Universal Assessment of Early Reading Ability in Title I Schools

A Dissertation submitted
to the Graduate School
Valdosta State University

in partial fulfillment of requirements for the degree of

DOCTOR OF SPEECH-LANGUAGE PATHOLOGY

in the Department of Communication Sciences and Disorders
of the Dewar College of Education and Human Services

November 2023

Laura Johnson Abercrombie

M.Ed., Valdosta State University, 2016
B.S.Ed., Valdosta State University, 2015

This dissertation, "Universal Assessment of Early Reading Ability in Title I Schools," by Laura Johnson Abercrombie, is approved by:

Dissertation Committee Chair

DocuSigned by:
Katherine Lamb
C6B89C2F426F400...

Katherine Lamb, Ph.D.
Assistant Professor of Communication Sciences and Disorders

Dissertation Research Member

DocuSigned by:
Matthew Carter
E9BB2060FBE042C...

Matthew Carter, Ph.D.
Professor of Communication Sciences and Disorders

Committee Member

DocuSigned by:
Barbara Radcliffe
2ADAA54865044BA...

Barbara Radcliffe, Ph.D.
Associate Professor of Teacher Education

Associate Provost for Graduate Studies and Research

Becky K. da Cruz

Becky K. da Cruz, Ph.D., J.D.
Professor of Criminal Justice

Defense Date

September 29, 2023

FAIR USE

DUPLICATION

I authorize the Head of Interlibrary Loan or the Head of Archives at the Odum Library at Valdosta State University to arrange for duplication of this thesis for educational or scholarly purposes when so requested by a library user. The duplication shall be at the user's expense.

Signature 

I refuse permission for this dissertation to be duplicated in whole or in part.

Signature _____

ABSTRACT

Identifying at-risk readers in early grades is an ongoing challenge for educators, especially those in Title I schools where environmental and economic disadvantages may negatively impact students' reading ability. The purpose of the current study was to compare methods of identifying at-risk readers in kindergarten through third grade in Title I schools. This study was conducted in two Title I elementary schools using Acadience Reading Diagnostic to measure reading achievement and growth in students with reading disabilities, students at risk of reading disabilities, and typically reading students.

Reading composite scores (RCS) were significantly different between the typical group and each of the other two groups (at-risk and identified groups) at all grade levels. Growth rate classifications were significant between groups only for kindergarten and first-grade participants. Among the three risk identification methods, there was fair agreement between low-achievement (LA) and low-growth (LG) methods and moderate agreement between dual-discrepancy (DD) and the other two methods. RCS means were found to be significantly lower than published norms. Growth rate classification means also fell significantly below published norms except for participants in first grade. Application of the LA and DD methods to local norms resulted in the identification of fewer than half the number of at-risk participants compared to published norms.

The dual discrepancy method was recommended to identify at-risk readers in these grades because both achievement scores and growth rates were significantly different among kindergarten and first-grade groups. Due to a large number of participants identified as at-risk based on the use of published norms and the lack of significant difference in growth rates among groups in second and third grade, it was recommended that the LA method based on local norms be used to identify at-risk readers in second and third grades.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF ABBREVIATIONS

DD	Dual discrepancy
LA	Low achievement
LG	Low growth
LSI	Letter-sound integration
MTSS	Multitiered system of supports
PA	Phonemic awareness
RAN	Rapid automatized naming
RCS	Reading composite score(s)
RTI	Response to intervention
SSC	Sound-symbol correspondence
WCPM	Words correct per minute
WPM	Words per minute

ACKNOWLEDGEMENTS

I would like to thank everyone who has contributed to the completion of my dissertation. This would not have been possible without the support and guidance of my dissertation chair, Dr. Katherine Lamb. I am also thankful to my committee members, Dr. Matthew Carter and Dr. Barbara Radcliffe, for taking the time to offer their expertise and feedback.

My parents have been a source of encouragement and support throughout this journey. I also want to thank my husband for his patience and understanding during my countless hours of researching and writing.

Many peers and colleagues have contributed in countless ways, big and small, and I am grateful for their support. Thank you all for going with me on this journey!

DEDICATION

This dissertation is dedicated to the memory of Joy Alexander, whose love of teaching inspired my love of learning.

Chapter I

INTRODUCTION

Reading ability is the product of many underlying foundational skills wherein weakness in a single skill may result in weak reading ability. A multitiered system of supports (MTSS) has been widely implemented in public schools to provide appropriate intervention to students with reading weaknesses. To provide appropriate intervention, students must be accurately identified.

Statement of the Problem

The current model for identifying at-risk readers in public schools is the multitiered system of supports, or MTSS, which involves three tiers of increasingly intensive intervention. The academic portion of MTSS is often called response to intervention, or RTI. Within this model, various assessment methods are used to measure and monitor student performance and growth to determine the appropriate tier of instruction. Students can move up or down the tiers depending on their responsiveness. This model has significantly improved the disproportionality of special education referrals (O'Connor et al., 2013).

However, "responsiveness" is difficult to define clearly. A first-grade student who increases their reading level from a beginning-of-kindergarten level to a mid-year first-grade level is clearly responsive to the given intervention. Alternatively, it is more difficult to determine whether a fourth-grade student who increases their reading level from beginning-of-year kindergarten to mid-year first grade is sufficiently responsive.

Formal diagnostic assessments are not administered to all students in public schools. However, reading screenings are administered to all students and may be one of the first tools

used to identify at-risk students. Although valuable, these screeners lack the scope, validity, and reliability to diagnose a reading disorder.

When using assessments to identify at-risk students, there is no consensus in the current literature on an appropriate cut score. Cut scores are predetermined scores that indicate the point at which performance is considered impaired. Cut scores vary significantly across studies: from the 5th to 25th percentile (Milburn et al., 2017; Milburn et al., 2018), the 16th percentile (McArthur & Castles, 2013; Papadopoulos et al., 2020), and the 25th percentile (Fedora, 2016).

Finally, the interpretation of these scores may be based on students' level of achievement and the amount of progress made over a period of time. Identifying at-risk readers based on achievement versus growth has limited agreement (Milburn et al., 2017). There is a lack of research on these identification methods specific to early elementary-age students in Title I schools.

Identifying at-risk readers is an ongoing challenge for educators, especially those in Title I schools where environmental and economic disadvantages may negatively impact students' reading ability. Lack of agreement in the methods used to identify at-risk readers results in a high degree of variability depending on the method of identification used. There is no agreed-upon method for interpreting data to identify at-risk readers within the RTI model.

Definitions of Terms

At-risk for a reading disability – For the purposes of this study, participants "at-risk" for a reading disability were those who received tier two or tier three reading intervention through RTI based on their classification at the end of the study year.

Dual discrepancy (DD) – DD included the combined criteria of low achievement (LA) and low growth (LG) (Milburn et al., 2017). Participants whose end-of-year reading composite scores

(RCS) fell at or below the sixteenth percentile and whose growth pathway was classified as "well below typical progress," which corresponds to the 20th percentile and below, met the criteria for DD.

Identified reading disability – For the purposes of this study, participants with "identified" reading disabilities included those with an individualized education program (IEP) who received services for reading based on their classification at the end of the study year.

Low achievement (LA) – In a previous study (Milburn et al., 2017), LA was defined as end-of-year reading scores on a single measure that fell below the specified percentile. For the purposes of this study, LA was defined as an end-of-year RCS at or below the sixteenth percentile.

Low growth (LG) – In a previous study (Milburn et al., 2017), LG was defined as less-than-expected growth based on a latent growth-curve model. For the purposes of this study, LG was defined as an Acadience reading pathway of progress of "well below typical progress," which corresponds to growth at or below the twentieth percentile.

No-risk classification – The no-risk classification in this study is reserved for participants who were not identified as at-risk readers using any of the three methods.

Pre-literacy skills – Pre-literacy skills refer to the early-developing, reading-related skills that provide a foundation for learning to decode. These skills include sound-symbol correspondence (SSC) and phonological awareness (Vehkavuori et al., 2021).

Sound-symbol correspondence (SSC) – SSC refers to an individual's knowledge and application of phonemes and their corresponding graphemes (Hulme et al., 2012). SSC may be assessed by presenting letters of the alphabet and prompting an individual to name the phonemes corresponding to each.

Typical readers – For the purposes of this study, typical readers were defined as participants who

received no reading-related intervention based on their classification at the end of the study year.

Word reading ability – Word reading ability includes a variety of skills, including phonemic decoding, sight word reading, real word reading, and nonsense word reading. Additionally, word reading ability may be measured via reading fluency, word reading accuracy, word lists, and connected text (Morris et al., 2013).

Purpose

The primary purpose of this study was to analyze various methods of interpreting reading assessment scores to identify struggling readers. Three methods of identification were used: low achievement (LA), low growth (LG), and dual discrepancy (DD) (Milburn et al., 2017). A secondary purpose of the study was to compare the use of published and local norms in determining risk status. In Title I schools, achievement and growth may not align with national averages.

This is important because educators must be able to quickly identify students who need intervention without undergoing comprehensive evaluations that are typically not available on a widespread scale in schools. Educators can more consistently identify students who are genuinely at-risk for reading disorders and provide appropriate interventions more quickly and more objectively if given specific benchmark scores and a method for interpreting those scores.

Research Questions

1. How do reading achievement scores and growth rates differ among kindergarten through third-grade students with identified reading and reading-related disabilities, students at risk of reading disabilities, and typical readers?
2. What is the level of agreement among three methods of classifying risk status (LA, LG, and DD) using published norms?

3. How does the use of local norms on a universal reading measure compare to the use of published norms in the identification of at-risk readers?

Research Hypotheses

1. It was hypothesized that reading achievement scores, but not reading-related growth, would be significantly different among participants with identified reading disabilities, at-risk for reading disabilities, and typical readers.
2. It was hypothesized that low to moderate agreement would exist among the three methods of identifying at-risk students (LA, LG, DD).
3. It was hypothesized that the use of local norms would result in a higher level of specificity, and the use of published norms would result in a higher level of sensitivity.

Significance of the Study

The results of this study have direct implications for the development of objective identification criteria. If either reading achievement scores or growth scores are significantly different among groups but not both, the significant scoring system would be more strongly associated with the current classification of at-risk readers in this population. Additionally, if reading scores in a Title I school are significantly different from those of the normative population, this would have implications for using local versus published norms.

Conceptual Framework

The first part of the conceptual framework is related to the methods used to identify at-risk readers. Milburn et al. (2017) analyzed the application of four different methods of identifying reading disabilities in preschool children. These included IQ-achievement discrepancy, LA, LG, and DD. IQ-achievement discrepancy requires a measure of IQ that is not widely assessed in RTI. Therefore, IQ-achievement discrepancy is not easily applied to all

students. However, universal screeners and assessments are widely used for RTI, providing an appropriate data source.

The other three methods can be applied to universal screening measures to determine whether students' performance indicates a possible reading disability. LA is generally defined as scores below a set benchmark, so students classified as LA are those who did not meet a predetermined benchmark. LG is generally defined as less-than-expected progress toward a goal in consideration of initial performance. The combination of these two methods results in a DD method. Students who meet the criteria for DD are at the most significant risk for a reading disorder because they are performing below benchmark and are not progressing as expected to catch up with peers.

There are costs and benefits to using each of these three methods. Achievement scores alone do not account for short- or long-term growth. Growth scores alone do not account for initial performance levels. The results of several studies suggest that growth may not be a reliable indicator of reading ability. In one study (Maki et al., 2020), reading growth rate based on progress monitoring from RTI was unrelated to eligibility for a specific learning disability in reading. In another study (Van Norman & Nelson, 2021), oral reading fluency growth was unrelated to variations in reading ability. This lack of significance may be due to a ceiling effect (Hendricks & Fuchs, 2020). Any growth measures should account for the initial performance level to avoid this effect. Hendricks and Fuchs (2020) also suggested that growth is more sensitive to changes in performance that is significantly below grade level. Therefore, growth rates may be more beneficial for students performing well below the average range.

The second part of the conceptual framework is related to the normative data source for the universal assessment measures. Using local versus published norms may benefit Title I

schools (Good & Kaminsky, 2020). Because the population of Title I schools is not representative of the national population, using normative data based on scores within Title I schools may result in greater specificity in identifying true reading disorders rather than weaknesses due to environmental variables. However, this could also exclude students from needed intervention. Synthesis of the three risk identification methods and using local versus national norms leads to several questions.

Summary of Methodology

Participants included emerging and early readers in kindergarten through third grade in two rural Title I K-5 elementary schools. Participants were classified as "identified" reading disability, "at-risk" for a reading disability, or typical readers. Acadience Reading Diagnostic (Good & Kaminsky, 2020) was administered at the beginning, middle, and end of the school year to assess participants' reading ability across several domains: pre-literacy skills, SSC, and word reading ability.

To answer research question one, a one-way analysis of variance (ANOVA) was used to determine if there was a significant difference in reading ability among groups. The Kruskal-Wallis test was used to determine if there was a significant difference in growth among groups. To answer research question two, participants' scores were classified as LA, LG, DD, or no deficit, and Cohen's kappa coefficient was calculated to determine the level of agreement. To answer research question three, single-sample *t*-tests were calculated to compare local reading achievement and growth with published norms. Based on local norms, participants were classified as LA, DD, or no risk.

Limitations

This study has several limitations. First, students placed in full-day resource classrooms were excluded from this study because they were excluded from testing by their respective schools. The exclusion of these students may have altered the results of this study. Regarding the assessment measures used, the rapid automatized naming (RAN) subtest of the Acadience Reading Diagnostic was not administered to participants despite this measure's value in identifying at-risk readers. Because this subtest was only recently available at the beginning of the study year, the school district chose not to include the RAN subtest at that time. Additionally, the accuracy of Acadience Reading in the prediction of reading disorders is limited (Good et al., 2019). For kindergarten through third grade, criterion-related validity of Acadience RCS compared to scores on the Group Reading Assessment and Diagnostic Evaluation (GRADE) (Williams, 2001) ranged from .40 to .80, depending on the grade and time of year. Predictive validity of beginning-of-year and middle-of-year RCS with later RCS ranged from .52 to .91. These ranges indicate moderate to strong correlations. Therefore, Acadience Reading scores should be interpreted with caution and used in conjunction with other measures of reading ability.

Chapter II

REVIEW OF THE LITERATURE

Domains of Reading Instruction

The National Reading Panel (2000) was formed to analyze and synthesize the large amounts of previously conducted reading research. It resulted in the identification of five components of reading instruction: phonemic awareness (PA), phonics, fluency, vocabulary, and comprehension.

PA refers to an individual's understanding and use of the phonemes in words (National Reading Panel, 2000). PA skills include segmenting, blending, adding, deleting, and changing sounds in spoken words. Strong PA skills are crucial to the ability to decode written words. While the phonemes in words correspond to specific letters, PA is specific to the sounds themselves. Therefore, PA tasks can be completed orally without written words.

Phonics instruction, however, refers to teaching sound-symbol correspondence (SSC) and is dependent on written letters and words. Phonics ranges from simple SSC to complex spelling patterns and multisyllabic word reading. When an individual can correctly identify each sound in a word based on the letters, they use PA skills to blend those sounds to form a word.

The ability to read single words is not sufficient to develop good readers. Individuals must be able to read words fluently and accurately. Fluency refers to the speed, accuracy, and prosody used to read a text. There are multiple measures of reading fluency. Regarding speed, a common measure is the number of words read correctly in one minute, typically from either a grade-level text or a text on the student's reading level. Accuracy is a percentage based on the

number of words read correctly and the total number of words attempted. Prosody is an essential component of fluency but is more difficult to measure. A rubric may be used to classify prosody with consideration for pitch and pausing but is rarely assessed in a standardized fashion.

Again, reading words quickly and correctly is not sufficient to develop good readers. Students must also understand the meanings of the words they are reading. This is why vocabulary instruction is essential. At the very least, students must be able to use context clues in the text to infer the definition of an unknown word.

Finally, comprehension refers to understanding a text, whether a sentence, passage, or book. Good comprehension is the result of strong PA, phonics, fluency, and vocabulary skills, in addition to background knowledge in the topic area of the text. Comprehension processes overlap significantly with learning processes. If students can successfully read and comprehend a text, they can learn from it.

Each of these five components of reading instruction cannot exist in isolation. They are all interconnected to produce overall reading ability. While strengths in one area can mediate the effects of weakness in another, strengthening all five areas is more likely to produce good readers.

Models of Reading Instruction

There is ongoing controversy over the most effective method of teaching reading, which is commonly called the "reading wars" (Pearson, 2004). This debate has resulted in abundant research on reading development in school-aged children and is based on the differences between top-down and bottom-up reading models (Amadi, 2019; Bergeron, 1990; Bowers, 2020; Brooks, 2023; Fletcher et al., 2021; Glavach & Pribyl, 2018; Manning & Kamii, 2000; Wyse & Bradbury, 2022). Top-down literacy instruction addresses language comprehension, background

knowledge, vocabulary, language structures, and literacy knowledge (Scarborough, 2001). In these models, the meaning of the text is emphasized rather than individual letters and words that must be decoded. This results in readers using their background knowledge and other information to predict the meaning of the text. In these models, word identification is dependent on the correct prediction of meaning.

An example of a top-down reading model is whole language reading instruction, in which the focus of instruction is a literacy-rich environment, including speaking, listening, reading, and writing, with an emphasis on the meaning of the text rather than decoding ability (Bergeron, 1990). The whole language approach utilizes three cueing systems to teach decoding within meaningful reading, writing, listening, and speaking activities: graphophonemic cues, semantic/meaning cues, and syntactic/language cues (Anderson & Kaye, 2016; National Reading Panel, 2000). Specific skills, such as sound-symbol correspondence (SSC; graphophonemic cues) are taught incidentally as students are exposed to them. Readers monitor for graphophonemic cues by looking at the spelling of the word. Semantic cues involve the meaning of a text and vocabulary. Readers monitor for semantic cues by determining whether a word makes sense within the context of the text. For example, if a reader is monitoring for semantic cues, they may read the word “horse” as “pony.” Syntactic/language cues involve the grammar of the text. Readers monitor for syntactic cues by determining whether a sentence follows grammatical rules. For example, if a student reads “The horses are big” as “The horse is big,” the sentence is still syntactically correct which indicates the student is monitoring for syntactic cues but is not monitoring for graphophonemic cues.

Conversely, bottom-up reading models emphasize word recognition (Amadi, 2019). These approaches include instruction in phonological awareness, decoding, and sight recognition

(Scarborough, 2001). An example of a bottom-up reading model is a phonics-based instructional program. Phonics-based instruction emphasizes SSC and PA and the structure and meaning of a text is deprioritized. These approaches result in readers decoding a text to determine its meaning rather than predicting or monitoring for semantic and syntactic cues.

Bottom-up reading models indicate the progression of reading begins at the bottom with letter sounds which ultimately results in construction of meaning from a text (i.e., reading comprehension). Stages include letter identification, sound-symbol correspondence (SSC), decoding and phonemic awareness, syllable patterns, whole word reading, reading connected text, and meaning (Gough, 1972). Contrary to top-down approaches where meaning drives word reading, in bottom-up approaches, meaning of a text is derived from successful decoding.

There is significant disagreement among researchers regarding which of these approaches, whole language or phonics, results in better reading outcomes (Amadi, 2019; Bergeron, 1990; Bowers, 2020; Brooks, 2023; Fletcher et al., 2021; Glavach & Pribyl, 2018; Manning & Kamii, 2000; Wyse & Bradbury, 2022). Glavach and Pribyl (2018) studied whole language reading instruction with struggling readers across twenty elementary schools. Participants were selected based on their unresponsiveness to phonics instruction. While statistical data included in the article was minimal, most participants demonstrated reading gains as measured by the percentage of words correct, timed reading, prosody, and school assessments. The authors suggested that whole language instruction is beneficial for students who are unresponsive to phonics instruction.

In a study comparing isolated phonics instruction and phonics instruction embedded within whole language, kindergarteners who were taught phonics within the whole language framework demonstrated more progress in both reading and writing than kindergarteners who

were taught using isolated phonics (Manning & Kamii, 2000). The results of each of these studies (Glavach & Pribyl, 2018; Manning & Kamii, 2000) suggest that neither whole language nor phonics in isolation is sufficient for teaching students how to read.

Regardless of the approach used, neither the top-down approach nor the bottom-up approach fully accounts for all reading processes. However, Gough & Tunmer (1986) developed the simple view of reading, which states that decoding times language comprehension equals reading comprehension. According to this theory, a deficit in either decoding (i.e., bottom-up processing) or language comprehension (i.e., top-down processing) could result in a deficit in reading comprehension. Deficits in one or both of these areas result in three possible reading disorders: hyperlexia, in which decoding is sufficient but language comprehension is poor; dyslexia, in which decoding is poor but language comprehension is adequate; and general reading disability, in which both decoding and language comprehension are poor.

Scarborough (2001) developed a more complex model stemming from the simple view of reading called Scarborough's Reading Rope. In this model, word recognition (bottom-up processes) includes decoding, phonological awareness, and sight recognition. Language comprehension (top-down processes) includes vocabulary knowledge, background knowledge, language structures, literacy knowledge, and verbal reasoning. As word recognition processes become increasingly proficient and language comprehension becomes increasingly strategic, reading becomes more skilled and fluent. This model illustrates the complex nature of learning to read and incorporates both top-down and bottom-up processes.

Additionally, the five domains of reading instruction (National Reading Panel, 2000) incorporate components of bottom-up and top-down reading models. PA and phonics are considered components of a bottom-up reading model, while vocabulary and comprehension

instruction are components of a top-down reading model. Fluent reading is the product of sufficient word decoding and comprehension. Providing instruction in these five domains effectively consolidates the top-down and bottom-up reading models.

In a study directly comparing whole language (top-down approach) and explicit phonics (bottom-up approach) in two groups of first-grade participants, there were no significant differences between groups with regard to reading fluency or spelling accuracy (Maddox & Feng, 2013). The researchers recommended a combined approach that incorporates components of both whole language and explicit phonics. Combining both types of approaches is referred to as balanced literacy.

Balanced Literacy

Balanced literacy is a method of reading instruction in which the focus is balancing teacher-led and student-led activities as well as balancing phonics and whole language (Frey et al., 2005; Robinson et al., 2016; Willson & Falcon, 2018). Although components of a balanced literacy approach vary, some examples include read-aloud, shared reading, independent reading, modeled writing, shared writing, and independent writing (Willson & Falcon, 2018). Read-aloud and modeled writing are teacher-led activities. Independent reading and writing are student-led activities. In shared reading and shared writing activities, the teacher and students engage in interactive learning. There is also a focus on the use of leveled texts to address reading and writing skills within context (Robinson et al., 2016).

Although balanced literacy, by definition, combines components of top-down and bottom-up reading processes, it does not fully meet criteria for systematic phonics instruction. Because SSC is taught responsively within the context of a text, it cannot also be taught systematically (Bowers, 2020). In a study designed to identify effective balanced literacy

practices in elementary schools, phonics accounted for 2.9% of instruction (Bitter et al., 2009). The attempt to incorporate components of whole language and systematic phonics resulted in instruction that does not fully meet criteria for either.

Because balanced literacy incorporates components of top-down and bottom-up approaches, students may benefit from balanced literacy even though only some of its components are effective. In a study analyzing specific balanced literacy practices, the following were found to be effective for improving reading achievement: higher-level questioning, higher-level discussions, composition, and discussion of writing (Bitter et al., 2009). Of the observed activities identified in the article, comprehension strategies, vocabulary, and phonics were not found to be effective. However, this may be due to a lack of implementation rather than the ineffectiveness of the concept itself. In a study comparing the components of balanced literacy, only teacher-directed activities resulted in increased reading comprehension (Guo et al., 2023). Student-directed activities and independent learning tasks did not contribute to reading comprehension. Despite teacher-led instruction being a critical component of balanced literacy and critical for student success, it was found that teacher-led reading and writing activities were implemented significantly less often than student-led reading and writing activities (Frey et al., 2005). Although teacher-led activities are one component of balanced literacy, they are more strongly emphasized in the structured literacy approach.

Structured Literacy

Many currently consider structured literacy to be the "best practice" for reading instruction, especially for students with dyslexia or other reading disorders. Structured literacy practices are supported by the vast body of research referred to as the "science of reading" (Kuo, 2023). In a study comparing direct instruction (a component of structured literacy) and balanced

literacy, students who received direct instruction significantly outperformed students who received balanced literacy instruction (Robinson et al., 2016). Structured literacy is a bottom-up reading approach and is based on six key practices:

> explicit, systematic, and sequential teaching of literacy at multiple levels – phonemes, letter-sound relationships, syllable patterns, morphemes, vocabulary, sentence structure, paragraph structure, and text structure; cumulative practice and ongoing review; a high level of student-teacher interaction; the use of carefully chosen examples and non-examples; decodable text; and prompt, corrective feedback. (Spear-Swerling, 2018, p. 202)

Structured literacy practices are found in programs/curriculums such as Orton-Gillingham (Gillingham & Stillman, 1997) and Lindamood Bell Phoneme Sequencing Program (Lindamood & Lindamood, 2011).

When applying structured literacy practices, instruction should be explicit, systematic, and sequential at all levels of instruction from phonemes to text structure. Learning to read requires explicit knowledge. Therefore, concepts should be clearly stated and taught directly rather than expecting students to retain information through exposure or implicit methods. Concepts should also be taught systematically rather than responsively. Instead of finding areas of weakness to address, instruction follows a pre-planned sequence. Finally, instruction should be sequential, with easier concepts taught before more complex concepts. At the phoneme level, instruction may progress in the following sequence: letter sounds, consonant-vowel-consonant (CVC) words, blends and digraphs, long vowels, r-controlled vowels, vowel digraphs, and multisyllabic words (Walpole & McKenna, 2017).

Structured literacy practices include the use of cumulative practice and ongoing review. When new concepts are introduced, students should have the opportunity to practice previously learned concepts and the new concepts cumulatively. For example, students learning r-controlled vowels should have opportunities to read words with blends and digraphs in their practice texts. Review of previously learned concepts should also be frequent.

A high level of student-teacher interaction ensures that students apply their reading skills appropriately and have opportunities to learn error correction strategies. It is well-documented that students do not learn to read on their own (Lyon, 1998). This interaction also allows teachers to provide prompt, corrective feedback to ensure students apply reading concepts correctly. Students should be provided with many examples and non-examples to support appropriate generalization of spelling rules. Finally, using decodable texts to practice reading skills limits exposure to words that are too complex and allows students to achieve a high level of reading accuracy.

Assessment of Reading Ability

Pre-literacy skills, sound-symbol correspondence, and word reading ability are commonly assessed in early readers and are roughly aligned to three of the five components of reading instruction, according to the National Reading Panel (2000). Assessment of pre-literacy skills includes measures of PA, assessment of phonics includes measures of sound-symbol correspondence and decoding, and assessment of reading fluency includes measures of word reading efficiency. While vocabulary and comprehension are undoubtedly essential components of a comprehensive assessment, they are beyond the scope of this study.

Pre-literacy Skills

Pre-literacy skills consist of early-developing, reading-related skills that provide a foundation for learning to decode and directly relate to future reading ability. These skills can be divided into letter knowledge (SSC) and phonological processing (Vehkavuori et al., 2021). Letter naming is not a component of reading but is indirectly related to reading ability through orthographic awareness. A predictive relationship between reading real words and nonwords indicates that orthographic cues assist in phonological memory for new words (Cunningham et al., 2020). This suggests the importance of SSC.

Emerging literacy skills are taught in kindergarten through second grade (Georgia Department of Education, 2015). In these grades, students typically receive instruction in PA, phonics, and to some extent, reading fluency. According to the English Language Arts (ELA) Georgia Standards of Excellence (GSE), foundational reading skills addressed in kindergarten include understanding basic print concepts such as the order of text and the alphabetic principle; phonological awareness, including rhyming and phonemic awareness; letter sounds; and reading high-frequency words. In first grade, foundational reading skills include features of sentences; phonological awareness, including phoneme discrimination and blending and segmenting phonemes; decoding words with consonant digraphs, regularly spelled one-syllable words, final -e and vowel digraphs, and syllables in two-syllable words; and reading fluency to support comprehension. Reading fluency standards continue in second and third grades. In second grade, print concepts and phonological awareness are no longer included as standards. However, decoding skills become more complex and include words with prefixes and suffixes and words with inconsistent spelling patterns. For third-graders, decoding instruction also involves the meanings of prefixes and suffixes and reading multisyllabic words. By fourth grade, phonics

standards no longer include specific skills. Instead, instruction addresses the application of decoding skills to read unfamiliar words (Georgia Department of Education, 2015). As students in kindergarten through third grade learn phonological awareness skills and phonics, phonological processing deficits may be revealed.

Phonological Processing in Early Readers. Phonological processing is a well-established area of weakness for children with reading disabilities. Eighty percent of struggling readers have deficits in the phonological domain of language that could prevent them from developing the word attack skills needed for fluent reading (Moats & Tolman, 2009). Phonological processing consists of phonological awareness, phonological retrieval, and phonological working memory (Wagner & Torgesen, 1987). Research studies have supported this theory of three separate but correlated phonological constructs (Kibby et al., 2014; Nelson et al., 2012). Additionally, Brandenburg et al., (2017) found that this theory transfers beyond the English language to German. However, Choi et al. (2016) argued that there is significant overlap in phonological awareness and phonological memory to the point that these constructs cannot be assessed independently. This is a logical conclusion because phonological working memory allows an individual to hold phonological information in their memory while completing phonemic awareness tasks. This argument is also supported by the findings of studies which are discussed later in this section.

Phonological processing refers to one's knowledge of the sound system of a language and the ability to retrieve those sounds. Children must have strong phonological representations of words in order to decode. In a study of phonological processing, eye movement during silent reading was used to measure dependence on phonological information in written words and nonwords. Children (ages seven to nine) demonstrated similar eye movements as adults, and

results indicated that the children processed phonological information during lexical reading in a manner similar to adults (Blythe et al., 2015). This suggests that phonological skills, and therefore, phonological representations, in children are developed early.

Each of the phonological processing constructs (phonological awareness, phonological retrieval, and phonological working memory) includes a variety of skills and may be assessed in multiple ways. Four studies were reviewed to determine the validity of phonological processing as an indicator of reading ability (Fedora, 2016; McArthur & Castles, 2013; Ozernov-Palchik et al., 2017; Papadopoulos et al., 2020). Fedora (2016) examined the double-deficit hypothesis (deficits in both PA and RAN) in predicting first grade reading achievement for first-grade students in Title I schools. They found that participants with a double deficit were not significantly more impaired than participants with a single deficit in either PA or RAN. They also found that there was no significant interaction between PA and RAN which indicates that they are independent influences on reading ability. McArthur & Castles (2013) examined the relationships among PA, phonemic decoding, and word reading in seven- to twelve-year-old participants with specific language impairment (SLI), specific reading disability (SRD), and both. They found that participants with SRD had impairments in phonological awareness but not participants with only SLI. Ozernov-Palchik et al. (2017) examined pre-literacy skills in kindergarteners, specifically how those skills change over time. Six reading profiles were identified: average, low-average, high, double-deficit risk, RAN risk, and PA risk. Participants' classification into these profiles remained consistent into their first-grade year. Papadopoulos et al. (2020) analyzed reading and spelling ability related to PA, RAN, and phonological working memory. This was a longitudinal study of participants beginning in kindergarten and ending in second grade. They found that good readers scored significantly higher than poor readers on

measures of PA and RAN. Results also indicated that participants with deficits in both PA and RAN had weakness in phonological working memory.

Developmentally, phonological and phonemic awareness skills generally follow a hierarchy in which awareness of larger components, such as syllables, is achieved before awareness of smaller units (i.e., phonemes). Phonological awareness is the more general term that refers to skills with larger components of words, such as rhyming, segmenting syllables, and combining single words into compound words. PA skills are a subset of phonological skills related to phonemes rather than larger word parts. PA skills include blending, segmenting, and manipulating phonemes in words (Fedora, 2016).

Because PA is a stronger predictor of reading ability than more general phonological awareness skills (Clayton et al., 2020; Cunningham et al., 2020; Loucas et al., 2016), PA is considered here. PA measures are a vital method for diagnosing and monitoring reading disabilities. PA refers to an individual's understanding of phonemes in language (McArthur & Castles, 2013; Stahl & Murray, 1994, as cited in Ozernov-Palchik et al., 2017) and contributes to word reading ability (Fedora, 2016). The most common measures of PA in these studies were elision and blending (Fedora, 2016; Ozernov-Palchik et al., 2017; Papadopoulos et al., 2020). Elision involves deleting sounds in a given word, and blending involves blending individual phonemes together to form a word. Other measures of PA in these studies included sound matching, phoneme discrimination, and alliteration (Fedora, 2016; McArthur & Castles, 2013). Sound matching involves matching a sound with a word that starts with the same sound, phoneme discrimination involves distinguishing between two given sounds, and alliteration involves identifying words that start with the same sound. Deficits in PA skills may result in a reading disability.

The speed of phonological retrieval refers to the rate at which an individual can access phonological representations of words (Fedora, 2016). RAN is a common measure of phonological retrieval and consists of presenting an individual with a series of pictures of objects, colors, digits, and/or numbers. The individual's response to naming the items is timed. RAN tasks were used to measure phonological retrieval across all studies. Object RAN was the most common measure. Other forms included digit and color RAN (Fedora, 2016; Ozernov-Palchik et al., 2017; Papadopoulos et al., 2020). Object and color RAN were likely selected instead of alphanumeric RAN due to the participants' developmental levels. Alphanumeric RAN does not provide an accurate representation of phonological retrieval in participants who have not yet mastered the names of numbers and letters. These studies included emerging readers who may not be sufficiently familiar with letters to complete alphanumeric RAN accurately.

Phonological working memory measures were not present in any of the four studies. However, phonological short-term memory was included in two studies, both of which used a nonword repetition task (McArthur & Castles, 2013; Ozernov-Palchik et al., 2017). Phonological short-term memory refers to an individual's ability to recall phonological information over short periods (Ozernov-Palchik et al., 2017). Papadopoulos et al. (2020) included verbal working memory in their study with measures of verbal-spatial relations and sentence questions. Although these measures are not as specific as those used to measure phonological working memory, they are related skills. Ozernov-Palchik et al. (2017) provide significant insight into the constructs relevant to reading disorders. One purpose of their study was to determine reading profiles in kindergarteners. Although phonological short-term memory was lowest in participants with a PA or double deficit, no separate phonological short-term memory profile was indicated. These results suggest that phonological short-term memory may not provide relevant diagnostic

information for reading disorders.

PA and RAN are Associated with Reading Ability. All the studies yielded results indicating that PA and RAN are reliable indicators of reading ability. Papadopoulos et al. (2020) assessed PA and RAN in good and poor readers. They found that good readers (PA: $M = 101.35$, $SD = 12.36$; RAN: $M = 102.02$, $SD = 13.13$) scored significantly higher on measures of PA and RAN than poor readers (PA: $M = 90.28$, $SD = 12.38$; RAN: $M = 88.92$, $SD = 7.80$). Participants were not grouped based on their reading profiles. Although the results of this study are limited for the purposes of the current discussion, they indicate a difference in PA and RAN in good versus poor readers.

Ozernov-Palchik et al. (2017) identified six reading profiles in kindergarteners using latent profile analysis: average, low-average, high, PA risk (word attack *z*-score = -.027; PA *z*-score = -.052; RAN *z*-score = 0.29), RAN risk (word attack *z*-score = 0.03; PA *z*-score = -0.14; RAN *z*-score = -0.42), and double-deficit risk (deficits in both PA and RAN; word attack z-score = -1.06; PA *z*-score = -0.50; RAN *z*-score = -0.60). The separation of PA and RAN into separate profiles suggests that these variables produce independent influences on reading ability. Participants' classifications had not changed by the end of first grade, suggesting that early reading profiles are stable, at least through the emerging reading phase. Further research is required to determine if these profiles are stable beyond this phase.

Fedora (2016) supports these findings. In this study, participants were divided into groups that align with the reading profiles identified by Ozernov-Palchik et al. (2017). Groups included no deficit, RAN deficit, PA deficit, and double deficit in both RAN and PA. Reading outcomes included measures of letter-word identification and word attack. Participants with a PA deficit ($n^2 = .26$) had poorer reading ability than participants with a RAN deficit ($n^2 = .09$). Participants

with a double deficit were not significantly more impaired than those with a single deficit. There was also no significant main effect between participants with a RAN deficit and participants with a PA deficit. This supports the findings in Ozernov-Palchik et al. (2017) that PA and RAN contribute differentially to reading ability. The results of both of these studies (Fedora, 2016; Ozernov-Palchik et al., 2017) support Choi et al.'s (2016) argument that phonological memory is not an independent factor of reading ability. Instead, it is enmeshed with phonological awareness.

It would be easy to assume that PA and RAN may also be deficient in children with language impairment due to the close relationship between oral and written language. McArthur & Castles (2013) examined PA and RAN in participants aged seven to twelve with specific language impairment (SLI), specific reading disorder (SRD), and comorbid SRD and SLI. Across all measures, participants with SRD, with or without SLI, had impaired PA (SRD only: $d = 1.26$; SRD and SLI: $d = 2.33$) and RAN (SRD only: $d = 0.86$; SRD and SLI: $d = 1.45$), while participants with SLI only had scores within the average range for all areas. Based on these results, measures of PA and RAN are specifically indicative of reading ability and less associated with oral language ability.

Across all four studies, PA and RAN had significant and differential effects on reading ability. These findings support the conclusion that PA and RAN are highly sensitive indicators of reading disorders in emerging readers. For this reason, they should be included in the screening and assessment batteries of reading ability. Because the research supports the use of various reading profiles, assessment in both areas is essential. Specifically, phonological processing measures may be used to identify specific reading profiles to select the most appropriate interventions for each reader. Assessment in only one area would have decreased sensitivity to

the various reading profiles.

A limitation of the current literature is the variation in cut scores. The cut score may be 1-1.5 standard deviations below the mean. These studies of PA and RAN provide theoretical information regarding the influence of phonological processing on reading ability. However, they do not offer a clear distinction between what is considered "no deficit" and "deficit" within each area. A cut score of the 16th percentile on measures of phonological processing and reading ability was used for grouping in the studies by Papadopoulos et al. (2020) and McArthur & Castles (2013), but a cut score of the 25th percentile was used in the study by Fedora (2016). By determining whether reading achievement differs between typical readers, readers at risk of a disability, and readers with identified disabilities, recommendations may be made for a specific cut score.

Sound-Symbol Correspondence (SSC)

The second component of a reading assessment is sound-symbol correspondence. While children are developing PA, they are also learning SSC, which refers to pairing phonemes with their corresponding letters. SSC is also referred to as letter-sound knowledge or phoneme-grapheme correspondence. This skill is taught directly during phonics instruction. Used together, PA and SSC predict word reading ability at least through the age of five years (Clayton et al., 2020; Hulme et al., 2012). Although SSC is a crucial skill, PA is a stronger indicator of word reading ability (Hulme et al., 2012). A ceiling effect of this skill prevents using SSC as a predictor of reading ability. However, poor sound-symbol correspondence results in poor decoding ability.

There is disagreement on whether SSC must be fully integrated to read successfully. Letter-sound integration (LSI) refers to the automaticity of SSC. Clayton & Hulme (2018) found

that an LSI priming task did not differentiate between children with dyslexia and typically developing controls. However, there was a significant difference in response times between the two groups. This difference indicates that children with dyslexia have longer processing times for sound-symbol correspondence. Delays in processing put a greater demand on verbal short-term memory and negatively impact decoding ability. The findings from a study by Clayton et al. (2020) also support the relationship between response time for SSC and word reading ability. Aravena et al. (2013) also measured LSI. The results of their study indicated that children with dyslexia required more processing time for letter sounds and made more errors during learning. However, SSC after a thirty-minute session was not significantly different from the typically developing control group. The lack of difference between groups may be due to a ceiling effect of SSC which supports the findings of Hulme et al. (2012). Overall, response time of sound-symbol correspondence was most associated with dyslexia. Children with dyslexia require greater response time to correctly identify letter sounds, even when they can correctly identify all letter sounds. Therefore, children with dyslexia are likely to identify fewer correct letter sounds than typically developing children given time constraints.

Word Reading Ability

The third component of reading assessment is word reading ability. To comprehend a text, children must be able to read words with sufficient proficiency so they can allocate cognitive resources toward comprehension rather than decoding. Assessment of word reading ability includes both decoding and sight word reading. These may include nonsense word fluency, real word fluency, and oral reading accuracy. Real word lists and reading passages may be read with a combination of decoding and sight word reading. Nonsense word fluency

indicates one's ability to decode efficiently, involves several phonological processes, and is more likely to be impacted by deficits in those areas.

Morris et al. (2013) suggest assessing word reading ability with both word lists and connected text. Higher fluency and accuracy with connected text compared to word lists may indicate an overdependence on context. RAN may provide a link between individual word reading and reading connected text because rapid phonological retrieval allows for more fluent reading ability (Protopapas et al., 2018). In other words, the ability to rapidly name a series of visual stimuli (RAN) and the ability to blend sounds together (PA) results in more fluent reading.

Reading fluency can be separated into text reading efficiency and prosody (Benjamin & Schwanenflugel, 2010; Dowhower, 1991; Kim et al., 2021). Kim et al. (2021) suggest that the assessment of word reading ability should include accuracy/rate and prosody because they each contribute differentially to word reading ability. When considered together, they are a reliable indicator of reading comprehension. Text reading efficiency includes both accuracy and speed. Prosody includes pausing and pitch. A true measure of reading fluency should include both word-reading efficiency and prosody. However, prosody can be subjective, making it difficult to measure accurately. Benjamin and Schwanenflugel (2010) reviewed methods of assessing prosody as part of oral reading fluency. The first method was to administer parsing tests in which participants place commas in a text that is read aloud. However, this skill may not transfer to expressive prosody. The next method reviewed was the use of rating scales which do not fully capture prosody due to a limited number of rating options. Additionally, the use of rating scales may be influenced by other components of reading fluency such as rate. The final option, which was selected by the researchers in the study, was the use of spectrographic measures. Spectrographic measures are used by researchers to provide a more objective measure of reading

prosody (Benjamin & Schwanenflugel, 2010; Dowhower, 1991; Kuhn et al., 2010) but are not commonly available in applied settings. Therefore, rating scales offer an easy-to-use alternative although they lack the validity and reliability needed for diagnostic reading measures (Kuhn et al., 2010). Despite the benefit of assessing prosody, word reading efficiency, including rate and accuracy, is easier to quantify, and therefore, easier to measure. This often results in the exclusion of prosody in reading assessments (Dowhower, 1991).

Word reading rate increases until around fourth grade when it plateaus (Morris et al., 2013). Morris et al. (2013) recommended the following reading fluency benchmark scores for second through sixth grade: second 80 words per minute (wpm), third 90 wpm, fourth 100 wpm, fifth 105 wpm, and sixth 110 wpm. Benchmarks were determined based on the wpm that fell between the 25th and 30th percentiles at each grade level.

These scores conflict with other recommended WPM minimums. Good et al. (2019) recommended the following end-of-year benchmarks: first 47 wpm, second 87 wpm, third 100 wpm, fourth 115 wpm, fifth 130 wpm, and sixth 120 wpm. Benchmarks differ from Morris et al. (2013) roughly by 10 wpm in each grade level. In an international study, WPM varied greatly across languages and did not predict reading comprehension (Dowd & Bartlett, 2019). In this study, the average WPM for second-grade participants was 51 and ranged from 30 to 96. The average is much lower than the benchmarks published by Morris et al. (2013) and Good et al. (2019). Based on the results of their study, Dowd and Bartlett (2019) recommend a shift away from wpm benchmarks as measures of reading ability.

Despite the disagreement on benchmark scores, Hasbrouck and Tindall (2017) published widely accepted and utilized oral reading fluency norms (Table 1) which may be beneficial in the assessment of oral reading fluency. These norms were created based on the compilation of data

from three widely used ORF assessments over a span of 25 years and include a much larger number of participants than other ORF studies.

Table 1

WCPM Norms (Hasbrouck & Tindall, 2017)

Grade	Spring WCPM 50th Percentile
1	60
2	100
3	112
4	133
5	146
6	146

Based on the inconsistencies of recommended reading fluency benchmarks, word reading accuracy may provide more insight into comprehension. In Morris et al.'s study (2013), word reading accuracy consistently fell at 95-96% accuracy for text that was sufficiently comprehended based on comprehension questions. In Dowd & Bartlett's study (2019), word reading accuracy of comprehended text exhibited a much narrower range across languages (94%-99%). Participants in this study could read as slowly as 30 words per minute and comprehend the text sufficiently. These findings (Dowd & Bartlett, 2019; Morris et al., 2013) suggest that word reading accuracy is a much more accurate predictor of reading comprehension than the number of words read per minute.

Word reading ability involves the combination of many skills and should be assessed as such. Recommendations by Protopapas et al. (2018) indicate that it is important to use words that are familiar to the reader when strictly measuring reading fluency. When measuring accuracy, reading speed is not important. Reading fluency, word reading accuracy, and language comprehension are enmeshed when reading grade-level text.

RTI in Reading

Assessment of reading ability in schools is often completed as part of the RTI framework (National Center on Response to Intervention, n.d.). This framework is commonly used in schools to identify students at risk for reading disorders and provide increasingly intensive interventions to students performing below grade-level expectations. The RTI team may support the student, make decisions regarding placement in the tiers, and determine what interventions will be provided. This team may include teachers, speech-language pathologists, intervention specialists, counselors, and administrators.

Tier one instruction is the general instruction provided in all classrooms. This should consist of an evidence-based curriculum. Based on the science of reading, this curriculum should be based on structured literacy practices. Instruction at this level should be designed to meet the needs of most students.

Tier two intervention is typically provided to students performing below expectations in small groups a few times each week. Intervention targets are dependent on the general area of weakness. For example, if several students in a class have poor PA skills, their small-group intervention should address PA.

Tier three intervention is typically individualized intervention offered multiple times per week to students who did not demonstrate sufficient growth with tier two intervention. Tier three intervention may be provided by a specialist and is specifically designed based on the strengths and weaknesses of the individual student. For a student with deficits in PA, instruction should address specific areas of PA such as blending.

Responsiveness to these interventions, specifically tier three interventions, is also used to determine if a referral is needed for special education eligibility testing. Educators must identify

which students will likely qualify for services to make accurate referrals and provide appropriate interventions. However, the measures and types of scores used to document students' performance in RTI are variable. Measures may include standardized testing, curriculum-based measures, or observation. Standardized assessments and classroom assessments both provide insight into students' language, cognition, and reading abilities.

The types of assessments used for individual students are much more limited. Educators are typically only able to administer standardized assessments universally. Formal evaluation of individual students requires parental consent and extra resources. Therefore, most classroom-administered assessments are short and limited to screening tests. These assessments may focus on pre-literacy skills, SSC, and word-reading ability. Research using these types of assessments is critical in understanding and identifying risk factors in students' reading ability.

Assessment of Reading Ability in Underserved Populations

Ensuring that the number of students identified as at-risk readers is proportionate with regard to race and gender of the general population is a consideration that must be made under the Individuals with Disabilities Education Act (IDEA) (Office of Special Education Programs, 2017). Therefore, it is essential to identify factors that influence disproportionality and establish systems for identifying at-risk students that are free of bias.

In a study of self-reporting as a measure of risk status, 7.2% of the variance in special education was explained by demographic variables, while only 1.2% of the variance in at-risk students was due to demographics (Dever et al., 2016). Subgroups overrepresented in special education were Black Americans, males, and students who qualified for free or reduced lunch. These groups should be the primary focus for reducing disproportionality.

In a study of Black boys and girls, boys were much more likely to have weak reading comprehension, but language and cognition were similar between genders (Washington et al., 2019). This indicates that the proportion of boys and girls receiving services for language and cognition should be similar even though more boys may receive intervention for reading. According to Scarborough's reading rope, skilled reading consists of language comprehension and word recognition (Scarborough, 2001). If boys and girls do not differ in language skills, this suggests that the boys' weak reading comprehension scores are due to word recognition ability. This deficit may be due to poor phonological awareness, decoding, and/or sight recognition.

Due to the disproportionality based on race, gender, and socioeconomic status in special education, it is important to ensure that assessments are not biased toward any specific subset of the population. The RTI model has been correlated with reduced disproportionality for males. O'Connor et al. (2013) found that more girls were identified with a learning disability through RTI referrals than non-RTI referrals. Therefore, RTI may support the proportionality of students in special education. The use of nonbiased universal screeners within the RTI framework is one method of minimizing disproportionality.

Universal Screeners

Brief assessments administered to all students in a specified population (e.g., kindergarten through third grade), are a common tool used in RTI to identify students at risk of a reading disorder. These universal screeners provide educators with data to guide decision-making in the RTI framework. These screeners can also be used as a measure of long-term growth.

Universal screeners for reading include AIMSweb (Pearson Education, 2011), FAST (FastBridge Learning, 2018), and Acadience Reading Diagnostic (Good & Kaminsky, 2020).

These screeners are commonly utilized in research and are widely used by school systems to identify at-risk readers. Table 2 outlines how these screeners address each area of reading ability: pre-literacy skills, SSC, and word reading. All screeners assess initial phonemes, letter naming, phoneme segmentation, letter sounds, nonsense word reading (decoding), and reading of connected text. However, there are a few key differences. FAST includes the most measures of PA, while Acadience is the only screener that includes RAN. FAST is the only measure that includes a separate assessment of sight word reading.

Table 2

How Universal Screener Subtests Address Reading Ability

Reading Ability	AIMSweb	FAST	Acadience
Pre-Literacy Skills	Initial sounds Letter naming fluency Phoneme segmentation	Onset sounds Word rhyming Word blending Word segmenting Letter names	First sound fluency Phoneme segmentation fluency RAN Letter naming fluency
SSC	Letter word sound fluency	Letter sounds	Nonsense word fluency (correct letter sounds)
Word Reading	Word reading fluency Nonsense word fluency Oral reading fluency	Decodable words Nonsense words Sight words Sentence reading	Nonsense word fluency Oral reading fluency (words correct, accuracy) Maze

Acadience Reading Diagnostic

The Acadience Reading assessment includes seven subtests. All subtest descriptions are found in the Acadience Reading Technical Manual (Good et al., 2019) and the Acadience Reading Assessment Manual (Good & Kaminsky, 2020).

The first sound fluency subtest is a measure of PA in which students state the first sound of a word presented orally. Three practice items are provided along with error correction strategies. Students have one minute to correctly identify the first sounds of words presented

orally by the test administrator. The first sounds include continuants, stops, and blends. Two points are given for correctly identifying the first sound, and one point is given for identifying the first consonant blend, consonant plus vowel, or consonant blend plus vowel. For example, if the student is presented with the word "swim," they are awarded two points for identifying /s/ and one point for identifying /sw/. If no response is provided within three seconds of presenting the word, no score is given for that word, and the next word is presented. The subtest is discontinued if the student receives zero points in the first five words.

The letter naming fluency subtest is a measure of letter knowledge in which students label upper- and lower-case letters. Students are presented with a page of upper- and lower-case letters organized in rows. Students have one minute to name each letter correctly. If no response is provided within three seconds, the letter is named by the test administrator. The subtest is discontinued if no letters are named correctly in the first row which consists of ten letters. The test is scored based on the number of letters correctly named in one minute.

The phoneme segmentation fluency subtest is a measure of PA in which students state each of the phonemes in a word presented verbally. Students have one minute to name each phoneme in words presented orally by the test administrator. One point is given for each word segment provided by the student. For example, if students segment "trap" as /t/ /r/ /a/ /p/, they are given four points. Full credit is provided for words in which each phoneme is segmented successfully. The subtest is discontinued if the student has zero correct sound segments in the first five words.

The RAN subtest includes measures of object, number, and letter RAN. This subtest is currently optional and does not contribute directly to the RCS. Each of the RAN tests includes practice items with error correction strategies. The test is discontinued if the student does not

correctly name all items in the practice trial. Additionally, the test is discontinued if the student makes four or more errors on the first two rows. The student is presented with a page with objects, numbers, or letters depending on the specific subtest administered and prompted to label all items on the page. Students have three seconds to identify each picture, letter, or number. The total time in seconds is recorded. On the RAN objects subtest, students may be corrected if they provide an incorrect name of an item (e.g., "kitty" instead of "cat"). Additionally, students may be redirected if they skip a line on the page. Individual skipped items are scored as incorrect. Each of the RAN subtests is scored based on the amount of time in seconds it takes the student to complete the tasks and the number of errors.

The nonsense word fluency (NWF) subtest is a measure of decoding in which students read CVC non-words. Students are administered two practice items with error correction strategies. After the practice items, students are presented with a sheet of CVC non-words and prompted to read the words or name the letter sounds if they cannot read the whole word. Students may read the word as a whole, for example, "sog," or they may name the sounds in the word, /s/ /o/ /g/. They have one minute to read the words. If the student does not read a word or provide a letter sound within three seconds, the word or sound is read by the test administrator. The subtest is discontinued if the student has no correct letter sounds in the first row. This subtest results in scores for correct letter sounds (CLS) score and whole words read (WWR).

The oral reading fluency (ORF) subtest is a measure of reading fluency and comprehension in which students read grade-level passages and retell the story. The student is presented with three grade-level passages and prompted to read each aloud. The student has one minute to read each story. The subtest is discontinued if no words are read correctly in the first line or if fewer than ten words are read correctly in passage one. This portion of the subtest is

scored by counting the number of words read correctly (WC), the number of errors, and reading accuracy in the form of a percentage. After reading each passage, the student is prompted to retell the story. This portion of the subtest is scored by counting the number of words in the retell and scoring the quality on a scale of one to four based on a rubric. According to the rubric, the quality of the retell response is classified as "one" if the student provides two or fewer details, "two" if the student provides three or more details, "three" if the student provides three or more details in a meaningful sequence, or "four" if the student provides three or more details in a meaningful sequence that captures a main idea. The story retell portion is optional if the student reads fewer than forty words correctly in the reading passage.

The maze subtest is a measure of reading comprehension that requires students to use comprehension strategies, word recognition skills, and prior knowledge. This is the only subtest that may be administered in groups. Students are presented with a grade-level passage in which approximately every seventh word is missing. In place of the word are three choices. Students have three minutes to read the passage and select the appropriate words by circling them. The score is the number of correct words minus half the number of incorrect words.

Scores from administered subtests, excluding RAN, are combined to calculate an RCS. The scores from each subtest are added together to calculate the RCS. NWF CLS is only included in the RCS for middle-of-year first grade. For NWF WWR, the score is multiplied by two. For ORF, the accuracy percentages are converted to an accuracy value ranging from zero to 120. For maze, the score is multiplied by four. Normative data for the end-of-year test administration can be found in Table 3.

Table 3

End-of-Year RCS Norms

Grade	25th percentile	50th percentile	75th percentile
Kindergarten	114	143	173
First grade	117	193	254
Second grade	202	273	126
Third grade	297	381	456

Because different subtests are administered at different parts of the year and different grade levels, the RCS cannot be compared between test administrations or between grade levels. However, the classification of the RCS relative to the benchmark can be generalized.

Scores below the 2nd percentile are considered the lower extreme. Scores between the 3rd and 8th percentiles are considered well below average. Scores between the 9th and 24th percentiles are considered below average. Scores between the 25th and 75th percentiles are considered average. Scores between the 76th and 90th percentiles are considered above average. Scores from the 91st to 97th percentiles are considered well above average. Scores at the 98th percentile and above are considered the upper extreme.

Classification of Risk Status

When implementing universal screeners in RTI, two types of scores are used to report performance and determine risk for disability: achievement scores and growth rates. Achievement scores indicate an individual's ability at a single point in time. When using achievement scores, students in RTI are determined "responsive" if they meet a specified score, usually set as equal to or less than one standard deviation below the mean or a specified benchmark. Growth rates are typically calculated by subtracting a current achievement score from a previous achievement score to calculate a growth rate over time (Hendricks & Fuchs, 2020). Using growth rates, students are considered responsive if they demonstrate sufficient

growth. In some cases, sufficient growth may be defined as any amount of progress. In other cases, sufficient growth may be defined as growth sufficient to catch up with same-grade peers. In all cases, unless there is a clearly defined criterion, sufficient growth is subjective. When using achievement scores, students identified as responsive fall within the average range. When using growth rates, students who demonstrate growth are not referred for further testing, regardless of their achievement scores. Substantial research has been completed on the use of achievement scores and growth rates when determining the need for intervention. Regardless of the model used, responsivity is generally considered a protective factor against disability.

Research has generally shown growth rates to be unrelated to disability (Al Otaiba et al., 2011; Maki et al., 2020; Schatschneider et al., 2008; Tindal et al., 2016; Van Norman & Nelson, 2021). Hendricks and Fuchs (2020) explain that this phenomenon is partly due to a ceiling effect. Students performing significantly below expectation have the potential for significantly more growth than students performing at expectation. Therefore, growth rates for low-performing students have the potential for much more variation.

Another explanation for these results may be a lack of differences in growth rates between students with and without weaknesses in reading abilities. Growth rates in word recognition and reading comprehension are not significantly different in students with typical or impaired language (Catts et al., 2008). In another study of reading ability, language ability, and cognition, gains made by students receiving tier-three intervention were not significantly different from gains made by students receiving typical instruction (Denton et al., 2013).

Although limited, growth rates have some use in the RTI model. Research has shown growth rates to be sensitive to changes in students performing significantly below grade level

(Field et al., 2019; Hendricks & Fuchs, 2020). However, these scores cannot be used in isolation, nor can they be accurately used for all students.

There are benefits to using either achievement scores or growth rates. This leads to the question of which is better suited for identifying students at-risk of reading disorders. Milburn et al. (2017) compared four methods of classifying risk status for reading disorders: IQ-achievement discrepancy, LA, LG, and DD (in which participants met criteria for both LA and LG). The IQ-achievement discrepancy was defined as at least one standard deviation between IQ and reading achievement. However, this method is not feasible for RTI use, because IQ is not widely assessed in schools. This leaves three other methods remaining.

In Milburn et al.'s study (2017), LA was defined as end-of-year scores at three levels (i.e., 5^{th}, 10^{th}, and 25^{th} percentiles). LG was defined as less-than-expected growth based on the initial level of achievement. The statistical methods used for determining LG accounted for the floor and ceiling effects that typically cause growth rates to be unreliable. DD criteria were the combination of LA and LG criteria. Among these three methods, there was weak to moderate agreement which presents a barrier to identifying at-risk students. Depending on the method used, identifying at-risk students is variable and unreliable. To provide the most sensitivity and specificity, Milburn et al. (2017) recommended using the DD model. This model is likely to identify the fewest students but has a high degree of specificity.

In another study, Milburn et al. (2018) compared the use of achievement, growth, and both to define "responsiveness" in RTI tier two. Participants were placed in tier two if their scores on any measure of literacy fell at or below the 25^{th} percentile. After 11 weeks of tier-two instruction, students were assessed again. Participants were classified as responsive based on achievement if their scores fell above the 25^{th} percentile. They were classified as responsive

based on growth if they demonstrated at least three standard score points of growth. Participants were also classified as responsive based on both criteria. Agreement among these methods was poor, and the authors indicated a need to establish a clear definition of responsiveness as it pertains to RTI.

Beyond achievement scores and growth rates, there are shared characteristics of students who typically have an inadequate response to the given intervention. These factors may indicate a need for a special education referral. Persistence of weakness, poor nonverbal cognition, poor reading comprehension, and more severe weakness are all related to inadequate responsiveness (Simmons et al., 2008). Students typically respond promptly when provided with an intervention (Simmons et al., 2008). Therefore, students who do not respond promptly are likely to have persistent difficulties that are resistant to intervention and result in long-term weakness (Catts et al., 2002). Low nonverbal IQ is related to poorer reading skills in the presence or absence of language impairment (Catts et al., 2008; Catts et al., 2002). Weak reading comprehension in early grades is associated with future weakness in word reading ability and reading comprehension (Catts et al., 2006). Finally, students with more severe impairments are more likely to have an inadequate response to intervention (Denton et al., 2013). This may be due to the previously discussed lack of difference in growth rates between students with and without disabilities.

Purpose

The primary purpose and first component of this study was to compare various methods (LA, LG, and DD) of interpreting reading assessment scores to identify students at risk for reading disorders in a Title I school. This is important because educators must be able to quickly identify students who need intervention without undergoing a comprehensive assessment that is

unavailable on a widespread scale in schools. Educators can more consistently identify students who are genuinely at-risk for reading disorders and provide appropriate interventions more quickly and more objectively if given specific benchmark scores and a method for interpreting those scores. To accomplish the purpose of this study, reading achievement and growth were compared among students with reading and reading-related disabilities, students at risk for reading disabilities, and typically developing readers. The level of agreement among the three methods (LA, LG, and DD) of identifying the risk of reading disabilities was established. Finally, comparing these methods using local norms versus published norms provided insight into how the norms affect the identification of at-risk students in Title I schools. To fulfill the purposes of this study, several questions were answered.

1. How do reading achievement scores and growth rates differ among kindergarten through third-grade students with identified reading and reading-related disabilities, students at risk of reading disabilities, and typical readers?
2. What is the level of agreement among three methods of classifying risk status (LA, LG, and DD) using published norms?
3. How does the use of local norms on a universal reading measure compare to the use of published norms in the identification of at-risk readers?

Chapter III

METHODOLOGY

The purpose of this study was to compare the use of three different methods of identifying at-risk readers in addition to comparing the use of published and local norms. Milburn et al. (2017) studied the use of LA, LG, and DD as methods of identifying at-risk students. In this study, these methods are applied to the identification of at-risk reading using a universal reading measure implemented within the RTI model in Title I public schools.

Research question one was, "How do reading achievement scores and growth rates differ among kindergarten through third-grade students with identified reading and reading-related disabilities, students at risk of reading disabilities, and typical readers?" Research question two was, "What is the level of agreement among three methods of classifying risk status (LA, LG, and DD) using published norms?" Research question three was, "How does the use of local norms on a universal reading measure compare to the use of published norms in the identification of at-risk readers?"

This study was conducted in two Title I elementary schools using Acadience Reading Diagnostic RCS to measure reading achievement and growth in students with reading disabilities, at-risk of reading disabilities, and typical readers. RCS from Acadience were used to classify the risk status of participants based on criteria for LA, LG, and DD. To answer research question one, RCS from each of the three groups (identified reading disability, at-risk for reading disability, and typical readers) were compared. To answer research question two, the level of agreement between each of the three risk classification methods (LA, LG, and DD) was

calculated. To answer research question three, RCS for the studied schools were compared to the published norms, and the risk classification methods were applied to participants based on local norms.

Participants

Participants included 514 emerging and early readers in kindergarten through third grade in two rural Title I K-5 elementary schools (Appendix A; Appendix B). Students who participated in universal screening during all three administrations during the school year were included as participants in the study. Of the participants, 46% identified as Black, 35% identified as White, 12% identified as Hispanic/Latino, 5% identified as two or more races, less than 1% identified as Asian or Pacific Islander, less than 1% identified as American Indian or Alaskan Native, and less than 1% identified as Native Hawaiian or other Pacific Islander. Male participants were 48.4% of the sample, and female participants were 51.5% of the sample.

Participants were categorized as "identified" with a reading or reading-related disorder, "at-risk" for a reading disorder, or typical readers based on the services they were receiving at the end of the academic year. Participants with identified reading or reading-related disorders were those receiving any type of special education service under the domain of reading through their Individualized Education Programs (IEPs). Participants at risk for a reading disorder were those who received tier two or tier three reading intervention through RTI. To qualify for tier two interventions, participants were referred by their classroom teachers. After at least nine weeks of tier two instruction, participants could be referred for tier three intervention by their RTI team, including the classroom teacher, parent(s), and RTI coordinator. Tier two and tier three reading interventions were implemented by the classroom teacher or intervention specialist. Typical readers were those who received no reading-related intervention. At the end of the academic year

studied, 31 participants met the criteria for the identified reading disorder group because they received reading services in their Individualized Education Programs. Additionally, 130 participants were in the at-risk group because they received reading intervention through RTI tier two or three. The remainder of the participants (n = 353) were in the typical reader group.

Measures

Acadience Reading (Good & Kaminsky, 2020) was used to assess participants' reading across a range of skills based on their grade. This measure was selected based on its use within the Title I schools involved in the study. Table 4 includes the subtests administered during each testing period of the academic school year.

Table 4

Acadience Reading Diagnostic Subtests Administered

Grade	Testing Period	First sound fluency	Letter naming fluency	Phoneme segmentation fluency	Nonsense word fluency	Oral reading fluency
Kindergarten	Beginning	X	X			
	Middle	X	X	X	X	
	End		X	X	X	
First grade	Beginning		X	X	X	
	Middle				X	X
	End				X	X
Second grade	Beginning				X	X
	Middle					X
	End					X
Third grade	Beginning					X
	Middle					X
	End					X

Assessment Administration Procedures

Acadience Reading was administered individually to all participants during each test administration window. The maze subtest was administered to each class as a group. The beginning-of-year administration occurred from August 9th through August 26th, the middle-of-

year administration occurred from January 11th through January 18th, and the end-of-year administration occurred from April 25th through May 5th. Four participants were assessed at the beginning of the year between August 26th and October 6th due to late enrollment. First-time test administrators attended a full training course before the first administration. Experienced test administrators received a brief refresher training course. Administration procedures were reviewed before each testing window. All subtests included written directions to be read orally by the test administrator. These included practice items and reminders to the participant. The Acadience subtests were administered and scored digitally according to the procedures detailed in the Acadience Reading Testing Manual (Good & Kaminsky, 2020).

Statistical Methods

For each participant, an RCS was calculated for each test administration. The RCS was used to answer the research questions. To answer research question one, which pertained to achievement scores and growth rates among groups, a one-way ANOVA was used to determine whether there was a statistically significant difference among the identified, at-risk, and typical groups at each grade level based on the end-of-year RCS. The one-way ANOVA was selected based on its utility in comparing group means and controlling for type I error (false positives). Pairwise comparisons were conducted to determine the significance between groups. To measure growth, RCS could not be compared between test administration periods or between grade levels because different subtests are administered. Therefore, to answer this part of question one, growth was measured by the pathway of progress. The pathways are determined by a participant's beginning and end-of-year RCS. Percentiles are calculated from all students with the same beginning-of-year score (Good & Stollar, n.d.). Categories include "well below typical progress," "below typical progress," "typical progress," "above typical progress," and "well

above typical progress." The categories were assigned numerical values 1-5 corresponding from least to greatest growth for statistical calculation. Categories, their numerical values, and the corresponding progress percentile ranges are defined in Table 5. The Kruskal-Wallis test was used to determine whether there was a statistically significant difference among the identified, at-risk, and typical groups at each grade level based on their pathway of progress. Pairwise comparisons were conducted to determine the significance between groups.

Table 5

Pathways of Progress for Growth and the Corresponding Values

Numerical Value	Pathway of Progress	Percentile Range (Good & Stollar, n.d.)
1	Well below typical progress	Below 20th percentile
2	Below typical progress	20th to 39th percentile
3	Typical progress	40th to 59th percentile
4	Above typical progress	60th to 79th percentile
5	Well above typical progress	80th percentile and above

To answer research question two, which pertains to the level of agreement among three methods of identifying at-risk readers, participants were identified as LA, LG, DD, and no risk. For the purposes of this study, LA was defined as an RCS at or below the sixteenth percentile based on published norms. The sixteenth percentile was selected as the cut score because participants at and below the sixteenth percentile fall one or more standard deviations (SD) below the mean. Milburn et al. (2017) used three different cut points (i.e., 5th, 10th, and 25th percentiles), and Milburn et al. (2018) used the 25th percentile as a cut score. Participants performing at the 25th percentile fall within the average range. On the other hand, use of the 5th or 10th percentiles may under-identify at-risk students. LG was defined as a pathway of progress well below typical progress. DD was defined as meeting the criteria for both LA and LG. Cohen's kappa coefficient was calculated to determine the level of agreement of each of the

three classification methods. Cohen's kappa was selected based on its utility for comparing binary ratings while also accounting for chance agreement.

To answer research question three, a single-sample *t*-test was conducted to compare the end-of-year RCS for each grade level to published norms to determine if reading achievement in Title I schools was significantly different from published norms. To determine whether reading growth was significantly different from published norms, a single-sample t-test was conducted using the numerical values corresponding to the pathways of progress. Single-sample t-tests were selected for both achievement scores and growth rates based on the utility of comparing participants' scores to a published mean.

To calculate local norms for reading achievement, participants' RCS were converted to a percentile based on the data for all participants at each grade level. These percentiles based on local data were used for the LA method. To compare published and local norms, participants were identified as LA using the same criteria as question two using local norms. The percentage of participants who met the criteria for LA was calculated based on local and published norms. Cohen's kappa coefficient was calculated to determine the level of agreement in classification based on local and published norms for the LA method. Because the pathways of progress for growth are specific to published norms, growth cannot easily be calculated by school districts based on local norms. Therefore, the LG method could not be applied to local data. To explore the possibility of using a combination of norms for the DD method, participants were classified as DD based on LA local norms and LG published norms. Cohen's kappa coefficient was calculated to determine the level of agreement in DD classification using only published norms (for both LA and LG) and a combination of published and local norms (local norms for LA and published norms for LG).

Chapter IV

RESULTS

The purpose of this study was to analyze three methods of classifying at-risk readers using a commonly used universal assessment, Acadience Reading. Additionally, the use of local and published norms was compared.

Reading Achievement and Growth in Typical, At-Risk, and Identified Groups

Research question one pertained to reading achievement scores and growth rates of kindergarten through third-grade participants with identified reading and reading-related disabilities, students at risk of reading disabilities, and typical readers. To address this question, reading achievement and growth were compared among typical readers, at-risk readers, and readers with identified disabilities. Means and standard deviations of reading achievement scores and growth rates can be found in Table 6.

Table 6

Achievement and Growth Score Means and Standard Deviations

Grade	Group	Number of Participants	RCS *M (SD)*	RCS Corresponding Percentile	Growth *M* (*SD*)
Kindergarten	Typical	95	121 (38)	30	2.43 (1.23)
	At-risk	26	85 (37)	11	1.69 (0.93)
	Identified	7	39 (20)	3	1 (0)
	Total	128	109 (42)	21	2.20 (1.46)
First grade	Typical	88	161 (82)	38	3.66 (1.14)
	At-risk	35	61 (58)	14	3.09 (1.25)
	Identified	7	18 (28)	6	2 (1)
	Total	130	126 (90)	27	3.42 (1.50)
Second grade	Typical	80	235 (87)	34	2.46 (1.28)
	At-risk	33	87 (92)	10	2.12 (1.19)
	Identified	9	73 (96)	9	1.89 (0.93)
	Total	122	183 (114)	21	2.33 (1.53)
Third grade	Typical	90	291 (107)	24	2.22 (1.31)
	At-risk	36	129 (94)	7	2.14 (1.15)
	Identified	8	55 (58)	3	2.25 (1.16)
	Total	134	234 (131)	14	2.20 (1.56)

A one-way ANOVA was used to determine the significance among groups. Reading achievement scores were significantly different among groups in all grade levels (kindergarten: $F = 23.19$, $p < .001$; first: $F = 30.52$, $p < .05$; second: $F = 39.60$, $p < .05$; third: $F = 46.02$, $p < .05$). Pairwise comparisons (post hoc Tukey HSD; Table 7) were significant in all grade levels between typical and at-risk groups and between typical and identified groups. The difference between at-risk and identified groups was only significant for participants in kindergarten.

Table 7

Pairwise Comparisons for Reading Achievement Between Groups

Grade	Typical vs. At-Risk	Typical vs. Identified	At-Risk vs. Identified
Kindergarten	*$Q = 3.86$, $p = .020$	*$Q = 8.80$, $p < .001$	*$Q = 4.94$, $p = .002$
First Grade	*$Q = 5.44$, $p < .001$	*$Q = 7.81$, $p < .001$	$Q = 2.37$, $p = .219$
Second Grade	*$Q = 7.34$, $p < .001$	*$Q = 8.00$, $p < .001$	$Q = 0.67$, $p = .884$
Third Grade	*$Q = 6.83$, $p < .001$	*$Q = 9.97$, $p < .001$	$Q = 3.14$, $p = .071$

*Indicates significance at $p < .05$.

RCS could not be compared among grade levels because different subtests were administered. Instead, mean RCS were converted to percentiles based on published norms for each grade level. These percentiles indicate that participants in first grade had the highest reading ability relative to published norms (27th percentile), and participants in third grade had the lowest reading ability relative to published norms (14th percentile). For each grade level, percentiles for kindergarten through second grade remained relatively stable (21st percentile – 27th percentile), but the percentile for third grade dropped to the 14th percentile. Additionally, each third-grade group had corresponding percentiles less than or equal to those same groups in other grade levels. For example, the third-grade at-risk group had a lower percentile than all other at-risk groups in different grade levels. Mean RCS for all groups at all grade levels fell below the average score. RCS for all typical groups fell within the average range, but RCS for all at-risk and identified groups fell below the average range.

The Kruskal-Wallis test revealed that growth classification was significantly different among groups in kindergarten ($H = 15.89$, $p < .001$) and first grade ($H = 12.94$, $p = .002$; Table 8). Pairwise comparisons (post hoc Mann Whitney U Test) indicated that typical readers in kindergarten and first grade demonstrated significantly more growth than at-risk and identified readers. Additionally, at-risk first-grade readers demonstrated significantly more growth than identified readers.

According to the Kruskal-Wallis test, growth classification was not significantly different among second-grade groups ($H = 2.47$, $p = .290$) nor third-grade groups ($H = 0.07$, $p = .968$). Although there was no significant difference between groups in second grade and third grade, typical readers in second grade demonstrated more growth than at-risk readers, and at-risk readers demonstrated more growth than identified readers. In third grade, the typical group

outperformed the at-risk group, but the identified group demonstrated more growth than the typical and at-risk groups.

Table 8

Pairwise Comparisons for Reading Growth Between Groups

Grade	Typical vs. At-Risk	Typical vs. Identified	At-Risk vs. Identified
Kindergarten	*z = 2.71, p = .007	*z = 3.18, p = .001	z = 1.67, p = .095
First Grade	*z = 2.23, p = .026	*z = 3.06, p = .002	*z = 1.97, p = .049

*Indicates significance at $p < .05$.

Scoring of growth is consistent across grades and can be compared. First-grade participants demonstrated the most growth (M = 3.42, SD = 1.50). Participants in kindergarten (M = 2.20, SD = 1.46) and third grade (M = 2.20, SD = 1.56) demonstrated the least growth. The typical and at-risk groups in first grade were the only groups with a growth score classification greater than three, which is considered average. All other groups had mean growth classifications less than average.

Agreement of Risk Classification Methods

Research question two pertained to the agreement among three methods of classifying risk status (LA, LG, and DD) using published norms. To address this question, the risk classification methods were applied to participants' end-of-year RCS and growth rate classifications. Use of the LA method resulted in the highest number of at-risk readers, while use of the DD method resulted in the lowest number of at-risk readers (Table 9). Use of the LA method identified 33.8%-39.3% of participants at each grade level with a range of 5.5% and average of 36.4%. Use of the LG method identified 6.9%-40.3% of participants at each grade level with a range of 33.4% and average of 30.4%. Use of the DD method identified 6.2%-32.8% of participants at each grade level with a range of 26.6% and average of 20.8%.

Table 9

Participants Identified Using Each Classification Method (Published Norms)

Grade Level	LA	LG	DD	No Risk
Kindergarten	46 (35.9%)	48 (37.5%)	42 (32.8%)	76 (59.4%)
First	44 (33.8%)	9 (6.9%)	8 (6.2%)	85 (65.4%)
Second	48 (39.3%)	45 (36.9%)	30 (24.6%)	59 (48.4%)
Third	49 (36.6%)	54 (40.3%)	27 (20.1%)	38 (28.4%)
Total	187 (36.4%)	156 (30.4%)	107 (20.8%)	258 (50.2%)

Cohen's kappa coefficients indicated fair agreement between LA and LG methods [κ = -0.319 (0.039)]. There was moderate agreement between LA and DD methods [κ = -0.482 (0.038)] and between LG and DD methods [κ = -0.576 (0.046)]. The agreement between the DD identification method and each of the other two methods was higher than the agreement between LA and LG. Even though growth was not significantly different among groups for second and third grade, the overall agreement between DD and LG was higher than the level of agreement between DD and LA.

Local Versus Published Norms

Research question three pertained to the use of local norms instead of published norms to identify at-risk readers. To address this question, risk identification methods were applied based on local norms rather than published norms. Reading achievement scores were compared to published norms using a single sample *t*-test. This revealed that achievement scores were significantly different from published norms at all grade levels (kindergarten: $t = -8.88$, $p < .001$; first: $t = -8.52$, $p < .001$; second: $t = -8.77$, $p < .001$; third: $t = -12.97$, $p < .001$). RCS means for all groups at each grade level fell below the published average scores (Table 6).

Reading growth rate classifications were compared to published norms using a single-sample *t*-test. This revealed that growth classifications were also significantly different than published norms at all grade levels (kindergarten: $t = -7.44$, $p < .001$; first: $t = 3.85$, $p < .001$;

second: $t = -5.97$, $p < .001$; third: $t = -7.36$, $p < .001$). Growth classification averages for all groups at each grade level fell below the published average score except for typical and at-risk readers in first grade (Table 6).

Because reading achievement and growth were significantly different from the published norms, the LA and DD methods were applied to local achievement norms. Cohen's kappa coefficients indicated fair agreement between the LA method based on published norms versus local norms [$\kappa = -0.35$ (0.035)] and between the DD method based on published norms versus local norms [$\kappa = -0.22$(0.031)].

Use of the LA method resulted in the highest number of at-risk readers, while use of the DD method resulted in the lowest number of at-risk readers (Table 10). Use of the LA method identified 15.7%-17.2% of participants at each grade level with a range of 1.5% and average of 16.5%. Use of the DD method identified 3.8%-17.2% of participants at each grade level with a range of 13.4% and average of 10.1%.

Fewer than half the number of students were identified as at-risk based on local norms compared to published norms for both LA and DD methods (Table 10). Around 36% of participants were identified as at-risk readers using the LA method based on published norms, while only around 16% of participants were identified using local norms. The same is true for the DD method. Around 20% of participants were identified using the DD method with published norms, but only about 10% were identified using local norms.

Table 10

Participants Identified Using Each Classification Method (Local Norms)

Grade Level	LA	DD	No Risk
Kindergarten	22 (17.2%)	22 (17.2%)	80 (62.5%)
First	21 (16.2%)	5 (3.8%)	105 (80.8%)
Second	21 (17.2%)	16 (13.1%)	72 (59%)
Third	21 (15.7%)	9 (6.7%)	67 (50%)
Total	85 (16.5%)	52 (10.1%)	324 (63%)

Chapter V

CONCLUSIONS, INTERPRETATIONS, AND RECOMMENDATIONS

Previous research on the identification of at-risk readers has yielded inconsistent results regarding the most appropriate way to determine risk (Al Otaiba et al., 2011; Hendricks & Fuchs, 2020; Maki et al., 2020; Milburn et al., 2017; Schatschneider et al., 2008; Tindal et al., 2016; Van Norman & Nelson, 2021). The purpose of this study was to analyze the use of three methods of identifying at-risk readers through the RTI model by applying these methods to a universal screening measure.

Significance of Achievement and Growth Scores

The first research question examined the difference in reading achievement and growth rates of participants with typical reading ability, those at risk for reading disabilities, and those with identified reading and reading-related disabilities. The lack of significance in RCS between identified and at-risk groups in first through third grades indicates that reading ability in these two groups is similar. One reason for this may be a delay in referring at-risk readers for a special education evaluation. According to the Georgia Department of Education (n.d.a), a minimum of 12 weeks of data from RTI interventions is required before a student may receive a special education evaluation and be found eligible for a reading-related specific learning disability. Depending on each school district's policies, this time may be even longer. This delay may have resulted in the incorrect grouping of some participants as at-risk instead of identified.

There is less of a delay before students can receive RTI interventions because they can begin as soon as a teacher identifies them. Because the at-risk students were identified by their teachers or other educators, there is a likely chance that many of these participants have a reading disorder and have not yet been evaluated. Two studies on the accuracy of teacher

identification of at-risk readers showed a moderate to strong correlation between teacher identification and actual reading scores (Feinberg & Shapiro, 2009; Martin & Shapiro, 2011). The significant difference in reading achievement between typical readers and those with any level of risk indicates that the methods used to identify at-risk readers (i.e., RTI and teacher referral) in these Title I schools have some level of agreement with Acadience Reading scores.

There were significant differences in growth rate classification between the typical group and each of the other two groups in kindergarten and first grade which suggests that growth may be an appropriate method of identifying reading-related risk in earlier grades. However, the lack of significance in growth rate classification between groups for second- and third-grade participants suggests that reading-related growth is not appropriate for identifying at-risk readers at these grade levels. This is evidenced by the fact that third-grade participants with identified reading and reading-related disorders demonstrated more growth than typical readers and at-risk readers. This has implications for the results of the remaining research questions.

Risk Classification Methods

The second research question was related to three methods of identifying at-risk readers based on their scores on the Acadience Reading diagnostic. The LA method had the least variability in the number of at-risk students identified at each grade level, while the LG method had the highest variability. This may be explained by the results of research question one. Because growth was not significantly different among groups, growth rate is much more variable across groups than reading achievement scores which may have resulted in greater variability of at-risk readers across grade levels.

Overall, there was fair to moderate agreement among the three risk identification methods which is consistent with the weak to moderate agreement among methods found in

Milburn et al.'s study (2017). Higher levels of agreement between DD and the other two methods compared to the agreement between LA and LG may be due to overlapping criteria. Participants who meet the criteria for DD must, by definition, also meet the criteria for LA and LG. However, this is not true for LA and LG, as these have separate criteria.

The question remains as to which of these three methods is most appropriate for identifying at-risk readers in Title I schools. Based on the lack of significant differences in growth rates in second and third-grade participants, the LG method is not appropriate for use with these grade levels. This finding is consistent with current literature on the relationship between growth rates and reading ability (Al Otaiba et al., 2011; Maki et al., 2020; Schatschneider et al., 2008; Tindal et al., 2016; Van Norman & Nelson, 2021). The most permissive method of identifying at-risk readers would be to support students who meet the criteria for either LA or LG. The LA method can be used to identify students who are currently performing below benchmark. In contrast, the LG method can be used to identify students who may be performing on grade level currently but are at risk of falling behind if their reading-related growth does not improve. However, providing intervention to this number of students, which is about 50% of participants in this study, is more appropriate for tier-one instruction. According to the widely used RTI pyramid, which provides recommended percentages of students who receive instruction at each tier (Georgia Department of Education, n.d.b), approximately 80% of students should have their needs met by tier-one instruction. The use of this method would result in half of the student population receiving RTI interventions. While each of these students is likely to benefit from the support, this number of students cannot be feasibly served in RTI tiers two and three.

Overall growth in first-grade participants fell significantly above published norms. This verifies that reading-related growth cannot be the only indicator of reading-related risk for several reasons. First, the mean RCS for all first-grade participants corresponds to the 37th percentile and falls significantly below the published average. Although first-grade participants demonstrated above-average growth, their reading achievement scores were below average. Additionally, there was a lack of significance in growth rate classification between second and third-grade groups. This lack of significance makes growth rate classification a poor indicator of reading ability at these grade levels. Finally, third-grade participants with identified reading disabilities had a better growth rate than typical and at-risk readers in third grade. The sole use of growth rate to identify at-risk readers in this grade level would exclude students with identified reading disabilities. These circumstances illustrate the risks of using LG as the only method of identifying at-risk readers.

Alternatively, sole use of the LA method would result in the identification of approximately 36% of participants. The identification of the highest number of participants using the LA method is consistent with Milburn et al.'s findings (2017). While providing an intervention to 36% of students is much more reasonable than the 50% of students who were classified as either LA or LG, Title I schools may not have the resources to provide evidence-based interventions to this number of students.

The most conservative method would be to only provide intervention for students who meet the criteria for DD. This includes approximately 20% of participants in this study. Of course, these students are at the greatest risk for a reading disability, but this method excludes low achievers with relatively good growth who need reading support. Additionally, growth rate,

which is part of the DD criteria, was not a reliable indicator of risk for second and third-grade participants.

There are costs and benefits associated with each of these three risk-classification methods. All three of these methods were applied based on the norms published with the Acadience Reading assessment. For schools with a large proportion of minority or underprivileged students, including Title I schools, the use of local norms offers an alternative method of identifying at-risk students.

The Costs and Benefits of Using Local Versus Published Norms

The third and final research question examined the use of local normative data instead of published norms for the three risk classification methods. Using local norms allows educators to compare students' performance to their peers rather than to the national normative sample. Except for growth rate classification for first-grade participants, participants in all grade levels and groups had achievement scores and growth rate classifications significantly below the published norms.

Regarding reading achievement scores, RCS means fell below the published averages for each grade level, even for participants identified as typical readers by their teachers. This discrepancy between reading ability in typical and underprivileged populations is well-documented in research (Buckingham et al., 2013). Additionally, the reading gap between national averages and underprivileged populations increases across grade levels. This is evidenced by the relative decrease in reading ability for third-grade participants in all groups compared to published norms.

When using local norms to apply the risk identification methods, fewer than half the number of participants were identified as at-risk compared to using published norms. The lack of

agreement between identification methods based on published norms and those based on local norms indicates a need for caution when selecting norms for identifying at-risk readers.

Conclusions

The RTI pyramid cannot be used prescriptively to determine what percentage of students in Title I schools have reading disabilities. However, the pyramid can be used to guide the types of interventions. Providing quality tier-one instruction to 80% of students, even those at low risk of reading disability, is the most feasible way to address widespread reading deficits. If tier-one instruction is designed to meet the needs of 80% of students, the remaining students can receive intervention in tiers two and three. Although Title I populations typically perform below the national average (Buckingham et al., 2013), many of these students' needs can be addressed through quality tier-one instruction.

Identifying At-Risk Readers

After receiving tier one instruction, many students will still need further instruction in tiers two and three of RTI. To identify which students would benefit from more intensive intervention, different methods should be used based on students' grade levels. For students in kindergarten and first grade, the DD method is recommended. These students are still in the beginning stages of learning to read when sufficient growth is crucial for their success. Based on the results of this study, both achievement scores and growth rates were significant indicators. Due to the nature of reading profiles in underprivileged students, it is recommended that the LA criteria be based on local norms. Because growth cannot be easily calculated based on local norms, LG criteria can be based on published norms. Students in kindergarten and first grade who fall at or below the 16th percentile based on local norms and have a growth classification of "well below average" would be eligible for more intensive interventions based on this approach.

In the current study, growth rates in second and third-grade participants were not a reliable factor of reading ability and should not be used as a criterion for receiving further intervention. The LA identification method based on local norms is recommended for second- and third-grade students. The LA method can be used to identify all students who are performing below the expected level. Additionally, the use of local norms allows school districts to focus their resources on a smaller number of students. For districts with available resources, the LA criteria should be based on published norms. This allows more students to receive the needed interventions and potentially improve their reading ability.

Instructional Recommendations

Using the recommended methods to identify at-risk readers will exclude many students who need reading intervention, especially if using local norms. To ensure that all students receive quality instruction, regardless of whether they were identified as at-risk readers, it is recommended that tier I instruction is designed to meet the needs of most students. To do this, instruction should be based on structured literacy practices. These include instruction that is explicit, systematic, and sequential; instruction at multiple levels (from phonemes to text structure); cumulative practice and ongoing review; student-teacher interaction; examples and non-examples; decodable text; and feedback (Spear-Swerling, 2018). The structured literacy approach is supported by the science of reading (Kuo, 2023).

Future Research

Future research should address how the widening gap in reading achievement as students age impacts the use of various methods of identifying at-risk readers across grade levels. Additionally, future research may address the current study's limitations by including students placed in full-day resource classrooms. Students are placed in this setting for a variety of reasons

which may or may not impact their reading ability. Therefore, including them in the research allows educators to apply reading-specific risk identification methods to this subset of the student population. Finally, including a measure of RAN in the universal screening measure may result in more accurate identification of at-risk readers. Due to the crucial role of RAN in reading ability, the inclusion of this measure may aid in the identification of at-risk readers with different reading profiles.

REFERENCES

Al Otaiba, S., Folsom, J. S., Schatschneider, C., Wanzek, J., Greulich, L., Meadows, J., Li, Z., & Connor, C. M. (2011). Predicting first grade reading performance from kindergarten response to tier 1 instruction. *Exceptional Children, 77*(4), 453-470. https://doi.org/10.1177/001440291107700405

Amadi, E. A. (2019). Bottom-up theory and phonics instruction: Implications for beginning reading. *European Journal of Applied Linguistics Studies, 1*(2), 89–100. http://dx.doi.org/10.5281/zenodo.3228773

Anderson, N. L., & Kaye, E. L. (2016). Finding versus fixing: Self-monitoring for readers who struggle. *The reading teacher, 70*(5), 543-550. https://doi.org/10.1002/trtr.1552

Aravena, S., Snellings, P., Tijms, J., & van der Molen, M. W. (2013). A lab-controlled simulation of a letter-speech sound binding deficit in dyslexia. *Journal of Experimental Child Psychology, 115*(4), 691-707. https://doi.org/10.1016/j.jecp.2013.03.009

Benjamin, R. G., & Schwanenflugel, P. J. (2010). Text complexity and oral reading prosody in young readers. *Reading Research Quarterly, 45*(4), 388-404. https://dx.doi.org/10.1598/RRQ.45.4.2

Bergeron, B. S. (1990). What does the term whole language mean? Constructing a definition from the literature. *Journal of Reading Behavior, 22*(4), 301-329. https://journals.sagepub.com/doi/pdf/10.1080/10862969009547716

Bitter, C., O'Day, J., Gubbins, P., & Socias, M. (2009). What works to improve student literacy achievement? An examination of instructional practices in a balanced literacy approach.

Journal of Education for Students Placed at Risk, 14(1), 17-44. https://doi.org/10.1080/10824660802715403

Blythe, H. I., Ascensión, P., & Dodd, M. (2015). Beyond decoding: Phonological processing during silent reading in beginning readers. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 41*(4), 1244-1252. http://dx.doi.org/10.1037/xlm0000080

Bowers, J. S. (2020). Reconsidering the evidence that systematic phonics is more effective than alternative methods of reading instruction. *Educational Psychology Review, 32*, 681–705. https://doi.org/10.1007/s10648-019-09515-y

Brandenburg, J., Klesczewski, J., Schuchardt, K., Fischbach, A., Büttner, G., & Hasselhorn, M. (2017). Phonological processing in children with specific reading disorder versus typical learners: Factor structure and measurement invariance in a transparent orthography. *Journal of Educational Psychology, 109*(5), 709–726. https://doi.org/10.1037/edu0000162

Brooks, G. (2023). Disputing recent attempts to reject the evidence in favor of systematic phonics instruction. *Review of Education, 11*(2). https://doi.org/10.1002/rev3.3408

Buckingham, J., Wheldall, K., & Beaman-Wheldall, R. (2013). Why poor children are more likely to become poor readers: The school years. *Australian Journal of Education, 57*(3), 190–213. https://doi.org/10.1177/0004944113495500

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*(2), 278-293. https://doi.org/10.1044/1092-4388(2006/023)

Catts, H. W., Bridges, M. S., Little, T. D., & Tomblin, J. B. (2008). Reading achievement growth

in children with language impairments. *Journal of Speech, Language, and Hearing Research, 51*(6), 1569–1579. https://doi.org/10.1044/1092-4388(2008/07-0259)

Catts, H. W., Fey, M. E., Tomblin, J. B., & Zhang, X. (2002). A longitudinal investigation of reading outcomes in children with language impairments. *Journal of Speech, Language, and Hearing Research, 45*(6), 1142–1157. https://doi.org/10.1044/1092-4388(2002/093)

Choi, D., Hatcher, R. C., Dulong-Langley, S., Liu, X., Bray, M. A., Courville, T., O'Brien, R., & DeBiase, E. (2016). What do phonological processing errors tell about students' skills in reading, writing, and oral language? *Journal of Psychoeducational Assessment, 35*(1-2), 24-46. https://doi.org/10.1177/0734282916669018

Clayton, F. J., & Hulme, C. (2018). Automatic activation of sounds by letters occurs early in development but is not impaired in children with dyslexia. *Scientific Studies of Reading, 22*(2), 137–151. https://doi.org/10.1080/10888438.2017.1390754

Clayton, F. J., West, G., Sears, C., Hulme, C., & Lervag, A. (2020). A longitudinal study of early reading development: Letter-sound knowledge, phoneme awareness, and RAN, but not letter-sound integration, predict variations in reading development. *Scientific Studies of Reading, 24*(2), 91–107. https://doi.org/10.1080/10888438.2019.1622546

Cunningham, A. J., Burgess, A. P., Witton, C., Talcott, J. B., & Shapiro, L. R. (2020). Dynamic relationships between phonological memory and reading: A five-year longitudinal study from age 4 to 9. *Developmental Science, 24*(1). https://doi.org/10.1111/desc.12986

Denton, C. A., Tolar, T. D., Fletcher, J. M., Barth, A. E., Vaughn, S., & Francis, D. J. (2013). Effects of a tier 3 intervention for students with persistent reading difficulties and characteristics of inadequate responders. *Journal of Educational Psychology, 105*(3), 633–648. https://doi.org/10.1037/a0032581

Dever, B. V., Raines, T. C., Dowdy, E., & Hostutler, C. (2016). Addressing disproportionality in special education using a universal screening approach. *The Journal of Negro Education, 85*(1), 59–71. https://doi.org/10.7709/jnegroeducation.85.1.0059

Dowd, A. J., & Bartlett, L. (2019). The need for speed: Interrogating the dominance of oral reading fluency in international reading efforts. *Comparative Education Review, 63*(2), 189–212. https://doi.org/10.1086/702612

Dowhower, S. L. (1991). Speaking of prosody: Fluency's unattended bedfellow. *Theory Into Practice, 30*(3), 165-175. https://doi.org/10.1080/00405849109543497

FastBridge Learning. (2018). FAST technical manual. Fast-Bridge Learning, LLC. http://support-content.fastbridge.org/FAST_Research/FAST_Technical_Manual_Version_FINAL.pdf

Fedora, P. M. (2016). Single and double deficits in early readers in rural, low-wealth communities. *Reading and Writing Quarterly, 32*(2), 101-126. https://doi.org/10.1080/10573569.2013.866529

Feinberg, A. B., & Shapiro, E. S. (2009). Teacher accuracy: An examination of teacher-based judgments of students' reading with differing achievement levels. *The Journal of Educational Research, 102*(6), 453–462. https://doi.org/10.3200/JOER.102.6.453-462

Field, S. A., Begeny, J. C., & Kim, E. K. (2019). Exploring the relationship between cognitive characteristics and responsivity to a tier 3 reading fluency intervention. *Reading and Writing Quarterly, 35(*4), 374-391. https://doi.org/10.1080/10573569.2018.1553082

Fletcher, J. M., Savage, R., Vaughn, S. (2021). A commentary on Bowers (2020) and the role of phonics instruction in reading. *Educational Psychology Review, 33*(3), 1249-1274. https://doi.org/10.1007/s10648-020-09580-8

Frey, B. B., Lee, S. W., Tollefson, N., Pass, L., & Massengil, D. (2005). Balanced literacy in an urban school district. *Journal of Educational Research, 98*(5), 272-280. http://www.heldref.org

Georgia Department of Education. (n.d.a). *Eligibility determination and categories of eligibility*. https://www.gadoe.org/External-Affairs-and-Policy/State-Board-of-Education/SBOE%20Rules/160-4-7-.05.pdf

Georgia Department of Education. (n.d.b) *MTSS snapshot: The power of the pyramid*. https://www.gadoe.org/wholechild/Documents/MTSS/SnapshotPowerPyramid.pdf

Georgia Department of Education. (2015). *Georgia standards of excellence: English language arts (ELA).* https://www.georgiastandards.org/Georgia-Standards/Documents/ELA-Standards-Grades-K-5.pdf

Gillingham, A., & Stillman, B. W. (1997). *The Gillingham manual*. Educators Publishing Service.

Glavach, M., & Pribyl, W. (Spring/Summer 2018). A whole language reading intervention: A case study. *Journal of the American Academy of Special Education Professionals*, 45-55. https://eric.ed.gov/?id=EJ1254613

Good, R. H., & Kaminski, R. A. (2020). *Acadience Reading K-6: Assessment manual*. Acadience Learning, Inc. https://acadiencelearning.org/wp-content/uploads/2020/08/AcadienceReading_Assessment_Manual.pdf

Good, R. H., & Stollar, S. A. (n.d.). *Improving reading outcomes through goal setting and progress monitoring* [PowerPoint slides]. Dynamic Measurement Group, Acadience Learning. https://acadiencelearning.org/wp-content/uploads/2020/01/Handout_ImprovingReadingOutcomes_2019.pdf

Good, R. H., Kaminsky, R. A., Dewey, E. N., Wallin, J., Powell-Smith, K. A., & Latimer, R. (2019). *Acadience Reading K-6: Technical manual*. Dynamic Measurement Group, Inc. https://acadiencelearning.org/wp-content/uploads/2020/01/Acadience_Reading_K-6_Technical_Manual.pdf

Gough, P. B. (1972). One second of reading. In J. F. Kavanagh, & I. G. Mattingly (Eds.), *Language by ear and by eye*. MIT Press.

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

Guo, L., Wang, J., Lee, J., & Lesley, M. (2023). Examining the differentiated impacts of balanced literacy: An analysis of reading comprehension skills. *Reading & Writing Quarterly: Overcoming Learning Difficulties*. http://dx.doi.org/10.1080/10573569.2023.2186293

Hasbrouck, J., & Tindal, G. (2017). *An update to compiled ORF norms*. Behavioral Research and Teaching. https://files.eric.ed.gov/fulltext/ED594994.pdf

Hendricks, E. L., & Fuchs, D. (2020). Are individual differences in response to intervention influenced by the methods and measures used to define response? Implications for identifying children with learning disabilities. *Journal of Learning Disabilities, 53*(6), 428-443. https://doi.org/10.1177/0022219420920379

Hulme, C., Bowyer-Crane, C., Carroll, J. M., Duff, F. J., & Snowling, M. J. (2012). The causal role of phoneme awareness and letter-sound knowledge in learning to read: Combining intervention studies with mediation analyses. *Psychological Science, 23*(6), 572–577. https://doi.org/10.1177/0956797611435921

Kibby, M. Y., Lee, S. E., & Dyer, S. M. (2014). Reading performance is predicted by more than

phonological processing. *Frontiers in Psychology, 5*. https://doi.org/10.3389/fpsyg.2014.00960

Kim, Y. S. G., Quinn, J. M., & Petscher, Y. (2021). What is text reading fluency and is it a predictor or outcome of reading comprehension? A longitudinal investigation. *Developmental Psychology, 57*(5), 718-732. https://doi.org/10.1037/dev0001167

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

Kuo, N. (2023). Instructing students with dyslexia through structured literacy. *International Journal of Education Technology and Science, 3*(1), 110-124. https://globets.org/journal/index.php/IJETS/article/view/97

Lindamood, P. C., & Lindamood, P. D. (2011). *The Lindamood phoneme sequencing program for reading, spelling, and speech* (4th ed.). Pearson Education.

Loucas, T., Baird, G., Simonoff, E., & Slonims, V. (2016). Phonological processing in children with specific language impairment with and without reading difficulties. *International Journal of Language and Communication Disorders, 51*(5), 581-588. https://doi.org/10.1111/1460-6984.12225

Lyon, G. (1998). Why reading is not a natural process. *Educational Leadership, 55*(6), 14–18.

Maddox, K., & Feng, J. (2013, October 18). *Whole language instruction vs. phonics instruction: Effect on reading fluency and spelling accuracy of first grade students*. Georgia Educational Research Association Annual Conference. Savannah, Georgia, United States. https://eric.ed.gov/?id=ED545621

Maki, K. E., Barrett, C. A., Hajovsky, D. B., & Burns, M. K. (2020). An examination of the

relationships between specific learning disabilities identification and growth rate, achievement, cognitive ability, and student demographics. *School Psychology, 35*(5), 343–352. https://doi.org/10.1037/spq0000400

Manning, M. & Kamii, C. (2000). Whole language vs. isolated phonics instruction: A longitudinal study in kindergarten with reading and writing tasks. *Journal of Research in Childhood Education, 15*(1), 53-65. https://www.proquest.com/scholarly-journals/whole-language-vs-isolated-phonics-instruction/docview/203887710/se-2

Martin, S. D., & Shapiro, E. S. (2011). Examining the accuracy of teachers' judgments of DIBELS performance. *Psychology in the Schools, 48*(4), 343–356. https://doi.org/10.1002/pits.20558

McArthur, G., & Castles, A. (2013). Phonological processing deficits in specific reading disability and specific language impairment: Same or different? *Journal of Research in Reading, 36*(3), 280–302. https://doi.org/10.1111/j.1467-9817.2011.01503.x

Milburn, T. F., Lonigan, C. J., & Phillips, B. M. (2018). Determining responsiveness to tier 2 intervention in response to intervention: Level of performance, growth, or both. *The Elementary School Journal, 118*(2), 310-334. https://doi.org/10.1086/694271

Milburn, T. F., Lonigan, C. J., Allan, D. M., & Phillips, B. M. (2017). Agreement among traditional and RTI-based definitions of reading-related learning disability with preschool children. *Learning and Individual Differences, 55*, 120–129. https://doi.org/10.1016/j.lindif.2017.03.011

Moats, L., & Tolman, C. (2009). Excerpted from Language essentials for teachers of reading and spelling (LETRS): The challenge of learning to read (Module 1). Sopris West.

Morris, D., Trathen, W., Frye, E. M., Kucan, L., Ward, D., Schlagal, R., & Hendrix, M. (2013).

The role of reading rate in the informal assessment of reading ability. *Literacy Research and Instruction, 52*(1), 52–64. http://dx.doi.org/10.1080/19388071.2012.702188

National Center on Response to Intervention. (n.d.). *What is RTI?* American Institutes for Research. https://files.eric.ed.gov/fulltext/ED526859.pdf

National Reading Panel (U.S). (2000). *Report of the National Reading Panel: Teaching children to read: an evidence-based assessment of the scientific research literature on reading and its implications for reading instruction*. U.S. Dept. of Health and Human Services, Public Health Service, National Institutes of Health, National Institute of Child Health and Human Development. https://www.nichd.nih.gov/sites/default/files/publications/pubs/nrp/Documents/report.pdf

Nelson, J. M., Lindstrom, J. H., Lindstrom, W., & Denis, D. (2012). The structure of phonological processing and its relationship to basic reading. *Exceptionality, 20*(3), 179-196. https://doi.org/10.1080/09362835.2012.694612

O'Connor, R. E., Bocian, K. M., Beach, K. D., Sanchez, V., & Flynn, L. J. (2013). Special education in a 4-year response to intervention (RTI) environment: Characteristics of students with learning disability and grade of identification. *Learning Disabilities Research & Practice, 28*(3), 98–112. https://doi.org/10.1111/ldrp.12013

Office of Special Education Programs. (2017). *IDEA part B regulations: Significant disproportionality.* U.S. Department of Education. https://sites.ed.gov/idea/files/significant-disproportionality-qa-03-08-17.pdf

Ozernov-Palchik, O., Norton, E. S., Sideridis, G., Beach, S. D., Wolf, M., Gabrieli, J. D. E., & Gaab, N. (2017). Longitudinal stability of pre-reading skill profiles of kindergarten children: Implications for early screening and theories of reading. *Developmental*

Science, 20(5). https://doi.org/10.1111/desc.12471

Papadopoulos, T. C., Spanoudis, G. C., Chatzoudi, D. (2020). A longitudinal investigation of the double dissociation between reading and spelling deficits: The role of linguistic and executive function skills. *Reading and Writing*, 33(4), 1075-1104. https://doi.org/10.1007/s11145-020-10029-1

Pearson Education (2011). AIMSweb default cut scores explained. State prediction user's guide. https://www.aimsweb.com/wp-content/uploads/AIMSweb_Default_Cut_Score_Guide.pdf

Pearson, P. D. (2004). The reading wars. *Politics of Education Association, 18*(1), 216–252. https://doi.org/10.1177/0895904803260041

Protopapas, A., Katopodi, K., Altani, A., & Georgiou, G. K. (2018). Word reading fluency as a serial naming task. *Scientific Studies of Reading, 22*(3), 248-263. https://psycnet.apa.org/doi/10.1080/10888438.2018.1430804

Robinson, L., Lambert, M. C., Towner, J., & Caros, J. (2016). A comparison of direct instruction and balanced literacy: An evaluative comparison for a Pacific Northwest rural school district. *Reading Improvement, 53*(4), 147-164. http://www.projectinnovation.com/reading-improvement.html

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

Schatschneider, C., Wagner, R. K., & Crawford, E. C. (2008). The importance of measuring growth in response to intervention models: Testing a core assumption. *Learning and Individual Differences, 18*(3), 308–315. https://doi.org/10.1016/j.lindif.2008.04.005

Simmons, D. C., Coyne, M. D., Kwok, O., McDonagh, S., Harn, B. A., & Kame'enui, E. J. (2008). Indexing response to intervention: A longitudinal study of reading risk from kindergarten through third grade. *Journal of Learning Disabilities, 41*(2), 158–173. https://doi.org/10.1177/0022219407313587

Spear-Swerling, L. (2018). Structured literacy and typical literacy practices: Understanding differences to create instructional opportunities. *Teaching Exceptional Children, 51*(3), 201–211. https://doi.org/10.1177/0040059917750160

Tindal, G., Nese, J. F., Stevens, J. J., & Alonzo, J. (2016). Growth on oral reading fluency measures as a function of special education and measurement sufficiency. *Remedial and Special Education, 37*(1), 28-40. https://doi.org/10.1177/0741932515590234

Van Norman, E. R., & Nelson, P. M. (2021). The importance of growth in oral reading fluency to predict performance on high-stakes assessments among students receiving supplemental intervention. *Journal of Applied School Psychology, 37*(1), 1–15. https://doi.org/10.1080/15377903.2020.1772432

Vehkavuori, S. M., Kamarainen, M., & Stolt, S. (2021). Early receptive and expressive lexicons and language and pre-literacy skills at 5;0 years – A longitudinal study. *Early Human Development, 156*. https://doi.org/10.1016/j.earlhumdev.2021.105345

Wagner, R. K., & Torgesen, J. K. (1987). The nature of phonological processing and its causal role in the acquisition of reading skills. *Psychological Bulletin, 101*, 192-212.

Walpole, S., & McKenna, M. C. (2017). *How to plan differentiated reading instruction* (3rd ed.). Guildford Press.

Washington, J. A., Branum-Martin, L., Lee-James, R., & Sun, C. (2019). Reading and language performance of low-income, African American boys in grades 1-5. *Reading & Writing*

Quarterly, 35(1), 42–64. https://doi.org/10.1080/10573569.2018.1535777

Williams, K. T. (2001). *Group Reading and Diagnostic Evaluation (GRADE)*. Pearson.

Willson, A. M., & Falcon, L. A. (2018). Seeking equilibrium: In what ways are teachers implementing a balanced literacy approach amidst the push for accountability? *Texas Journal of Literacy Education, 6*(2), 73-93. http://www.texasreaders.org

Wyse, D., & Bradbury, A. (2022). Reading wars or reading reconciliation? A critical examination of robust research evidence, curriculum policy and teachers' practices for teaching phonics and reading. *Review of Education, 10*(1), 1-53. https://doi.org/10.1002/rev3.3314

APPENDIX A

Institutional Review Board Exemption Report

Institutional Review Board (IRB)
For the Protection of Human Research Participants

PROTOCOL EXEMPTION REPORT

Protocol Number: 04411-2023 **Responsible Researcher(s):** Laura Abercrombie

Supervising Faculty: Dr. Katherine Lamb

Project Title: *Universal Assessment of Reading Ability in Title-1 Schools.*

INSTITUTIONAL REVIEW BOARD DETERMINATION:

This research protocol is **exempt** from Institutional Review Board (IRB) oversight under 45 CFR 46.101(b) of the federal regulations, **categories 1 & 4**. If the nature of the research changes such that exemption criteria no longer apply, please consult with the IRB Administrator (tmwright@valdosta.edu) before continuing your research study.

ADDITIONAL COMMENTS:

- *Upon completion of the research study, collected data must be securely maintained and accessible only by the researcher(s) for a minimum of 3 years. At the end of the required time, collected data must be permanently destroyed.*

☒ ***If this box is checked, please submit any documents you revise to the IRB Administrator at tmwright@valdosta.edu to ensure an updated record of your exemption.***

Elizabeth Ann Olphie **04.11.2023**
Elizabeth Ann Olphie, IRB Administrator

Thank you for submitting an IRB application.
Please direct questions to irb@valdosta.edu or 229-253-2947.

Revised: 06.02.16

APPENDIX B

Letter of Cooperation

[redacted]

Board of Education

Date: 3/10/2023

Re: Letter of Cooperation for [redacted] County Schools

Dear Laura Abercrombie,

This letter confirms that I, as an authorized representative of [redacted] County Schools, allow you access to conduct study-related activities at the listed sites, as outlined below.

- **Research Site(s)**: [redacted]
- **Study Purpose:** The purpose of the study is to analyze various methods of identifying at-risk readers using universal reading measures (i.e., Acadience Reading Diagnostic).
- **Study Activities:** The study will include access to Acadience Reading Diagnostic score reports, MTSS/RTI spreadsheets, student IEPs, and demographic data.
- **The Sites' Support**: The site agrees to distribute and collect consent forms as required by the VSU internal review board and provide access to the necessary data as described above.
- **Data Management:** Student information will be collected including Acadience Reading Diagnostic information, placement in MTSS/RTI and special education, and demographic data. After all data is consolidated into one spreadsheet, identifying data including student name, date of birth, and student ID numbers will be deleted. Participant ID numbers will be generated for the purposes of the study. The participants, sites, and district will not be disclosed.

I understand that this site's participation will only take place during the study's active Institutional Review Board (IRB) approval period. All study-related activities must cease if IRB approval expires or is suspended. I understand that participation will be voluntary; Brooks County Schools' students will not be penalized or rewarded for their participation.

If I have any concerns related to this project, I will contact Laura Abercrombie. For concerns regarding IRB policy or human subject welfare, I may contact the VSU IRB at (229)249-2614.

Regards,

DocuSigned by:
[redacted]
804FC48AF1A84D1...

3/13/2023

Signature of Research Site Authorized Representative ***Date***

[redacted]

K-8 curriculum Director

Printed Name of Research Site Authorized Representative ***Job Title of Site Representative***