two questions focus on understanding differences in students identified with and without dyslexia characteristics through comparison:

1. What is the relationship of the criterion scores used for Reading Recovery intervention on the *Observation Survey of Early Literacy Achievement* (OS) test and the *Slosson Oral Reading Assessment* (SORT-R) and how do they compare to the Total Score from the *Feifer Assessment of Reading* (FAR)?
2. Is there a significant difference in the mean OS criterion scores and each of the four FAR indices at pre-intervention?
3. Is there a more prevalent FAR subtype among students identified for Reading Recovery?
4. What is the effectiveness of the Reading Recovery intervention as assessed by changes to pre-and-post intervention scores of the OS, SORT, and the FAR?

Theoretical Framework

Reading Recovery

Reading Recovery is based on a literacy processing theory of instruction (Clay, 2001; Doyle, 2019). The theoretical base derives from cognitive psychology and aligns with the cognitive apprenticeship model of instruction (Rogoff, 1990). Clay's complex theory focuses on the specific perceptual and cognitive behaviors involved in reading and writing and explanations of the integration of complex in-the-head processes. Thus, descriptions of progress involve perceptual and cognitive working systems directed to complete meaning based reading and writing tasks. Phonological and orthographic information is taught in the service of reading and

writing continuous texts, and is addressed briefly in isolation as well. Based on this theory, reading is "a message-getting, problem-solving activity, which increases in power and flexibility the more it is practiced. It is complex because within the directional constraints of written language, verbal and perceptual behaviors are purposefully directed in some integrated way to the problem of extracting sequences of information from texts to yield meaningful and specific communications" (Clay, 2001, p. 1).

Clay's complex literacy processing theory rejects the idea that a singular, underlying cause of reading difficulty exists. Rather, students experience learning challenges due to a variety of factors that are unique to the individual learner. This view, therefore, engenders a number of instructional implications. Teachers must design a unique series of individual lessons for each student so they can respond to each child's idiosyncratic response repertoire, thus forging the quickest possible route to reading and writing proficiency (Clay, 2001; 2016). Instruction aims to accelerate the pace of learning for each child and build a firm foundation for subsequent literacy learning (Clay, 2001; 2016). Keeping in mind the complex neural networks proficient readers form, teachers first help novice readers actively construct primitive working systems using whatever existing systems they have upon entry to school, including their oral language, ways of constructing meaning, background experiences, and knowledge of print. As existing working systems become more effective, students extend their network of strategies for problem solving, with the goal being a self-extending system in which children continue to learn from their successful efforts at reading and writing (Clay, 2001).

Dyslexia

Elliott and Grigorenko (2014) discuss the controversies surrounding the varied definitions of dyslexia, whether based on discrepancies, causal explanations, or response to

intervention. The prevailing view of dyslexia is explained by the phonological deficit hypothesis, that is, deficits involving how words are sounded out. Dyslexia experts work from a simple theory of reading which stands in contrast to Clay's complex literacy processing theory. The Simple View of Reading (SVR) explains that reading comprehension, the ability to understand printed text, is determined by two cognitive abilities: decoding and language comprehension (Gough & Tunmer, 1986; Hoover & Gough, 1990; Hoover & Tunmer, 2018). The SVR is represented with the simple equation R=D x L, where decoding (D)and language comprehension (L) hold equal importance in terms of their contribution to reading comprehension (R). Although decoding is explained as the ability to recognize words in print, it is often more narrowly defined as alphabetic coding which relates letter sequences to the phonological structures underlying the word's pronunciation and thereby links to the reader's mental lexicon (Hoover & Tunmer, 2018).

The SVR is a static model, which describes reading at a single point in time, rather than a model of reading development (Hoover & Tunmer, 2018). As such, the SVR does not explain how reading develops over time only that one's level of attainment at any given point depends solely "on the multiplicative combination of the levels of the two components" (Hoover & Tunmer, 2018, p. 306).

The SVR holds that language comprehension and decoding are critical to reading success, but it does not make specific recommendations about instructional protocols (Hoover & Tunmer, 2018). Nevertheless, young children entering school typically have developed their language comprehension (L) to a greater degree than their word recognition (D). Therefore, it stands to reason that in order to increase reading proficiency, decoding skills (D) must be strengthened.

Moats and Tolmab (2019) cite instructional recommendations that support the Simple View of Reading and are often applied to students with dyslexia. These recommendations are based on ten foundational ideas:

1. The Simple View of Reading (Gough & Tunmer, 1986; Hoover & Gough, 1990, Hoover & Tunmer, 2018)
2. Orthographic processing and the brain (Dehaene, 2009; 2013)
3. Ehri's phases of word reading development (Ehri, 1996; 2014; Ehri & Snowling, 2004)
4. Letter-by-letter processing and orthographic mapping (Raynor et al., 2001)
5. Four-part processing model for word recognition (Seidenberg, 2013; Seidenberg & McClelland, 1989)
6. Scarborough's Reading Rope (Scarborough, 2001)
7. Phonological and phonemic awareness (Kilpatrick, 2015)
8. Vocabulary's importance to reading comprehension (Adolf & Perfetti, 2014; Stahl & Nagy, 2006)
9. Lexical quality of stored word image (Perfetti, 2007)
10. Mental model of the text (Johnson-Laird, 1983) or situation model (Kintsch, 1998)

Many states are requiring school districts to provide explicit instruction in the Science of Reading skills, and to provide professional development for teachers in the Science of *Teaching* Reading. In 19 states, students in colleges of teacher education are required to take a state test specifically on the Science of Reading skills and concepts related to teaching reading (Gewirtz, 2020).

Dyslexia interventions are commonly based on the Orton-Gillingham (OG) approach, which is a multisensory approach grounded in synthetic phonics (Ritchey & Goeke, 2006). Two meta analyses offer similar conclusions about the effectiveness of OG methods, which are consistent with the What Works Clearinghouse report. Ritchey and Goeke reviewed published research on OG, and found that some studies showed the intervention resulted in significantly higher performance post-intervention; however, they concluded that the available research on OG is "currently inadequate, both in number of studies and in the quality of the research methodology, to support that OG interventions are scientifically based" (Ritchey & Goeke, 2006, p. 182). A more recent meta-analysis studied the effects of OG interventions on reading outcomes for students identified with word-level reading disabilities (WLRD) or at risk for WLRD (Stevens et al., 2021). Unlike Ritchey and Goeke's study, this newer study excluded research with college participants and OG instruction delivered in whole-class settings. Researchers found no statistically significant improvement to foundational skills, vocabulary, or reading comprehension outcomes for students involved in OG interventions. In addition, they called for more high-quality research with larger sample sizes that might provide the evidence necessary to inform both practice and policy.

Nevertheless, currently 28 states have adopted policies requiring specific intervention approaches; 10 require "multisensory" approaches; 17 require "evidence based" approaches, and 10 stipulate "explicit" or "direct instruction" (National Center on Improving Literacy, 2021).

Brief Description of Relevant Research

The effects of Reading Recovery have been examined in high-quality experimental studies, quasi-experimental studies, as well as numerous qualitative studies (Watson & Askew, 2009). *The Observation Survey of Early Literacy Achievement* (Clay, 2013), used as one of the

assessments in this study, was reviewed by the National Center on Intensive Intervention (NCII) and found to have the highest possible ratings for classification accuracy as a screening tool and technical standards. The U. S. Department of Education Institute of Education Sciences (2013), What Works Clearinghouse (WWC) found Reading Recovery to have positive effects on general reading achievement and potentially positive effects on alphabetics, reading fluency, and comprehension for beginning readers. A recent four-year, mixed methods, external evaluation, conducted by the Consortium for Policy Research in Education (CPRE) at the University of Pennsylvania, and funded through the Investing in Innovation (i3) grant from the U.S. Department of Education's Office of Innovation and Improvement, found Reading Recovery to be highly successful in outcomes for students, implementation fidelity, and instructional strength in the teachers who delivered the intervention lessons (May et al., 2016). This i3 study was one of the most comprehensive studies ever implemented in the field of education.

The What Works Clearinghouse has reviewed research related to several programs frequently associated with instruction for students with dyslexia; however, the Students with a Specific Learning Disability review protocol requires research studies that often are more rigorous than those submitted, thus resulting in inconclusive findings. As of July 2010, the U.S. Department of Education WWC (2010) found no studies of "Unbranded Orton-Gillingham-based Interventions" that fell within the scope of the review protocol and met WWC's evidence standards. Similarly, the WWC reviewed four studies of the Dyslexia Training Program, from Texas Scottish Rite Hospital; however, none of those studies met the evidence standards for the review protocol for students with Learning Disabilities.

The U. S. Department of Education WWC (2008, 2009) also reviewed a professional development intervention based on Language Essentials for Teachers of Reading and Spelling

(LETRS), a professional development curriculum that embraces the Simple View of Reading (Garet et al., 2008). Teachers participated in one of three groups: standard district professional development, eight days of training from the LETRS curriculum, or eight days of training from the LETRS curriculum along with weekly, one-to-one support from a trained instructional coach. Teachers in both LETRS groups showed increased knowledge of reading instructional techniques and use of explicit instruction; however, students' reading test scores did not increase under either condition.

Methods

This study employed two designs: (1) a correlational design to determine concurrent validity among the *Observation Survey of Early Literacy Achievement (OS), the Feifer Assessment of Reading (FAR),* and *the Slosson Oral Reading Test-Revised,* and (2) an ex post facto, quasi-experimental design involving a pretest and posttest for all students with three assessments. The purpose of this pilot study was to determine if first grade students who have been identified at-risk for dyslexia can be distinguished from students who have initial reading and writing difficulties but do not have dyslexia characteristics.

Setting

The study took place in two public school districts in a southern state that implemented Reading Recovery as the first intervention for first grade students. See Table 1. In the first district considered remote by the state department of education, three elementary schools participated. They comprised 2.7%, 0.7%, and 2.0% African American students; 64%, 63%, and 85% Hispanic students; and 20%, 31%, and 7.4% White students respectively as the three largest

populations in the schools. Educationally disadvantaged percentages for each school in District 1 were 69%, 43%, and 83% respectively.

Five schools participated in the study from District 2. They are comprised of 9.4%, 7%, 11%, 30%, and 9% African American students; 30%, 18%, 55%, 20%, and 31% Hispanic students; and 53%, 67%, 30%, 41% and 58% White students, respectively, attending as the three largest ethnic groups. The schools ranked as 38%, 19%, 76%, 62%, and 42% economically disadvantaged respectively.

Participants

In the United States, Reading Recovery teacher leaders are affiliated and supported by a university training center (UTC). A UTC trains teacher leaders in the theory, practice, and implementation of Reading Recovery; provides site visits and ongoing professional development; and supports state and national data collection. Teacher leaders provide Reading Recovery teacher clinical training at the district or consortium level. The two districts that participated in this study were both affiliated with one UTC.

When selecting the sample, researchers first looked at a pool of districts within the UTC network whose Reading Recovery outcomes were equivalent to or higher than national outcomes. This ensured that the teaching and implementation were likely to be done with fidelity. From that pool, two districts agreed to participate as a purposeful sample. Research permission was secured from the district administrators and school principals to recruit Reading Recovery teachers and first grade students.

The 10 teachers in this study were trained Reading Recovery teachers. For their initial training, the teachers completed 6 hours of graduate coursework taught by a certified Teacher

Leader employed by a school district and affiliated with a Reading RecoveryUniversity Training Center. Throughout the training year, teachers were required to teach two rounds of 4 children, individually and daily, while participating in weekly clinical and peer-critiquing experiences. The teachers learned to observe closely and describe student and teaching behaviors as well as analyze children's reading behaviors and relate those behaviors to more general theories of literacy. The teachers-in-training also learned to design a series of lessons tailored to the specific needs of an individual child and to make effective moment-by-moment instructional decisions. After the training year, the teachers continued to participate in 6-8 required professional development sessions annually to refine and further develop their skills to effectively teach children who are 'at risk' of failing to learn how to read.

Students in a Reading Recovery intervention are selected based on the scores on the *Observation Survey of Early Literacy Achievement* (Clay, 2013) taken by a first grade school cohort. First-grade students scoring lowest on the survey receive the intervention first. Reading Recovery trained teachers each teach two rounds of 4 students in individual lessons, one round in the fall (12-20 weeks) and one round in the spring (12-20 weeks). Teachers participating in the study secured permission from parents of the 36 students in the first round of their tutoring sessions; however, due to students moving and variation in testing, the sample size varied. Students tested similarly on all pretest measures in both school districts including the *Observation Survey of Early Literacy Achievement*, the *Feifer Reading Assessment* (Feifer & Nader, 2016), and the *Slosson Oral Reading Test-Revised* (Slosson & Nicholson, 2002).

Screening for dyslexia characteristics for kindergarten and first grade students is mandated in the state in which this study took place. The districts followed state guidelines to select their screening instruments. Fifty percent of students selected for Reading Recovery as the

lowest achieving literacy learners in first grade were also found to have dyslexia characteristics based on the district screening instruments. For the purposes of this study, we refer to two groups: students *with* dyslexia characteristics and students *without* dyslexia characteristics, based upon the district screening. One student's parents opted out of the district's dyslexia screening, one student moved during the study, and one school did not provide dyslexia screening information.

This study took place during the first semester of school. Within two weeks of school starting in the fall, students in 1st grade cohorts within the participating schools were assessed with *An Observation Survey of Early Literacy Achievement* (OS; Clay, 2013). Once the students were selected for the intervention, the Reading Recovery teachers also administered the Slosson Oral Reading Test-Revised (SORT-R; Slosson & Nicholson, 2002). Additionally, within one week of selection, students were assessed with the *Feifer Assessment of Reading* (FAR; Feifer & Nader, 2016) by either graduate students or researchers. Graduate students were not part of the study, but they assisted in the pre/post assessment with the FAR in one district. Two of the researchers pre/post-tested students in the second district. Graduate students and researchers participated in the online training videos to learn to give the FAR with fidelity.

Intervention

Students received daily, 30 minute, one-to-one lessons with a teacher trained in Reading Recovery supplemental to their classroom reading instruction. Students in the intervention were tested with the OS, SORT, and the FAR within a one-week window again at the end of 20 weeks. The post-test OS and SORT were given by a teacher trained in Reading Recovery who was not the intervention teacher for the child to reduce teacher bias.

Assessment Instruments

The Observation Survey of Early Literacy Achievement (OS; Clay, 2013) is an individually administered, standardized, assessment comprising six tasks. Letter Identification measures letter knowledge, Word Reading measures automatic word recognition, Hearing and Recording Sounds in Words measures phonological awareness and how it is linked to letter knowledge, Writing Vocabulary measures how many words a child can write independently in 10 minutes, and the Text Reading Level measures oral reading competency.

The OS received the highest possible ratings by the National Center for Response to Intervention (NCRTI) on all five of NCRTI's technical standards and can be used by "school psychologists, special educators and others as an evidence-based screening instrument to identify children at risk for literacy failure" (D'Agostino, 2012, p. 53). The OS is appropriate for K-2 emergent readers in the classroom and is used by teachers trained in Reading Recovery for student selection and to make decisions about progress at the end of a student's series of lessons. It takes approximately 30 minutes to give.

The *Slosson Oral Reading Test-Revised* (SORT-R; Slosson & Nicholson, 2002) is intended to be used as a screening instrument with preschool to adult persons to determine a reading level. The SORT-R comprises 200 words in groups of 20 to determine a grade level. Internal reliability for the SORT-R yielded coefficients above .95% and the highest reliability was found in the youngest age groups. This instrument is often used by the research community as an additional measure of reading achievement because it is reliable, quick, and easy to use. It is used by Reading Recovery as an external, standardized measure of reading achievement and is given at the beginning and end of a series of lessons.

The *Feifer Assessment of Reading* (FAR; Feifer & Nader, 2016) is a standardized, comprehensive assessment that has been validated on preschool to college age students. It comprises fifteen subtests measuring different aspects of vocabulary, phonemic awareness, decoding skills, rapid naming, orthographic processing, morphological processing, word memory, reading fluency (word and story; silent and oral), and comprehension skills. This assessment is recommended for the general screening of dyslexia among other purposes and takes approximately one hour to administer.

The FAR was chosen for the study because of its reported ability to measure specific subtypes of dyslexia for first grade students. The FAR consists of four subindicies: the phonological index, the fluency index, the mixed index, and the comprehension index. For first graders, some of the subtests in each index were excluded as per the testing materials. The phonological index included tests on phonemic awareness, isolated word reading fluency, oral reading fluency, and positioning sounds; it excluded the nonsense word decoding test. The fluency index included tests on rapid automatic naming, verbal fluency, visual perception, and orthographical processing; it excluded the irregular word reading fluency test. The mixed index combined the phonological index and the fluency index. The comprehension index included tests on semantic concepts, word recall, print knowledge (only for pre-Kindergarten through first grade students); it excluded tests on morphological processing and silent reading fluency: comprehension. As noted above, there are specific instances in which the FAR is a markedly different test for students in first grade and students in second grade and above.

Statistical Analyses

Each research question will be addressed separately; however, in order to decrease the risk of overall Type I error, a Bonferroni correction will be used to adjust our critical *p*-value of $p < .05$ to $p < .01$. This helps to decrease the chance of finding false positive results. Numerous studies use correlations to indicate concurrent validity for the OS, so a Pearson product correlation was used to indicate concurrent validity for all measures (D'Agostino, 2012; Gómez-Bellengé et al., 2005; Holliman et al., 2010). To indicate any significant differences among the OS and the FAR subtests, scores were converted to z-scores and compared using repeated measures ANOVA. In order to identify a more prevalent FAR subtype, a chi-square analysis was used to see if our results differed from expected results. Lastly, we used an ANOVA to indicate differences among students with and without dyslexia characteristics in regard to the OS, FAR, and SORT-R and a repeated measures ANOVA to see their individual growth. Further, we compared our sample's effect sizes to the effect sizes of a national random sample and a UTC random sample each of which was calculated by the International Data Evaluation Center at the Ohio State University.

We conducted an ANOVA to ensure that the students in the different school districts and dyslexia status grouping were the same at the onset of the intervention. Maintaining a Bonferroni correction to decrease our Type I error ($\alpha < 0.01$), we found that students in the two school districts were similar on the pre-test for the FAR, $F(1, 31) = 2.589$, $p = .118$, the SORT-R, $F(1,34) = 6.923$, $p = .118$, and the OS, $F(1,34) = 7.861$, $p = .010$. Additionally, we found that students identified with dyslexia characteristics and students identified without dyslexia characteristics were similar on the pre-test for the FAR, $F(1, 31) = .062$, $p = .805$, the SORT-R, $F(1, 31) = .660$, $p = .423$, and the OS, $F(1, 28) = 1.755$, $p = .196$.

Results

The purpose of this study is to determine if first grade students who have been identified at-risk for dyslexia can be distinguished from other students who have initial reading and writing difficulties. In order to better understand this, we asked four research questions to determine (1) the concurrent validity between the *Feifer Assessment of Reading* (FAR; Feifer & Nader, 2016), the *Observation Survey of Early Literacy Achievement* (OS; Clay, 2016), and the *Slosson Oral Reading Test-Revised* (SORT-R; Slosson & Nicholson, 2002), (2) whether there was a significant difference in OS scores and each of the four FAR indices at pre-intervention, (3) the extent to which the FAR categories are present among children selected for Reading Recovery, and (4) the effectiveness of the Reading Recovery intervention for students both with and without dyslexia tendencies.

Concurrent Validity

Our first research question is what is the relationship of the criterion scores used for Reading Recovery intervention by the OS and the SORT-R and how does that compare to the Total Score from the FAR. In order to answer this question, we conducted Pearson product correlation coefficients. Even with a small sample size and after analyzing a scatterplot of the data, the assumptions of linearity, homoscedasticity, and normality for a Pearson product correlation were assumed once missing data were removed due to students moving. Confirming previous studies of concurrent validity (D'Agostino, 2012), the relationship between the OS and the SORT-R was strong and positive ($r = .874$, $p < .001$). While the relationship between the FAR and SORT-R ($r = .701$, $p < .001$) did not meet our standard for confirming concurrent validity ($r > .80$), we did find that the relationship between the FAR and OS was strong and positive ($r = .807$, $p < .001$), which confirmed concurrent validity (See Table 2).

Understanding Dyslexia Identification

Our second and third research questions dealt with the identification of students with dyslexia characteristics who were in Reading Recovery. Our second research question asked if there was a significant difference in the mean OS criterion scores and each of the four FAR indices at pre-intervention. The four FAR indices indicate a FAR subtype of dyslexia: dysphonetic dyslexia, surface dyslexia, mixed dyslexia, or reading comprehension deficits. The FAR measures dysphonetic dyslexia through the phonological index (FAR-PI), surface dyslexia through the fluency index (FAR-FI), mixed dyslexia through the mixed index (FAR-MI), and reading comprehension deficits through the comprehension index (FAR-CI). Any measure that has a standard score less than 85 indicates that particular form of dyslexia is present.

Our third research question asks whether there is a more prevalent FAR subtype, dysphonetic dyslexia, surface dyslexia, mixed dyslexia, or reading comprehension deficits, identified among the students who were in Reading Recovery. All FAR subtypes indicated a significant difference; however, due to the mathematical model for the chi-square test, no statistic can be given for the presence of mixed dyslexia due to all subjects having that quality (the chi-square test assumes that there will be members in each category). There was a significant difference for the presence of dysphonetic dyslexia (χ^2 [1, n = 33] = 25.485, $p < .001$) with more students with dyslexia characteristics ($n = 31$) than not ($n = 2$). There was also a significant difference for the presence of surface dyslexia (χ^2 [1, n = 33] = 22.091, $p < .001$) with more students with dyslexia characteristics ($n = 30$) than not ($n = 3$). Lastly, there was a significant difference for the presence of students with reading comprehension deficits (χ^2 [1, n = 33] = 13.364, $p < .001$) with more students with dyslexia characteristics ($n = 27$) than not ($n = 6$). These data suggest that, according to the FAR, our sample had a disproportionate number of students with dyslexia characteristics (See Table 3).

Because the FAR indicated that there was a high incidence of students with dyslexia characteristics, a statistic to identify differences was needed to understand how these identified forms of dyslexia may interact. To ascertain if there was a significant difference among the OS and each of the four FAR indices, scores were converted to z-scores. Missing scores were excluded from data analysis. Once all parameters were verified, a repeated measure ANOVA was conducted to see if there was a difference on each of the five test scores (OS, FAR-PI, FAR-FI, FAR-MI, and FAR-CI). No significant differences were found when comparing tests, F (1,27) = .053, p = .815, η_p^2 = .002 or the intersection between test and dyslexia status, F (1,27) = .140, p = .711, η_p^2 = .005. These data indicate that students scored similarly on the OS and each FAR subtest at the beginning of the intervention.

Effectiveness of Reading Recovery

Our fourth research question pertains to the effectiveness of the Reading Recovery Intervention as assessed by changes to pre- and post-intervention scores on the OS, SORT, and FAR. All data were deemed normal and of equal variance once accounting for students without scores due to moving. While there were no significant differences in the sample population in regard to school district or school-identified dyslexia status pre-intervention, we did notice a significant difference post-intervention both with the whole group and individual groupings by dyslexia status.

A repeated measures ANOVA was used to show the difference between pretest and posttest scores. For the FAR, there was a significant difference between pretest and posttest scores, F(1,25) = 222.199, p <.001, η_p^2 = .899, and no significant interaction between time and dyslexia status, F(1,25) = 2.132, p = .157, η_p^2 = .079. These students showed very large gains on their FAR scores with very large effect sizes except for the interaction between time and dyslexia

status. For the OS, there was a significant difference between the pretest and posttest scores, $F(1,29) = 507.654$, $p < .001$, $\eta_p^2 = .946$, and a significant interaction between time and dyslexia status, $F(1,29) = 12.078$, $p = .002$, $\eta_p^2 = .294$. These students showed very large gains on their OS scores with very large effect sizes; however, there was a difference between students identified with dyslexia characteristics and students without identified dyslexia characteristics. On the SORT-R, there was a significant difference between pretest and posttest scores, $F(1,29) = 173.888$, $p < .001$, $\eta_p^2 = .857$, and a significant interaction between time and dyslexia status, $F(1,29) = 10.914$, $p = .003$, $\eta_p^2 = .273$. These students showed very large gains on their SORT scores with very large effect sizes, but there was a difference between students identified with dyslexia characteristics and students without identified dyslexia characteristics. All students saw gains, but because of the interactions detected with the OS and the SORT, additional analyses are needed to understand how dyslexia characteristics interacted with these scores.

In regard to the OS, students without dyslexia characteristics scored higher ($M = 515.46$, $SD = 16.974$) than students with dyslexia characteristics ($M = 473.93$, $SD = 34.132$), $F(1,25) = 15.627$, $p < .001$. In regard to the SORT-R, students without dyslexia characteristics scored higher ($M = 33.08$, $SD = 12.757$) than students with dyslexia characteristics ($M = 19.64$, $SD = 10.853$), $F(1,25) = 8.729$, $p < .01$. In regard to the FAR, no statistically significant difference was detected between students without dyslexia characteristics ($M = 1014.92$, $SD = 64.994$) and students with dyslexia characteristics ($M = 964.79$, $SD = 49.436$), $F(1,25) = 5.137$, $p = .032$.

While these findings might be interpreted to indicate that the intervention is a powerful tool for students without dyslexia characteristics, comparing our sample to a national sample and a regional sample from a university training center indicates that Reading Recovery may have had a positive impact on both groups in regard to the OS. When we look at the effect sizes of

each group, they all indicate extremely large differences (Cohen, 1992). See Table 4. Even though our sample's effect size (d = 3.94) compares to the national sample (d = 1.48) and the university training center's sample (d = 1.38), it should be noted that our sample (M = 360.35, SD = 34.641) started with lower scores and less variability than the national sample (M = 436.4, SD = 53.7) and the university training center sample (M = 426, SD = 57.4). These findings indicate that the Reading Recovery intervention was effective for both students with and without dyslexia characteristics; however, students without dyslexia characteristics saw greater gains, which merits further study in how Reading Recovery compares to other interventions for students with dyslexia characteristics.

Limitations

The small number of students in this pilot study and brief duration of the study are clear limitations to the generalizability of the study. Originally planned as a yearlong study, researchers intended to follow students through the end of their first grade year. Due to COVID-19 and the resulting school closures, end-of-year scores were not available. Therefore, no follow-up data are available about students' progress in the months following their interventions. Nationally, Reading Recovery students who complete a full series of lessons in the first half of the school year and continue with good classroom instruction continue to make progress with their grade cohorts (International Data Evaluation Center, 2019). A larger replication study that follows students throughout the entire school year is warranted.

The measurement tools also have limitations. The *Observation Survey* has ceiling effects on several of the tasks, such as Letter Identification, which contains a finite set of items that most children master by the end of first grade. These ceiling effects prevent differentiation of students scoring at the top end of the scale, which complicates efforts to measure literacy growth. On the

other hand, the *Feifer Assessment of Reading* has floor effects. At the beginning of first grade, it identified all students as having dyslexic characteristics, but for the most part did not distinguish the different types of dyslexia.

Discussion

Based on the findings from this study, the criterion scores on *The Observation Survey of Early Literacy Achievement* (OS; Clay, 2013) and the *Slosson Oral Reading Test- Revised* (SORT-R; Slosson & Nicholson, 2002) showed concurrent validity. Similarly, both the OS and SORT-R showed concurrent validity with the total scores on the *Feifer Assessment of Reading* (FAR; Feifer & Nader, 2016). There also were no significant differences in the mean OS criterion scores and each of the FAR indices at pre-test. These scores indicated statistical agreement among the three assessments used in this study at pre-test. Based on these statistical analyses, we were able to show that all of the students in the study were similar pre-intervention.

At the beginning of the school year all first grade students were screened for dyslexia characteristics by designated personnel in their districts. Students in the study were placed in Reading Recovery as their first intervention because they were the lowest performing in literacy in their first grade cohorts. The district screening scores indicated that approximately 50% of the students showed dyslexia characteristics and 50% did not, yet were still the lowest performing in their classrooms. However, all children in the study showed dyslexia characteristics according to the FAR indices. From the beginning of the year to midyear, the number of students in this study who demonstrated difficulty according to the FAR subtypes/indices dropped by more than 50%. This finding held true for every subtype. (For individual student's scores, see Appendix A.) All of these findings highlight the challenge of clearly defining dyslexia, especially at the level of emergent reading. As stated earlier, schools in this state were required to screen K-1 students for

dyslexia characteristics, but they were not provided guidance about which screening assessments to use because a valid assessment has not been determined. According to the International Literacy Association (2016), "[s]ome have advocated for an assessment process that determines who should and should not be classified as dyslexic, but this process has been shown to be highly variable across states and districts in the United States, [and]of questionable validity" (p.3.).

Assessment instruments used for this study warrant further discussion. Although the OS has ceiling effects on some tasks, it seems that some of the FAR tasks may have floor effects, with many first graders in this study finding some subtests quite challenging at the beginning of the year (i.e., the concept of *synonyms* and *antonyms* on the Semantic Concepts subtest). Moreover, every child in the study tested with the FAR showed some dyslexia characteristics. The FAR may be a less sensitive measure for beginning first graders and, therefore, not as reliable for identifying students who will have significant difficulty learning to read when administered at the beginning of their schooling. Because the FAR subtests change after first grade, it is not clear how this impacts dyslexia identification beyond first grade. The FAR assessment for second graders includes different subtests such as the addition of tests on nonsense word reading, irregular word reading, morphological processing, and silent reading fluency: comprehension while dropping the test in print knowledge, which indicates a slightly different test being administered and thus a different conception of each type of dyslexia at this stage in development. Following students into second grade would warrant additional studies on the impact of Reading Recovery on students with dyslexia characteristics beyond the short term gains shown.

Reading Recovery as an intervention showed strong positive outcomes for children in this study who began the year as the lowest-achieving readers, regardless of whether they were

identified as having dyslexia characteristics. The Reading Recovery intervention served to reduce the incidence of false positives after 20 weeks of instruction. Reading Recovery has served as a response to intervention for the lowest achieving first grade students regardless of their language abilities, social economic status, assigned labels, or perceived deficits (Doyle, 2018). All students in the study made progress in the intervention based on significant pre-post effect sizes; however, students without school-identified dyslexia characteristics made greater progress. The outcome of this study is consistent with more than three decades of research that demonstrates Reading Recovery's effectiveness with diverse populations (Watson & Askew, 2009), including language ability, social economic status, perceived deficits, or assigned labels such as dyslexia.

Implications for Educators

Dyslexia is a mercurial concept. There is no agreement on the definition of dyslexia (Elliott, & Grigorenko, 2014), and these data do not support distinguishing different types of dyslexia at the beginning of first grade. There is a need for sensitive literacy assessments in the early grades. Conventional tests for dyslexia may not distinguish young children with and without dyslexia. Teachers should be critical of tests used as screeners, particularly if they direct children to one-size-fits all instruction.

Children come into school with varied literacy experiences and skills. Emergent readers who are having difficulty should have an opportunity for a short-term, early intervention before they are identified for inclusion in a long-term dyslexia program. Highly skilled Reading Recovery teachers, who have been trained to observe closely and respond to individual needs in one-to-one instruction, can help most children make gains in reading.

In conclusion, current state literacy policies are well intentioned; however, our students are short-changed when mandated identification and intervention practices do not address the diverse, individualized needs of children. Becoming literate is the bedrock of all subsequent learning. In terms of positive student outcomes and dollars spent, effective and short-term early interventions can help many students become independent readers and writers, while also serving as a means of identifying students who will need more sustained support. More research about dyslexia identification and intervention is needed for policymakers to make informed decisions to equip schools with varied ways to effectively support students so that schools can teach *all* children to be literate.

References

Adolf, S. M. & Perfetti, C. A. (2014). Individual difference in word learning and reading ability. In C. A. Stone, E. R. Silliman, B. J. Ehren, & G. P. Wallach (Eds.). *Handbook of language and literacy: Development and disorders (*pp. 246–264). Guilford Press.

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

Clay, M. M. (2013). *An observation survey of early literacy achievement*. Heinemann.

Clay, M. M. (2016). *Literacy Lessons designed for individuals*. Heinemann.

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*(1), 155–159. https://doi.org/10.1037/0033-2909.112.1.155

D'Agostino, J. V. (2012). Technical review committee confirms highest NCRTI rating for *Observation Survey of Early Literacy Achievement*. *The Journal of Reading Recovery, 11*(2), 53–56.

Davis Dyslexia Association International. (2021). State Dyslexia Laws — Dyslegia: A Legislative Information Site https://www.dyslegia.com/state-dyslexia-laws/

Dehaene, S. (2009). *Reading in the brain: The new science of how we read.* Penguin Viking.

Dehaene, S. (2013). Inside the letterbox: How literacy transforms the human brain. *Cerebrum: The Dana Forum on Brain Science, 7.* https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3704307/

Doyle, M. A. (2009). A dynamic future. In B. Watson & B. Askew (Eds.), Boundless horizons: Marie Clay's search for the possible in children's literacy (pp. 287–306). Heinemann.

Doyle, M. A. (2018). Communicating the power of Reading Recovery and Literacy Lessons instruction for dyslexic learners: An ethical response, *The Journal of Reading Recovery.* Columbus, OH, Reading Recovery Council of North America.

Doyle, M. A. (2019). Marie M. Clay's theoretical perspective: A literacy processing theory. In D. Alvermann, N. J. Unrau, M. Sailors, & R. Ruddell (Eds.), *Theoretical models and processes of literacy* (7th Ed., pp. 84–100). Routledge.

Ehri, L. C. (1996). Development of the ability to read words. In R. Barr, M. Kamil, P. B. Mosenthal, & P. D. Pearson (Eds.). *Handbook of reading research* (Vol. 2, pp. 383–418). Lawrence Erlbaum.

Ehri, L. C. (2014). Orthographic mapping in the acquisition of sight word reading, spelling memory, and vocabulary learning. *Scientific Studies of Reading, 18(1),* 5–21. https://doi.org/10.1080/10888438.2013.819356

Ehri, L. & Snowling, M. (2004). Development variation in word recognition. In A. C. Stone, E. R. Silliman, B. J. Ehren & K. Apel (Eds.). *Handbook of language and literacy: Developmental disorders* (pp. 443–460). Guilford Press.

Elliott, J. G. & Grigorenko, E. L. (2014). *The dyslexia debate.* Cambridge University Press.

Feifer, S., & Nader, R. G. (2016). *Feifer assessment of reading*. Academic therapy publications.

Garet, M. S., Cronen, S., Eaton, M., Kurki, A., Ludwig, M., Jones, W., Uekawa, K., Falk, A., Bloom, H. S., Doolittle, F., Zhu, P., & Sztejnberg, L. (2008). *The impact of two professional development interventions on early reading instruction and achievement*. NCEE 2008-4030. National Center for Education Evaluation and Regional Assistance. https://files.eric.ed.gov/fulltext/ED502700.pdf

Gerwitz, K. (2020). Reading instruction: A flurry of new state laws. *Education Week.* https://www.edweek.org/teaching-learning/reading-instruction-a-flurry-of-new-state-laws/2020/02

Gómez-Bellengé, F. X., Rodgers, E., Schulz, M., & Wang, C. (2005). Examination of the validity of the Observation Survey with a comparison to the ITBS. Paper presented at the Annual Meeting of the American Educational Research Association, Montreal, Quebec, Canada.

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

Holliman, A. J., Hurry, J., & Douëtil, J. (2010). Standardisation of the Observation Survey in England and Wales, UK. Institute of Education.

Hoover, W. A., & Gough, P. B. (1990). The simple view of reading. *Reading and writing, 2*(2), 127–160. https://doi.org/10.1007/BF00401799

Hoover, W. A., & Tunmer, W. E. (2018). The Simple View of Reading: Three Assessments of Its Adequacy. *Remedial & Special Education*, *39*(5), 304–312. https://doi.org/10.1177/0741932518773154

International Data Evaluation Center (2019). *2018-2019 Reading Recovery National Summary Report for the United States.* The Ohio State University.

International Literacy Association. (2016). Dyslexia [Research advisory]. Newark, DE.

Johnson-Laird, P. (1983). *Mental models: Toward a cognitive science of language, inference, and consciousness.* Harvard University Press.

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

Kintsch, W. (1998). *Comprehension: A paradigm for cognition.* Cambridge University Press.

Kirby, J., & Savage, R. (2008). Can the simple view deal with the complexities of reading? *Literacy*, *42*(2), 75–82. https://doi.org/10.1111/j.1741-4369.2008.00487.x

May, H., Sirinides, P., Gray, A., & Goldsworthy, H. (2016). *Reading Recovery: An evaluation of the four-year i3 scale-up*. Consortium for Policy Research in Education. https://www.cpre.org/sites/default/files/reading_recovery_final_report.pdf

Moats, L. C., & Tolman, C. A. (2019). *LETRS* (3rd ed., Vol. 1). Voyager-Sopris Learning, Inc.

National Center on Improving Literacy. (2021). *State of Dyslexia*. https://improvingliteracy.org/state-of-dyslexia

Perfetti, C. (2007). Reading ability: Lexicon quality to comprehension. *Scientific Studies of Reading,* 11(4), 357–383. https://doi.org/10.1080/10888430701530730

Pinnell, G. S., DeFord, D. E., & Lyons, C. A. (1988). Reading Recovery: Early intervention for at-risk first graders (Educational Research Service Monograph). Educational Research Service.

Raynor, K., Foorman, B. F., Perfetti, C. A., Pesetsky, D., & Seidenberg, M. S. (2001). How psychological science informs the teaching of reading. *Psychological Science in the Public Interest,* 2, 31–73. https://doi.org/10.1111/1529-1006.00004

Reading Recovery Council of North America. (2021). Research & Evaluation. https://readingrecovery.org/reading-recovery/research-evaluation/

Ritchey, K. D., & Goeke, J. L. (2006). Orton-Gillingham and Orton-Gillingham—based reading instruction: A review of the literature. *The Journal of Special Education*, *40*(3), 171–183. https://doi.org/10.1177/00224669060400030501

Rogoff, B. (1990). *Apprenticeship in thinking: Cognitive development in social context*. Oxford University Press.

Scarborough, H. (2001). Connecting early language and literacy to later reading (dis)abilities: Evidence, theory and practice. In S. B. Neuman & D. K. Dickinson (Eds.). *Handbook of early literacy research* (Vol. 1, pp. 97–110). Guilford Press.

Schwartz, R. M. (2005). Literacy learning of at-risk first-grade students in the Reading Recovery early intervention. *Journal of Educational Psychology,* 97(2), 257–267. https://doi.org/10.1037/0022-0663.97.2.257

Seidenberg, M. S. (2013). The science of reading, and its educational implications. *Language Learning and Development*, 9(4), 331–360. https://doi.org/10.1080/15475441.2013.812017

Seidenberg, M. S., & McClelland, J. L. (1989). A distributed, developmental model of word recognition and naming. *Psychological Review,* 96(4), 523–568. https://doi.org/10.1037/0033-295X.96.4.523

Slosson, R. L., & Nicholson, C. (2002). *Slosson oral reading test-revised*. Slosson Educational Publications.

Stahl, S. A., & Nagy, W. E. (2006). *Teaching word meanings.* Lawrence Erlbaum.

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

U. S. Department of Education Institute of Education Sciences. (2008). *What works clearinghouse: WWC quick review of the report "The impact of two professional development interventions on early reading instruction and achievement."* https://files.eric.ed.gov/fulltext/ED503728.pdf

U.S. Department of Education Institute of Education Sciences. (2009). *What works clearinghouse: WWC quick review of the report "The impact of two professional development interventions on early reading instruction and achievement" (Revised)*. https://ies.ed.gov/ncee/wwc/Docs/QuickReview/pdr_rev0309.pdf

U. S. Department of Education Institute of Education Sciences. (2010). *What works clearinghouse intervention report: Orton-Gillingham-based strategies (Unbranded)*. https://ies.ed.gov/ncee/wwc/Docs/InterventionReports/wwc_ortongill_070110.pdf

U.S. Department of Education Institute of Education Sciences. (2013). *What works clearinghouse intervention report: Reading Recovery.* https://ies.ed.gov/ncee/wwc/Docs/InterventionReports/wwc_readrecovery_071613.pdf

vanDerHeyden, A. M., & Burns, M. K. (2017). Four Dyslexia Screening Myths That Cause More Harm Than Good in Preventing Reading Failure and What You Can Do Instead. *Communique*, 45(7), 25–28.

Watson, B., & Askew, B. (2009). *Boundless horizons: Marie Clay's search for the possible in children's literacy.* Heinemann.