Examining the match between reading curricula and curriculum-based measures on the Common Core grapheme-phoneme correspondences.

By

Neena M. Saha

Dissertation
Submitted to the Faculty of the
Graduate School of Vanderbilt University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

in

Special Education

May 8, 2020

Nashville, Tennessee

Approved:

Laurie Cutting, PhD

Chris Lemons, PhD

Joe Wehby, PhD

Emily Phillips-Galloway, PhD

DEDICATION

To my mom and dad, the first scientists I ever knew,
For passing on your endless curiosity

ACKNOWLEDGEMENTS

The research presented in this dissertation was supported in part by Grant H325D140073 from the Office of Special Education Programs, U.S. Department of Education. That said, nothing in this article reflects the positions of the federal government, and therefore no official endorsement should be inferred. The research presented in this article was also supported in part by the Peabody College of Vanderbilt University's Professional Development funds.

Thank you to Dr. Joe Wehby, Dr. Chris Lemons, and Dr. Emily Phillips-Galloway. As committee members and mentors, their guidance throughout this study, and my time at Vanderbilt is greatly appreciated. Thank you especially to my advisor, Dr. Laurie Cutting, who gave me guidance and support but also the freedom to explore questions I was deeply interested in. This led me down paths and reach heights I never thought possible.

Thank you to my husband, Daniel Hudson, for your unconditional support and love throughout these last five years, and of course, for keeping us solvent. Also, thank you to my son, Alexander Saha Hudson, you make my life meaningful in a way that no paper, dissertation, or degree ever could. Last but not least, thank you to Diane and Larry Hudson for watching Alex late into the evening while I worked. The peace of mind knowing he was safe and loved alleviated much of my mom guilt.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

Investigating the Decoding Complexity of Popular Reading Curriculum-Based Measurement Oral Reading Fluency Passages

Introduction

The ability to read fluently and with comprehension is of critical importance to function in today's society and several positive academic and social outcomes are associated with skilled reading ability (e.g., Kern & Friedman, 2009). Therefore, it is unsettling that national reading assessments show roughly 30% of children in the United States are not reading at a basic level (U.S. Department of Education, 2017). In order to prevent and remediate reading deficits, early and ongoing assessment is critical. One of the most valid and reliable methods of tracking reading development is through the use of Reading Curriculum-based Measurement (R-CBM; Miura Wayman, Wallace, Wiley, Tichá, & Espin, 2007). Indeed, R-CBM is the most commonly used type of assessment framework in response to intervention models (Graney & Shinn, 2005; Griffiths, VanDerHeyden, Skokut, & Lilles, 2009), and is also one of the predominant methods of progress monitoring in a data-based individualization (DBI; www.intensive intervention.org) process (Danielson & Rosenquist, 2014).

Curriculum-Based Measurement. Curriculum Based Measurement (CBM) refers to a specific framework for assessment developed at the University of Minnesota. CBM was a revolutionary idea because it combined two previously disparate doctrines of assessment: behavioral time-series trajectories with traditional general outcome measurement. It circumvented the issues associated with mastery measurement and simultaneously provided a feasible way to track the progress of several students (Fuchs, 2017). There is a large literature on the efficacy of CBM for tracking reading development, response to intervention, and progress monitoring (Fuchs, 2017). While there are different types of R-CBM (letter naming fluency, letter sound fluency, maze comprehension, etc.) this paper focuses on oral reading fluency (ORF) R-CBM passages, also sometimes referred to as passage reading fluency. In ORF R-CBM the dependent variable of interest is words read correctly per minute (WCPM). Several researchers have shown that WCPM is a valid measure of overall reading ability that is sensitive to change (e.g., Fuchs, Fuchs, Hosp & Jenkins, 2001).

The development of standardized passages. An important development in the field of R-CBM was the creation of standardized passages (aka 'generic' or 'curriculum-independent' passages). Research conducted in the early 1990s (e.g., Fuchs & Deno, 1994) found that the technical adequacy of R-CBM was maintained even if passages were sampled from outside of a given student's curriculum. These findings led to the realization that a teacher no longer had to select texts for assessment, but rather, an outside party, such as educational publishers, could create a set of standardized passages, one for each grade. This was important in terms of social validity, because it removed the burden of passage selection from teachers and put it on publishers, saving teachers time.

Today, there exist several different standardized R-CBM passage sets created by different publishers. These include, but are not limited to: DIBELS, Acadience (formerly known as "DIBELS Next" but referred to as "Acadience" throughout this paper), FastBridge, and AIMSwebPlus (see Table 1 for background information on these publishers).

Uses of R-CBMs for clinical decisions & questionable psychometric properties. While the development of standardized R-CBMs undoubtedly streamlined the assessment

Table 1
List of publishers of R-CBMs examined in this paper.

CBM	Publisher	Year Published	Number of benchmark passages in first grade	Number of progress monitoring passages in first grade
AIMSwebPlus	Pearson	2015	6	20
Acadience ("DIBELSNext")	Acadience	2010	6	20
DIBELS 8th edition	University of Oregon	2018	3	20
FastBridge	FastBridge Learning	2017	3	20

process, researchers have questioned how the passages are equated (e.g., Miura Wayman et al., 2007; Santi, Barr, Khalaf, & Francis, 2016). Passages that are equated are important to maintaining the integrity of the decisions made based upon the results of the CBM assessment framework. R-CBMs have been used to predict performance on high-stakes tests (e.g., Christ & Ardoin; 2009, Deno, 2003; McGlinchey & Hixson, 2004) and even been put forth as a way to measure performance of educators at the school level (Good et al., 2003). In addition to such high-stakes testing, R-CBMs are also used for relatively "low-stakes" testing such as making instructional changes for non-responders through a process known as progress monitoring (National Center on Response to Intervention, 2013). These small changes, however, could add up to a significant misuse of resources over time if decisions are biased by form effects.

Part of R-CBMs unique appeal is that it is used for both summative and formative assessment. Summative assessment provides important data on screening and eligibility (i.e. additional, more intensive reading instruction), while formative assessment consists of repeated measurement over time, the resulting data informing instructional decisions (Poncy, Skinner, & Axtell, 2005). Summative assessment compares students to one another (relative scores) while formative assessment compares individual growth over time (absolute scores). While using passages of different difficulty (i.e. passages that are not equated) might not change the rank order of student's WCPM relative scores, it could influence an individual's absolute scores (Francis et al., 2008; Petscher & Kim, 2011), undermining the process of progress monitoring.

The difference between absolute and relative scores is important, because the primary reliability statistic often reported by publishers of standardized R-CBMs - alternate form reliability - does not take into account absolute score differences across passages (Cummings, Park, & Bauer Schaper, 2013), leaving a critical psychometric property (absolute score reliability) understudied (Francis et al., 2008). The research that does exist (e.g., Ardoin & Christ, 2009; Francis et al., 2008; Hintze & Christ, 2004) found large differences in an individual's WCPM depending on the particular passage they were reading even when the passages were administered in a short time period to control for growth effects. These differences altered the observed trajectory of the child's growth rate. If growth trajectories are influenced by form effects then the process of R-CBM for progress monitoring is undermined (Santi et al., 2016). In fact, some researchers (Ardoin, Christ, Morena, Cormier, & Klingbeil, 2013, pg. 12) have gone as far as stating, "The conclusion across multiple studies seems apparent: CBM-R progress monitoring is not an evidence-based practice for modeling growth of individual students' gains in reading."

Current equating methods for R-CBM passages. Currently, publishers take several different approaches for equating passages. Some approaches take what we will refer to as a "bottom up" approach, or attempting to measure different aspects of text (e.g., word frequency, sentence length, etc.) a priori, or *while* developing the passages. Others take a more "top down" approach, utilizing statistical methods of equating passages *after* the passages have been developed. While both approaches have merit, each also has potential pitfalls. Of note, no approach to date has attempted to capture decoding difficulty of passages, which may be especially important for beginning as well as struggling readers (the importance of controlling for decodability of text is discussed after current equating techniques). Below we review the current "bottom up" and "top down" ways passages are equated.

A priori methods of reducing variability.

Readability formulas. The technical manuals of popular standardized R-CBM passages, such as DIBELS and AIMSwebPlus (DIBELS 8th edition Administration and Scoring guide,

Table 2
Different passage equating techniques across publishers.

Passage set	Description of Passage Equating Procedures
AIMSwebPlus (NCS Pearson Inc., 2012) and (Development Manual, 2016)	• Teachers and paraprofessionals were trained to write passages that were appropriate in length and that contained the grade-appropriate number of syllables and sentences per 100 words, based on the Fry Readability formula • Passages were discarded that fell below grade level as indicated by the Lexile readability formula. • Benchmark passage levels were aligned to common readability indices (Lexile, Fry, Flesch-Kincaid, Powers, Spache and the SMOG), with correlations ranging from 0.83 to 0.97. Similarly, progress monitoring passage levels were aligned to common readability indices (e.g., Lexile, Flesch-Kincaid), with correlations ranging from 0.81 to 0.91 • Shinn (2012-unpublished data) updated AIMSweb 1.0 passages to include about 60 words of highly decodable text for first grade probes (pg. 23 AIMSwebPlus Development Manual) • Grades 2-8 were updates but are basically the same as AIMSweb
FastBridge Learning Formative Assessment for Teachers (Technical Manual Version 3.0 2016-2017 & Christ & Ardoin 2009 pg. 60)	• 50 3rd grade passages were selected from 9 textbooks using methods described by Shinn (1989) and Hintze et al. (1998), passages were narrative, proper nouns were reviewed by 2 undergraduates and difficult passages were replaced • Of the original 50 passages, 20 were selected by examining the Euclidean distance between passage variance using student ORF performance, passages with the lowest Euclidean distance were kept
DIBELS 8th edition (Administration and Scoring Guide, 2019)	• Aspiring and published children's authors were given word-length limits and Flesch-Kincaid readability levels to stay within but were told to create diverse narrative and expository texts • Passages were reviewed by an external group of parents and teachers and discarded/changed as needed • Passages were field-tested on students in multiple grades and discarded/changed as needed
"Acadience" (Technical Manual, 2013)	• Passages were designed to be authentic text and are not written entirely in decodable text • Page 28 of the Technical Manual lists the 15 requirements of the passage construction process including: • Avoiding unusual and repetitive sentence structures • All passages had to meet readability criteria as determined by the DIBELS passage Revision Utility software which examines word length, rare words, and sentence length and provides guidance when the range of a passage is outside the DMG Passage Difficulty Index • The DMG Passage Difficulty Index is described in detail on pg. 30 of the DIBELS Next Technical Manual, but briefly; it assesses decoding difficulty of a passage by examining (a) characters per word (b) percent of words with three or more syllables (c) percent of words with seven or more characters, and (d) number of syllables per word. Unlike a readability formula which looks at these in combination, the DMG Passage Difficulty Index looks at these in isolation and makes sure that each passage falls within the appropriate range. • A readability study was then conducted and the passages with the smallest average residuals, smallest standard deviation of the residuals and highest alternate-form reliability plus 5 other factors were chosen (Powell-Smith, Good, & Atkins, 2010).

2019; AIMSwebPlus Efficacy report, 2016), state that passages were equated using readability formulas (see Table 2 for passage equating techniques used by the different publishers).

For example, a research report published by Pearson (the creators of AIMSweb and AIMSwebPlus; NCS Pearson Inc., 2012, pg. 24) state that, “Teachers and paraprofessional were trained to write passages that were appropriate in length (approximately 250 words for Grades 1 and 2, 300 words for Grades 4-8) and that contained the grade-appropriate number of syllables and sentences per 100 words, based on the Fry readability formula.”

This approach may be problematic for several reasons. First, readability formulas tend to use only a few, broad, variables and do not capture salient variables known to contribute to reading (Bruce & Rubin, 1988; Davison & Kantor, 1982). For example, most of the frequently employed readability formulas such as Flesch-Kincaid and Fry only take into account two variables (Fry: number of sentences and the number of syllables; Flesch-Kincaid: average words/sentence and average syllables/word). Furthermore, the variables that are used in the readability formulas are rather broad, such as the number of words per sentence.

While readability formulas may be sufficient for older children, they have been criticized as not being sensitive in the primary grades, where fine-grained distinctions (subtler than grade level) are required to match readers to appropriate text, such as a reader’s decoding level (e.g., Cunningham et al., 2005). To this end, in an effort to examine the validity of popular readability formulas, Hiebert and Pearson (2010) compared five readability indices in seven types of kindergarten through second grade texts: Degrees of Reading Power, Fry, Spache, Lexile, and Coh-metrix. They examined whether the readability formulas matched the progression of difficulty as laid out by the publishers (basically, they examined the content validity of the formulas). Notably, while they found that the formulas were consistent with the rankings of several categories of texts, they were not consistent for the category of decodable texts, which varied widely across readability indices, suggesting that readability formulas are not valid for beginning reader texts (texts that, in theory, should have high repetition and carefully scaffolded introduction of phonic elements). This is concerning, given the well-established finding that phonics-based approaches are a necessary element of beginning reading instruction (e.g., Ehri, Nunes, Stahl, & Willows, 2001). Furthermore, studies that have specifically examined whether readability formulas are accurate predictors of ORF, a measure closely linked to students' decoding abilities, have found no relationship, especially among the readability formulas used to equate R-CBM passages (Ardoin, Suldo, Witt, Aldrich, & McDonald, 2005; Christ & Ardoin, 2009; Compton, Appleton & Hosp, 2004; Poncy et al., 2005). Additionally, other research has shown that readability formulas do not yield passages of similar difficulty (Dunn & Eckert, 2002; Hintze, Shapiro & Lutz, 1994).

***Statistical equating using student performance*.** Since the use of readability formulas is problematic for the aforementioned reasons, some researchers have put forth alternative or additional measures of equating passages known as statistical equating (Albano & Rodriguez, 2011; Santi et al., 2016; Stoolmiller et al., 2013). Statistical equating is a "top-down" approach: rather than ensuring text factors are similar across passages ("bottom-up"), statistical equating examines student performance on passages and provides corrections to certain passages that have been deemed too difficult or too easy. Corrections are numerical adjustments to the student's WCPM score that can be provided to practitioners in the form of a lookup table. Santi et al. (2016) and Stoolmiller (2013) provide excellent reviews of four types of statistical equating methods with examples applied to real data: linear equating, equi-percentile equating, latent variable linear equating, and latent variable non-linear equating. The benefits and limitations are discussed thoroughly in Santi et al. (2016) but briefly, these forms of equating involve collecting ORF data on hundreds (or thousands depending on the method) children and anchoring each passage to a common scale. However, Santi et al. (2016) state that one disadvantage of this approach is the burden on R-CBM test developers to run these analyses and provide lookup tables for teachers to convert a child's WCPM score to a scaled score.

There are also several additional limitations associated with statistical equating. For example, Albano & Rodriguez (2011, pg. 57) state: "the use of numerous forms of an assessment, as with progress monitoring, makes it very difficult to obtain a single, large sample of students across all forms." Even if a large, representative sample could be secured there still remains the issue of how to administer several passages in a short time period so as not to introduce growth effects. Another limitation is the burden of an additional step in the R-CBM process, a step that can introduce practitioner error. Researchers have reported that teachers are already pressed for time and struggle to use R-CBM data to inform instructional decisions. Moreover, the derivation of the lookup tables is not intuitive to understand, and teachers might be wary of using them if clear professional development is not provided alongside the introduction of the tables. Cummings, Park and Bauer Schaper (2013, pg. 93) address this point when they mention that equating results are "difficult to interpret" and "could potentially remove some of the authenticity of R-CBM."

While software can alleviate the extra step of converting the child's raw WCPM score, software alone probably cannot answer all the questions that a teacher might have. Cummings et al. (2013, pg. 93) state, "There is no substitute for careful passage authoring field testing and passage review by content area experts. In fact, equating is only a successful procedure when it is applied to test forms that are designed to be exactly the same (Kolen & Brennan, 2004)."

Moreover, statistical equating might still yield inaccurate results and even present more error than it eliminates (Stoolmiller, Biancarosa, & Fien, 2012).

Euclidean distance. Another type of "top-down" method for examining passage variability is the use of cluster analysis, and was used by the developers of the FastBridge oral reading passages (see Table 2 for passage equating techniques used by the different publishers). Instead of providing corrections to student WCPM scores in the form of lookup tables, Christ and Ardoin (2009) used the method in an *a priori* fashion: testing passages on students and eliminating anomalous passages before creating their official set of passages. Christ and Ardoin (2009) examined the proportion of variance across passage sets that had been developed using four different procedures: random, readability, mean, and Euclidean Distance. They used Shapiro's (1996) guidelines for selecting a random set of passages. The readability passage set used readability formulas as described in Hintze & Christ (2004). The mean passage set was constructed based on student performance means as described in Poncy et al. (2005). Lastly, the Euclidean Distance set used cluster analysis. Their results showed that the proportion of variance was lowest (1%) for the Euclidean Distance set as compared to the random passage set (10%).

While FastBridge may be unique in its application of cluster analysis to passages, it is important to note that FastBridge is not the only standardized R-CBM publisher that used student performance in an iterative fashion to eliminate anomalous passages. Indeed, DIBELS 8th edition and Acadience passages were also field-tested on students to determine outliers.

Procedural methods of reducing variability. Procedural methods are another way of reducing form effects after the passages have already been constructed. Procedural methods refer to specific actions intended to minimize variance during the administration of the passages. For example, instead of using a single passage WCPM score, three passages are administered, and the median of the three passages is used as an estimation of the true score. However, an obvious limitation of this method is the additional strain on teachers' time as it requires administration of two additional passages.

In summary, there are problems with both "bottom-up" and "top-down" methods of equating standardized R-CBM ORF passages. Top-down methods require large samples that may not be feasible to obtain, while bottom-up methods such as readability formulas appear to be too coarse to capture the influence of sub-lexical features of words such as orthographic transparency (e.g., Ellis & Hooper, 2001) or phonotactic probability (e.g., Vitevitch & Luce, 1999).

The importance of certain text characteristics for early grade R-CBMs: decoding difficulty. While WCPM has been shown to be a valid measure of general reading ability (Fuchs et al., 2001), it is known that WCPM can be influenced by different text factors (e.g., Barth, Tolar, Fletcher, & Francis, 2014). One factor known to contribute to text difficulty (as measured by student ORF performance) is decoding complexity, or 'decodability' of a text (Mesmer et al., 2012). A recent network meta-analysis found that text complexity tools that incorporated sub-lexical components were more strongly correlated with student's oral reading performance (Saha & Cutting, 2019).

Decoding complexity of text is especially important to measure and control for in early grade texts (Mesmer et al., 2012), as it is arguably the main reading skill that children are acquiring at this stage (Ehri, 1995; National Reading Panel, 2000). There is substantial evidence to support the idea that emerging readers engage in slow, deliberate processing of sub-lexical (aka 'intra-word') variables, such as grapheme-phoneme correspondences, rather than coarser text variables. In their theories on the development of reading skill, Adams (1994), Ehri (1995), and Perfetti (1992) all acknowledge that beginning readers differ from skilled readers in their sensitivity to letter information. Attention to sub-lexical processing is also highlighted in practice of teaching reading, not just theories of reading development. For example, explicit instruction in grapheme-phoneme correspondences is covered by several major first-grade core reading programs (see Chapter 2) and the widely-adopted Common Core State Standards (CCSSs: Common Core State Standards Initiative, 2010 Appendix A) even go as far as to list specific grapheme-phoneme correspondences that first-grade children should know by the end of the year.

Despite the evidence that R-CBM passages should be controlled for decodability in the early grades, it is unclear if R-CBM passages are actually equated on decoding difficulty. Indeed, Toyama et al. (2017) found ‘considerable’ variability on within-grade passage means (measured via Lexile, Flesch-Kincaid, the Reading Maturity Metric, and the TextEvaluator) including Acadience oral reading passages despite a thorough process for passage creation and selection (See Table 2 for passage construction and equating information on the R-CBM passages investigated in this study). Furthermore, Poncy et al. (2005) examined the variance in student performance (WCPM) on DIBELS oral reading fluency passages from 2002 using 37 third-graders. They found that 10% of the variance was due to passage variability (form effects) and 9% of the variance remained unexplained. This resulted in standard error of measurement swings of up to 18 WCPM from passage to passage. Such variability could lead to skewed data and decisions based on incorrect data.

Given that R-CBM ORF is intended to measure growth in early reading skill (decoding skill) it is critical to examine text features that influence decoding skill so that differences in decoding complexity of texts do not interfere with determining the true score of WCPM for a given child. Or, as Toyama, Hiebert, and Pearson (2017, pg. 140) state: “if you want to measure change, don’t change the measure.” Variation in passage difficulty should take into account as fine-grained text features as possible, and control for differences in difficulty. Fine-grained attention to sub-lexical features of words and how such features contribute to text complexity is therefore an important but understudied variable in ORF R-CBM passages. Perhaps this type of analysis is not found in the research because an automated, quantitative measure of decoding difficulty was not available. To address this limitation in the literature, the current study analyzes the decoding complexity across standardized R-CBM passages using a new measure of decoding difficulty that takes into account several sub-lexical features of words known to influence word-reading accuracy and fluency. This new measure, the Decoding System Measure, is described in detail in the methods. The overarching focus of this aim was to determine if passages used in R-CBMs with beginning readers (1st grade) either equated through bottom up or top down processes were actually also equivalent on decoding difficulty. Given prior research (Hiebert & Pearson, 2010; Toyama et al., 2017), we hypothesized that there would be significant variability in decoding, despite equivalence being demonstrated via other methods.

Research Questions

While the primary focus of this aim is on passage equivalence *within* sets of R-CBMs, a secondary aim was to examine differences *across* publishers in overall decoding complexity. The first two research questions pertain to this primary aim (examining passage-to-passage variance in decoding complexity), while the last question pertains to the ancillary focus of this paper.

(1) Do first-grade oral reading fluency passages within a publisher come from a common distribution in terms of decoding complexity?

(2) Do first-grade passages (including benchmark and progress monitoring passages) within a publisher have outliers in terms of decoding difficulty (i.e., have a DSyM passage mean that is above or below one averaged standard deviation of the remaining sample)?

(3) Does the mean decoding complexity of first-grade R-CBM oral reading passages vary across publishers (AIMSWebPlus, FastBridge, Acadience, and DIBELS 8th edition)?

Methods

Selection of R-CBMs. The NCLII Tools Chart for Progress Monitoring (National Center on Intensive Intervention, 2018) was used to select four commonly used R-CBMs with strong evidence of reliability and validity: AIMSwebPlus, Acadience, DIBELS 8th edition, and FastBridge. The ORF passages for first-grade were acquired either through direct purchase or free download via online websites (in the case where materials are provided free of charge).

The Decoding System Measure: A Tool for Assessing Text Decodability. The Decoding System Measure (or, DSyM; Cutting, Saha & Hasselbring, 2018) is a quantitative measure of word-level decoding complexity that incorporates four variables: (1) word frequency, (2) the discrepancy between the number of letters in a given word and the number of phonemes, (3) the conditional probability of the grapheme-phoneme correspondences for vowels in the word, and; (4) the number of blends in a given word. Points are assigned for each one of the components and added together, higher word scores indicate that a word is harder to decode. For example, the word “mom” has a DSyM score of 1.15, whereas the word, “favorites” has a DSyM score of 5.02.

Initial validation evidence of the DSyM across five studies and four unique participant samples found that the DSyM predicted children’s word reading accuracy and fluency above and beyond current passage and word-level measures of text difficulty (Saha, DelTufo, Bailey, & Cutting, in preparation).

The Components of the DSyM.

Word frequency. The word frequency component score is calculated by subtracting the word frequency percentile score from the Standard Frequency Index corpus (Zeno, Ivens, Millard, & Duvvuri, 1995). So, for example, if the word, "you" had an SFI score of .80, then .8 would be subtracted from 1 to get .2. This number (.2) would comprise the word frequency score component and would then be added to the remaining three components. More points are awarded to rarer words.

Letter-sound discrepancy. The second component is the discrepancy between the number of letters and phonemes in a word. This is calculated by subtracting the number of phonemes from the number of letters. For example, again using the word "you", the two phonemes (/y/ and /oo/) would be subtracted from three letters yielding a score of 1 for this component. Less transparent words (i.e. words with a higher ratio of letters to sounds) are awarded more points.

Conditional probability. The third component is the conditional probabilities of the grapheme-phoneme correspondences for the vowels in the word. This metric is calculated by subtracting the frequency of a particular grapheme-phoneme vowel match (i.e. "ou" - /oo/ in "you") from 1. So, for example, if "ou" - /oo/ has a frequency in the Berndt corpus (Berndt, Reggia, & Mitchum, 1987) as .40, then .4 would be subtracted from 1 yielding .6, which would form the conditional probability score component. Since consonants are highly regular (and did not increase the R^2 in initial exploratory analyses), they were not taken into account as part of this component. Vowels with rarer sound matches (such as the letter 'a' making the /ih/ sound in the word "spinach") are awarded more points.

Number of blends. The last component, the number of blends, is calculated by tallying the number of consonant blends (or clusters) within a word. Digraphs are not included in this component, because they are already accounted for in the discrepancy component. Words with more blends are awarded more points.

There are two deviations from the above scoring procedure and those concern the words “a” and “I.” For the words, “a” and “I” a score of .5 was assigned.

Scoring of passages. All passages were scored using the Decoding System Measure’s scoring system which is currently an online web application. Scoring involved copying and pasting the text from the document into the web browser. Since the scoring application is still in beta version, word scores were checked by the first author for mistakes. On rare occasions a word was not scored because it was not in the relational database that the DSyM web app uses to score. In these cases, the word was hand-scored by the first-author. DSyM word scores were then averaged within a passage to get a mean DSyM passage score for all the progress monitoring and benchmark passages. All of the data is available upon request by contacting the author.

Analysis

Examining passage equivalence within publishers (research questions 1 & 2).

ksamples Anderson-darling test. The kSamples Anderson-Darling test (package ‘kSamples’; Scholz & Zhu, 2019) was used to investigate whether passages from the same publisher came from a common distribution. The Anderson-Darling test is a non-parametric statistical test used to determine if a particular sample comes from a certain type of distribution. For example, the test answers the question: are the words in Acadience benchmark passage 1 from a normal distribution? However, the *kSamples* Anderson-Darling test is a measure of agreement between two distributions for whether a number of random samples come from the same (unspecified) distribution. Put simply, the kSamples Anderson-Darling test allows for an overall comparison of several passages to determine if they are from a common underlying distribution. The results reported here were achieved by bootstrapping the sample 10,000 times. These analyses included both the benchmark and the progress monitoring passages.

Bootstrapping. In order to determine which first-grade passages were outliers within a publisher, a bootstrapping approach was used. Each individual passage mean was removed from the sample one at a time and the removed passage mean was then compared to see if it was in between a specific interval. The interval was created by adding and subtracting the *averaged* standard deviation from the mean of the remaining passages. The averaged standard deviation was derived by calculating the mean of the four standard deviations of the full set of passages across all four publishers and included all benchmark passages. An average standard deviation was used to create the interval, rather than the standard deviation of the individual set of passages, so as not to unfairly penalize passage sets with low overall variation. For example, Acadience passages had low variability compared to the other publishers, yet if the standard deviation of Acadience (which was relatively quite small) was used to create the interval, several Acadience passages would end up falling outside the interval, when in reality, Acadience had the least overall variability. Therefore, the averaged standard deviation (.21) was added and subtracted from the passage set means.

Examining decoding complexity across publishers (research question 3).

RDI boxplots. *T*-tests adjusted for multiple comparisons were conducted to analyze mean decodability differences across publishers. RDI graphs were also created to help visualize the differences. RDI refers to graphs that plot the raw data, descriptive statistics (such as the mean), and inferential statistics (such as confidence intervals) all on the same graph. RDI boxplots, with the 95% frequentist confidence interval, were created to show variability among passages within a publisher, and then across publishers. This allows for an intuitive visual assessment of group differences and provides the same information as inferential statistical tests such as *t*-tests. For example, if the 95% frequentist confidence intervals (the rectangular boxes) do not overlap across two publishers, then these publishers are statistically different. RDI boxplots were created in R (R Core Development Team, 2019) using the package 'yarrr' (Phillips, 2017).

Results

How does the mean decoding complexity of R-CBM oral reading passages vary across publishers (AIMSwebPlus, FastBridge, Acadience, and DIBELS 8th edition)? There were significant differences across all six comparisons among the four publishers (see Table 3). Acadience (M = 2.05, SD = .10) passages were harder to decode than DIBELS 8th edition {(M = 1.92, SD = .18), $t(47) = 3.02$, $p = .004$)}, AIMSwebPlus {(M = 1.76, SD = .16), $t(50) = 7.74$, $p < .001$)}, and FastBridge {(M = 1.64, SD = .14), $t(47) = 11.73$, $p < .001$)}. FastBridge passages were easier to decode than DIBELS 8th edition $t(44) = 5.93$, $p < .001$), and AIMSwebPlus $t(47) = 2.77$, $p = .008$). Finally, DIBELS 8th edition passages were harder to decode than AIMSwebPlus passages $t(47) = 3.35$, $p = .002$).

The RDI boxplots in Figure 1 graphically display the results of the *t*-tests. The DSyM means (the individual data points each represent one passage mean) vary by publisher across first grade. The mean of each of the passage means (including benchmark passages) is represented by the black bar in the center of the shaded rectangle and ranges from 1.64 (FastBridge) to 2.05 (Acadience). AIMSwebPlus and FastBridge generally contain passages that are easier to decode than DIBELS 8th edition, and Acadience. AIMSwebPlus is lower and does not overlap with

Table 3

P-values and effect sizes from T-tests of means between all four R-CBM publishers.

	DIBELS 8th		Acadience		FastBridge	
	p-value	cohen's *d*	*p*-value	cohen's *d*	*p*-value	cohen's *d*
AIMSwebPlus	0.002	0.955	< 0.001	2.170	0.008	0.781
FastBridge	< 0.001	1.730	< 0.001	3.320		
Acadience	0.004	0.874				

Note. All *p*-values were significant at the corrected value alpha level of .008

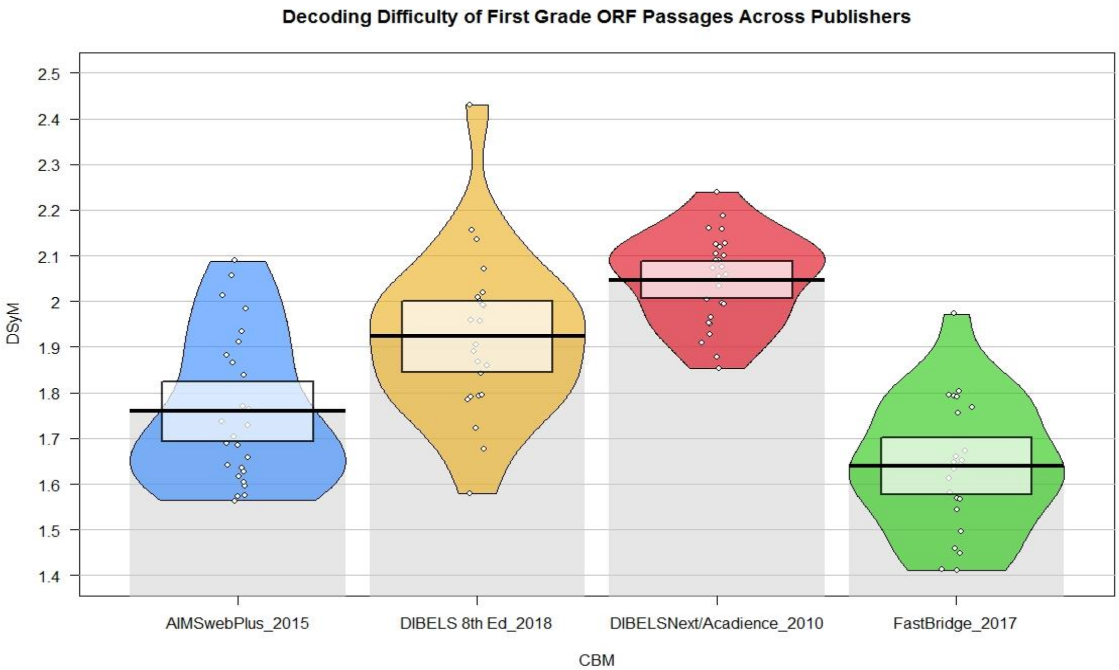


Figure 1. Decoding Difficulty (DSyM Mean) of R-CBM ORF passages across all publishers.

DIBELS 8th edition, or Acadience. The same is true for FastBridge: FastBridge's 95% confidence interval is lower than, and does not overlap with DIBELS 8th edition or Acadience.

DIBELS 8th edition contained the passage with the highest DSyM score (Passage 2, DSyM = 2.43). This passage contained several difficult words such as toothbrush, bristles, supplies, circular, and mouthwash. By way of comparison, the passage with the lowest DSyM score was a tie between passage 14 and passage 16, both part of the FastBridge set (DSyM score = 1.41).

Do first-grade oral reading fluency passages within a publisher come from a common distribution? KSamples Anderson Darling tests were performed to examine passage-to-passage variability (see Table 4). The *p*-values for all publishers were significant: AIMSwebPlus (AD = 25.89, $p < .001$), Acadience (AD = 8.776, $p < .001$), FastBridge (AD = 13.04, $p < .001$), and DIBELS 8th edition (AD = 22.19, $p < .001$). This finding suggests that first grade passages do not belong to a common population, no matter which publisher was examined.

Which first-grade passages (including benchmark and progress monitoring passages) within a publisher are outliers (have a DSyM passage mean that is above or below one averaged standard deviation of the remaining sample)? Individual passage means are displayed in the 'forest plots' in Figure 2. The passages are labelled in order on the y-axis according to the passage number assigned by the publisher. These plots echo the results of the RDI boxplots (Figure 1): Acadience, while more difficult to decode, has much less variability from passage to passage than AIMSwebPlus, DIBELS 8th edition, and FastBridge.

Yet, unlike Figure 1, Figure 2 displays the passages in order, so that consecutive passage-to-passage variability can be examined, which is important because progress monitoring gauges student performance on three or four consecutive passages (discussed further below).

Table 4

K-Samples Anderson-Darling test statistics for first-grade progress monitoring & benchmark passages.

	AD value	AD test statistic	*P*-value	Simulated *p*-value
AIMSwebPlus	123.30	25.89	<.001	<.001
Acadience	58.33	8.776	<.001	<.001
FastBridge	68.42	13.04	<.001	<.001
DIBELS 8^{th} ed.	101.00	22.19	<.001	<.001

Note. Results are based on 10,000 simulations

The results of the bootstrapping analyses are displayed in Table 5. All of the publishers except Acadience have at least three passages that fell either above or below one averaged standard deviation of the remaining passages. AIMSwebPlus and DIBELS 8th edition each had 4 passages outside a standard deviation, while FastBridge had 3. The individual passage means and are displayed graphically in Figure 2.

One can see from this figure that Acadience has low passage to passage variability (the individual passages fall almost along a vertical line), but higher overall decoding complexity (the vertical line is centered around DSyM score of 2). AIMSwebPlus, DIBELS 8th edition and FastBridge all have more variability but generally lower decoding complexity (easier to decode). From a psychometric point of view, examining passage equation (the primary aim of this paper and the concern of research questions 1 & 2), the decoding complexity level is less important than the passage-to-passage variability. However, the different levels of the passages have important practical implications which will be discussed further below.

Discussion

Summary. This is the first study to present the results of text decoding difficulty of first-grade ORF R-CBM passages across publishers using a quantitative, automated measure of text decoding difficulty. Specifically, a new, automated tool of decoding text complexity (the DSyM) was used to examine the differences across AIMSweb, AIMSwebPlus, Acadience, DIBELS 6^{th} Edition, DIBELS 8th edition, and FastBridge passages, with a particular emphasis on whether passages within a publisher could be considered alternate forms.

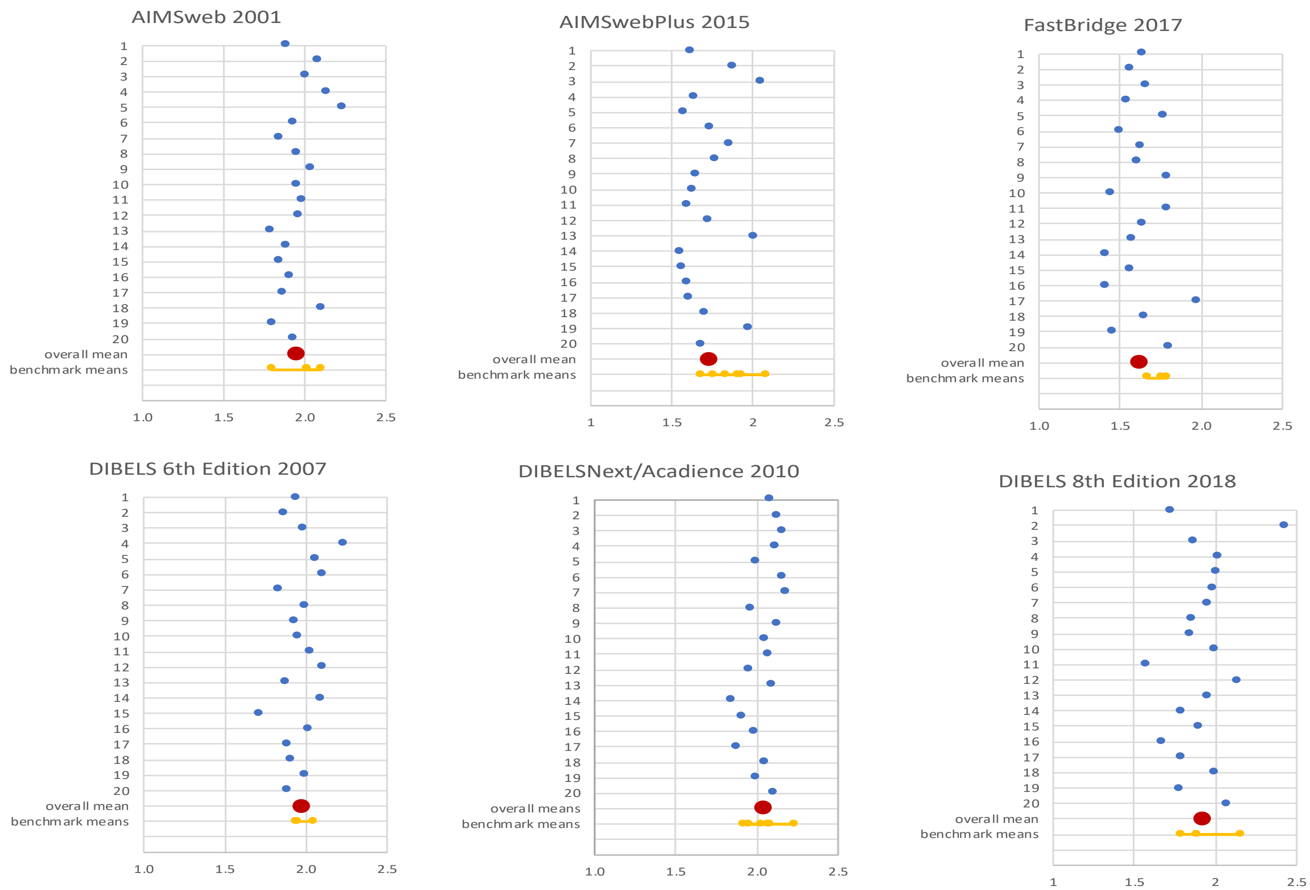


Figure 2. Plots of the DSyM passage means for progress monitoring passages for Grade 1 ORF R-CBMs across publishers.

Table 5
Means and results of the bootstrapping analyses by passage and publisher for Grade 1.

Passage Number	AIMSwebPlus		DIBELS 8th edition		FastBridge		DIBELSNext “Acadience”
Mean	1.76 (1.55, 1.97)		1.92 (1.71, 2.14)		1.64 (1.43, 1.85)		2.05 (1.83, 2.26)
1	1.63		1.72		1.64		2.09
2	1.88		2.43	High	1.57		2.12
3	2.06	High	1.87		1.66		2.16
4	1.64		2.02		1.54		2.12
5	1.57		2.01		1.77		2.00
6	1.74		1.99		1.50		2.16
7	1.87		1.96		1.63		2.19
8	1.77		1.86		1.61		1.97
9	1.66		1.84		1.79		2.13
10	1.64		2.00		1.45		2.06
11	1.60		1.58	Low	1.79		2.07
12	1.73		2.14		1.65		1.95
13	2.01	High	1.96		1.58		2.10
14	1.56		1.79		1.41	Low	1.85
15	1.57		1.90		1.57		1.91
16	1.60		1.68	Low	1.41	Low	1.99
17	1.62		1.79		1.97	High	1.88
18	1.70		1.99		1.65		2.05
19	1.98	High	1.78		1.46		2.00
20	1.69		2.07		1.80		2.10
BM 1	1.68		1.89		1.79		2.03
BM 2	1.94		2.16	High	1.76		1.95
BM 3	1.84		1.79		1.67		2.07
BM 4	1.76		N/A		N/A		1.93
BM 5	2.09	High	N/A		N/A		2.24
BM 6	1.91		N/A		N/A		2.09
Total Low		0		2		2	0
Total High		4		2		1	0
Total		4		4		3	0

Note. DIBELS 8th edition and FastBridge each only had three benchmark passages.

The results showed there were statistically significant differences in decoding complexity across publishers in first-grade. Across time, there was a trend for lower decoding complexity, but more variability within publishers. Furthermore, there was substantial variability from passage to passage within a given publisher. Statistical tests showed that passages were not from a common distribution and bootstrapping analyses identified several passages that fell outside an interval defined by the averaged standard deviation. AIMSwebPlus and DIBELS 8th edition each had 4 passages outside a standard deviation while Acadience had none. The lack of passage-to-passage variability in Acadience perhaps is explained by the level of attention that was given to passage construction. It is possible that Acadience first-grade passages showed the least variability due to the extra *a priori* text features they utilized, many of which were more fine-grained than the variables in readability formulas that capture 'coarser' sentence level measures. For example, in addition to readability formulas, the Acadience technical manual (DIBELS Next Technical Manual, 2013) states that four additional variables were used to equate passages on decoding complexity: (1) the number of letters per word, (2) the percent of words with 3 or more syllables, (3) the percent of words with 7 or more letters, and (4) the number of syllables per word (see Table 1 for information on the R-CBMs examined in this paper and Table 2 for a description of the equating techniques used by the different publishers).

Limitations. One limitation of this study is the use of one standard deviation (in either direction high or low) as a metric of passage non-equivalence. This was indeed arbitrary, and perhaps too liberal. One could easily argue that a window of plus or minus half a standard deviation should be used, given the high-stakes decisions that R-CBMs are used to determine. That said, this being an initial study, with a new measure (the DSyM), we feel that one standard deviation is an appropriate starting point and future studies could explore different metrics of passage non-equivalence.

A second limitation is that the passages contained different numbers of words. FastBridge passages were generally shorter while Acadience passages tended to be longer. In order to address this limitation and control for passage length, the analyses were repeated using only the first 60 words of each passage. The number 60 was chosen because it is roughly the number of words per minute that a first-grader of average reading ability is able to accomplish by the end of the year (Hasbrouck & Tindal, 2006). The correlations between the DSyM scores on the full passages with the DSyM scores on just the first 60 words of each passage were as follows: DIBELS 8th edition (r = .50), Acadience (r = .57), AIMSwebPlus (r = .60), and FastBridge (r =.82). The relatively low correlations suggest that the decoding difficulty of the first 60 words is not necessarily representative of the entire passage. The re-calculated averaged standard deviation using just the first 60 words was .16 (compared with .21 when the full passage was scored in the original analyses). This could be due to the intentional use of graded difficulty within the passage. Indeed, there is some evidence that at least one publisher purposefully constructed the first 60 words to be of easier decodability – to prevent floor effects for poor readers, while the rest of the passage was more difficult- to prevent ceiling effects (AIMSwebPlus Development Manual, 2016). It is unknown if other publishers used this same process, but according to the published material the author was able to locate (see Table 2), it appears that the other publishers did not use this approach.

In terms of the bootstrap analyses on just the first 60 words, there are more passages across publishers that fall outside the +/- averaged standard deviation (.16) interval (see Table 6).

Table 6
Means and results of the bootstrapping analyses by passage and publisher for Grade 1 for the first 60 words in the passages.

Passage Number	AIMSwebPlus		DIBELS 8th edition		FastBridge		DIBELSNext "Acadience"	
Mean and range before removing any passages	1.25 (1.09, 1.41)		1.95 (2.11, 1.79)		1.55 (1.39, 1.71)		1.97 (1.81, 2.13)	
1	1.25		1.49	Low	1.55		2.17	High
2	1.24		2.24	High	1.34	Low	1.88	
3	1.36		1.85		1.72	High	2.05	
4	1.05	Low	1.68		1.40		2.02	
5	1.06	Low	1.87		1.70		2.07	
6	1.24		1.96		1.61		2.27	High
7	1.51	High	1.82		1.46		2.04	
8	1.40		1.84		1.52		1.98	
9	1.38		1.87		1.65		1.87	
10	1.07	Low	2.03		1.38	Low	1.87	
11	1.17		1.82		1.90	High	2.11	
12	1.22		1.98		1.60		1.82	
13	1.40		2.18	High	1.45		1.90	
14	1.26		2.07		1.38	Low	1.68	Low
15	1.19		1.96		1.43		1.92	
16	1.28		1.94		1.44		1.83	
17	1.18		1.80		1.80	High	1.72	Low
18	1.26		2.14	High	1.56		1.85	
19	1.30		2.07		1.38	Low	2.18	High
20	1.26		2.43	High	1.69		2.07	
Total Low		3		1		4		2
Total High		1		4		3		3
Total		4		5		7		5

While AIMSwebPlus had the same number of passages that fell outside the interval as in the original analyses, there was an increase in outlier passages across the rest of the publishers. Notably, Acadience, which did not have any passages outside the interval in the first analyses, now has 5 passages outside the interval when only looking at the first 60 words. These results are problematic in terms of practical implication, suggesting that for poor readers (those reading less than 60 WCPM) there is more variability in decoding difficulty from passage to passage, which could lead to error-prone progress monitoring data for the children who need it most. The results presented here echo the sentiment put forth by Ardoin et al. (2013) that more research is needed on form effects in R-CBMs.

Another limitation of this study is that the DSyM is a text measure, and does not take into account student performance. That said, initial validation evidence for the DSyM shows it predicts children's word reading errors across four independent samples of children (Saha et al., in preparation). Furthermore, and perhaps more importantly, even statistical equating methods such as equi-percentile equating assume that passages are similar in order for the method to work (Kolen & Brennan, 2008). Therefore, prior to any student performance data, there need to be more rigorous, research-based methods of passage equation, and the DSyM provides a method for passage equating at the sub-lexical level. Validation is an ongoing process, and examining the DSyM scores on R-CBM passages in relation to student performance on those same passages is an interesting and necessary future direction.

A final limitation is that the DSyM mean was used as a measure of location to represent a given passage's decoding complexity. However, there are over 60 measures of central tendency (Andrews et al., 1972) and it is possible that another measure might serve as a better summary for a given text. Future work could examine how these different measures of location are correlated with actual student performance to determine the best summary measure for a given piece of text.

Additional future directions could examine differences between publishers across grades, rather than just grade 1 and examine level with more scrutiny, rather than passage-to-passage variability. Our results showed that Acadience had the least passage-to-passage variability, but its passages were also the highest in terms of decoding difficulty and often contained difficult to decode words. This might be problematic for struggling readers, as some research has shown that R-CBM ORF is not a valid measure unless a certain number of WCPM can be achieved.

Practical Implications. There are several practical implications of the results discussed above. First, all R-CBMs are not equal in terms of decoding complexity and switching from one publisher to another (at a school district level, or at the individual level as in the case of a child moving to a new school) could influence student results and high-stakes decisions.

Particularly troubling is the fact that some benchmark passages (which are used for universal screening) are among the outlier passages (see Table 5). This could lead to over-identification (since the passages were outliers on the 'High' side of the interval) of children needing additional reading support.

Third, given that three of the four publishers had at least three passages that fell outside one averaged standard deviation of the remaining passages' mean, it is important to discuss the issue of order of passage implementation. Decisions about instructional modifications are often based on student performance on three or four *consecutive* WCPM scores. While a full discussion of best practices in R-CBM administration is beyond the scope of this paper, generally, teachers judge whether student performance is above or below a goal line. If three (sometimes four) WCPM score in a row are below the goal line, a modification is warranted.

Therefore, it could be problematic that FastBridge has two passages (not consecutive, but close to it) that fall below one averaged standard deviation (see Table 5). One can envision a scenario where a student suddenly is above their goal line and looks like they are making adequate progress, but in reality, this is due to passage effects thus providing a teacher with a false sense of security.

That said, while the four R-CBM publishers examined in this paper all order their passages (i.e. provide a number to each passage labelled 1-20) it is unknown if this order is actually implemented in practice (vs. teachers using a random order if they believe the passages to be equated). Using a random order could lead to administering three consecutive passages that fall outside one standard deviation, thus obscuring true student performance.

Conclusion

In a 2011 paper, Albano and Rodriguez stated "the issue of passage nonequivalence remains unresolved (pg. 44)." In a 2013 paper, Shapiro stated, "Most importantly, the quality of the passage set needs to be carefully considered, and the equivalence of passages cannot be assumed. Careful examination of published passage sets and the data used to substantiate their equivalence must be an important consideration before choosing a passage set. (pg. 65)" In this study we investigated the problem of passage nonequivalence by carefully examining nuanced linguistic features of the text, namely, sub-lexical features of words. We used a new, automated, quantitative measure of decoding difficulty: the DSyM. Our results showed that there was a large amount of variability in decoding difficulty of first-grade passages both across and within publishers, three with passages that fell outside the averaged standard deviation interval. It is important that both practitioners and publishers are aware of text differences in R-CBM passages that might contribute to unintended variance in student performance and lead to decisions based on incorrect data.

CHAPTER II

Examining the Scope & Sequence of Explicit Instruction of Grapheme-Phoneme Correspondences Across a Variety of Reading Curricula

Introduction

Over thirty years of research has shown that systematic and explicit phonics instruction improves children's reading skills and is one of the best methods for ameliorating reading difficulties. Indeed, both Pearson (1999) and Foorman, Francis, Davidson, Harm, and Griffin (2004) note that all of seminal documents in beginning reading (e.g., reading—*Learning to Read: The Great Debate* [Chall, 1967/1996], *Becoming a Nation of Readers* [Anderson, Hiebert, Scott, & Wilkinson, 1985], *Beginning to Read* [Adams, 1990], and the National Research Council's *Preventing Reading Difficulties in Young Children* [Snow, Burns, & Griffin, 1998]) suggest (or outright state) that reading curricula should teach grapheme-phoneme correspondences (GPCs) and allow children to practice taught GPCs on decodable text. However, noticeably lacking from these important reports, is a discussion of *which* GPCs should be taught and *when*.

Historically, researchers have published guides to help practitioners select appropriate research-based reading curricula. For example, the Oregon Reading First Center for Teaching and Learning (http://oregonreadingfirst.uoregon.edu/inst_curr_review.html) lists nearly a hundred interventions that teach phonics and evaluates them on 19 different rules such as "Sequences the introduction of letter sounds, letter combinations, and word parts in combinations, and word parts in ways that minimize confusion." Similarly, the Florida Center for

Reading Research (https://fcrr.org/resources/), and the Institute for Educational Sciences (Foorman, Smith, & Kosanovich, 2017) also publish guides to help practitioners select and evaluate reading curricula. However, the author is not aware of any guide that has examined the scope and sequence of explicit phonics (GPC) instruction across different reading curricula. Explicit phonics instruction refers to the idea that teachers are directly modeling the graphemes and phonemes present in the language (i.e. saying the phoneme out loud while pointing out the specific grapheme that constitutes the sound). Specifically, part b of the definition of explicit phonics instruction used by Mesmer & Griffith, (2005, pg. 369) is also used here; “instruction that is direct, precise, and unambiguous.”

The lack of a research guide examining differences in explicit phonics instruction is interesting given that the English Language Arts Common Core State Standards (ELA CCSSs; Common Core State Standards Initiative, 2010) list over a hundred specific grapheme-phoneme correspondences (GPCs) in Appendix A, and imply that these should be taught in first grade. For example, a first-grader should know that the sound /o/ (long “o”) can be made by the following graphemes: o_e, oa, oe, ow, and o. See Tables 7 (consonants) and 8 (vowels) for a list of the CCSSs GPCs.

While there has been much controversy surrounding the ELA CCSSs and several researchers have stated that more empirical evidence is needed (e.g., Hiebert & Mesmer, 2013; Mesmer, Cunningham, & Hiebert, 2012), the fact remains that nearly 42 states have adopted them, and publishers openly state that they re-structured their curricula to align with them (based on marketing materials put out by the educational publishers). Yet, there is little research showing how reading curricula actually align with the CCSSs, and specifically with regards to explicit phonics instruction.

Table 7

CCSSs Appendix A list of consonants.

Phoneme	Grapheme	example
P	p	pit
B	b	bit
M	m	mitt
M	mb	comb
M	mn	hymn
T	t	tickle
T	tt	mitt
T	ed	sipped
D	d	die
D	ed	loved
N	n	nice
N	kn	knight
N	gn	gnat
K	k	kite
K	c	cup
K	ck	duck
K	ch	chorus
K	lk	folk
K	q	quiet
G	g	girl
G	gh	pittsburgh
NG	ng	sing
NG	n	bank
F	f	fluff
F	ff	fluff
F	gh	tough
F	ph	sphere
F	lf	calf
V	v	van
V	ve	dove
S	s	sit
S	ss	pass
S	sc	science
S	ps	psychic
Z	z	zoo
Z	zz	jazz
Z	se	nose
Z	s	as
Z	x	zylophone
TH	th	thin, breath, ether
TH	th	this, breathe, either
SH	sh	shoe
SH	ss	mission
SH	s	sure
SH	ch	charade
SH	sc	precious
SH	ti	notion
SH	si	mission
SH	ci	special
ZH	s	measure
ZH	z	azure
CH	ch	cheap
CH	t	future
CH	tch	etch
J	j	judge
J	dge	judge
J	ge	wage
L	l	lamb
L	ll	call

L	le	single
R	r	reach
R	wr	wrap
R	er	her
R	ur	fur
R	ir	stir
Y	y	you
Y	u	use
Y	eu	feud
Y	i	onion
W	w	witch
W	qu	queen
WH	wh	where
H	h	house
H	wh	whole

Note. Phonemes are depicted using the Carnegie Mellon Pronunciation Dictionary Rules (http://www.speech.cs.cmu.edu/cgi-bin/cmudict#about).

Table 8
CCSSs Appendix A list of vowels.

Phoneme	Grapheme	example
IY	ee	see
IY	e_e	these
IY	_e	me
IY	ea	eat
IY	ey	key
IY	y	happy
IY	ie	chief
IY	ei	either
IH	i	sit
IH	y	gym
EY	a_e	make
EY	ai	rain
EY	ay	play
EY	ea	great
EY	y	baby
EY	eigh	eight
EY	ei	vein
EY	ey	they
EH	e	bed
EH	ea	breath
AE	a	cat
AY	i_e	time
AY	ie	pie
AY	y	cry
AY	igh	right
AY	i	rifle
AO	o	fox
AO	wa	swap
AO	al	palm
AH	u	cup
AH	o	cover
AH	oo	flood
AH	ou	tough
AA	aw	saw
AA	au	pause
AA	all	call
AA	wa	water
AA	ought	bought
OW	o_e	vote
OW	oa	boat
OW	oe	toe
OW	ow	snow
OW	o	open
UH	oo	took
UH	u	put
UH	ou	could
UW	oo	moo
UW	u_e	tube
UW	ue	blue
UW	ew	chew
UW	ui	suit
UW	ou	soup
Y.UW	u	use
Y.UW	ew	few
Y.UW	u_e	cute
OY	oi	boil
OY	oy	boy
AW	ou	out
AW	ow	cow
ER	er	her
ER	ur	fur
ER	ir	sir

Note. Phonemes are depicted using the Carnegie Mellon Pronunciation Dictionary Rules (http://www.speech.cs.cmu.edu/cgi-bin/cmudict#about).

For example, do the major reading curricula all cover the same GPCs? If they differ, how do they differ?

Rationale

Several reading curricula (core and intervention) have been touted as systematic and explicit by curricula evaluation guides, or effective at improving certain student outcomes by the What Works Clearinghouse, but the author found no research describing how reading programs differ in which GPCs are covered, and when. This aim addresses that need by examining the scope and sequence of explicit GPC instruction in first-grade across a variety of reading curricula.

Research Questions

The following research questions were investigated: (1) Which CCSSs GPCs are explicitly taught (per the instructions to teachers in the teacher edition or scope and sequence) across different first-grade reading curricula? (2) How do the CCSSs GPCs that are taught differ across reading curricula? (3) How does the order of presentation (i.e. the sequence) of GPCs differ across curricula?

Methods

Sample. Since the goal of this chapter was to investigate differences in the scope and sequence across a wide variety of reading curricula, the author chose to examine 12 different programs representing different types of instruction: core and supplemental. Furthermore, the sample was designed to include reading curricula that varied in terms of their evidence of effectiveness for beginning readers as listed by the What Works Clearinghouse (see Table 9 for information on the included reading curricula). Several of the major educational publishers were included (Pearson, McGraw-Hill, Fountas & Pinnell, & Scholastic, etc.). Teacher's editions for

Table 9
Reading curricula publisher information, source of data, and evidence of effectiveness.

Name	Reading Program Information	Source of Information	What Works Clearinghouse Evidence
Explode the Code	Fuertes, C. (2005). Explode the Code for English Language Learners. Cambridge: Educators Publishing Service.	Teacher's Editions, Curriculum Library	No Results Found
Horizons	Engelmann, S., Engelman, O., & Davis, K. (1998). Horizons Learning to Read Teacher Guide Level A. Columbus, Ohio: McGraw Hill.	Teacher's Editions, Curriculum Library	As of July 2007 no studies of Direct Instruction/Horizons were found that fell within the scope of the Beginning Reading review protocol and met WWC design standards. Therefore, the WWC is unable to draw any research based conclusions about the effectiveness or ineffectiveness of Direct Instruction/Horizons to improve outcomes in this area.
Journeys	Tennessee Journeys Common Core Grade 1 Teacher's Edition (2014). Orlando, FL: Houghton Mifflin Harcourt Publishing Company.	Teacher's Editions, Curriculum Library	No Results Found
Levelled Literacy Intervention	Fountas, I. C., & Pinnell, G. S. (2009). Leveled Literacy Intervention: Green System, Levels AJ, Lessons 1-110. Boston, MA: Houghton-Mifflin Harcourt.	Teacher's Editions, Curriculum Library	Yes (Leveled Literacy Intervention had positive effects on general reading achievement, potentially positive effects on reading fluency, and no discernible effects on alphabetics for beginning readers.)
Open Court	Open Court Introduction to Sound/Spellings (2018). New York City, NY: MacMillan/McGraw-Hill.	Online scope and sequence*	*Open Court Reading*© was found to have potentially positive effects on general reading achievement and comprehension for beginning readers.
Phonics from A to Z	Blevins, W. (1998). Phonics from A to Z: A practical guide. New York, NY: Scholastic Inc.	Teacher's Editions, Curriculum Library	No Results Found
Reading Mastery	Engelman, S., & Bruner, E., (1995). Reading Mastery I Rainbow Edition Teacher's Guide.	Online scope and sequence*	No studies of *Reading Mastery* that fall within the scope of the Beginning Reading review protocol meet What Works Clearinghouse (WWC) evidence standards. The lack of studies meeting WWC evidence standards means that, at this time, the WWC is unable to draw any conclusions based on research about the effectiveness or ineffectiveness of *Reading Mastery*on beginning readers in grades K–3. Additional research is needed to determine the effectiveness or ineffectiveness of this intervention.
Reading Street (2008)	Scott Foresman Reading Street Common Core (2008). London, United Kingdom: Pearson.	Teacher's Editions, Curriculum Library	No Results Found
Reading Street (2013)	Scott Foresman Reading Street Common Core Grades K-6 (2013). London, United Kingdom: Pearson.	Teacher's Editions, Curriculum Library	No Results Found
Road to Reading	Blachman, B., & Tangel, D. (2008) Road to Reading: A Program for Preventing & Remediating Reading Difficulties. Baltimore, MD: Paul H. Brookes, Inc.	Purchased Personal Copy	No Results Found
S.P.I.R.E.	Clark-Edmands, S. (2005). S.P.I.R.E. A Specialized Program Individualizing Reading Excellence. Level 1 Teacher's Guide. Second Edition. Cambridge: Educators Publishing Service.	Teacher's Editions, Curriculum Library	As of July 2007 no studies of S.P.I.R.E. were found that fell within the scope of the Beginning Reading review protocol and met WWC design standards. Therefore, the WWC is unable to draw any research based conclusions about the effectiveness or ineffectiveness of S.P.I.R.E. to improve outcomes in this area
Treasures	Tennessee Treasures, A Reading Language Arts Program, Teacher's Edition Grade 1(2008). New York City, NY: MacMillan/McGraw-Hill.	Teacher's Editions, Curriculum Library	No Results Found
Wonders	Wonders: A K-6 Literacy Curriculum (2020). New York City, NY: MacMillan/McGraw-Hill.	Online scope and sequence*	No Results Found

Note. *The online scope and sequences were used when it was cost-prohibitive to purchase the teacher's editions. Please refer to the methods section for more information on how each was located. In order to determine What Works Clearinghouse Evidence, the search tool under 'Intervention Reports' was used https://ies.ed.gov/ncee/wwc/Publication#/ContentTypeId:1. The Common Core State Standards final version was released in June 2010 (http://www.corestandards.org/about-the-standards/development-process/ and states started to replace their standards in the subsequent years. As of 2015, 42 states had adopted them.

first grade core and supplemental reading curricula were located through the campus curriculum library, via interlibrary loan, or purchased. In some cases, purchasing a new set of teacher's editions was cost prohibitive so the phonics scope and sequence was located online or by e-mailing the publisher if it could not be secured via interlibrary loan (see Table 9 for the source of information). When the scope and sequence was ambiguous (and many were), the first author reached out for clarification via e-mail to the publisher. These e-mail exchanges did not always result in clarification of which specific GPC was being taught. In this case, the benefit of the doubt was given to the publisher. Due to the somewhat ambiguous nature of these decisions, a second coder was used.

Recording of GPCs. A comprehensive list of GPCs in the English language was modeled after Berndt, Reggia, and Mitchum (1987, Appendix B). The Berndt (1987) list of GPCs was used as a reference because it was the most comprehensive list of GPCs the author was able to locate, comprised of 369 GPCs in the English language. GPCs that were explicitly taught were given an ordered number (based on when it was taught in the reading curricula) next to the appropriate GPC on the Berndt list. Only GPCs that were explicitly taught were assigned a number. For example, if a word was listed in the lesson plan that contained a certain GPC, but the actual lesson plan did not indicate that the teacher state the sound while indicating the letter(s) (i.e. pointing, underlining, etc.) then this particular GPC was not counted as being explicitly taught. An example of this can be found on pg. 162 of the Green System Lesson Guide for lessons 1-60. The word "our" is listed as a necessary word card for the lesson, and students practice building it, but nowhere in the lesson does it instruct teachers to indicate that the letters "ou" combine to make the sound /OW/.

High frequency words or sight words that the curricula instructed children to memorize were not counted as being explicitly taught since the graphemes were not pointed out nor were the individual phonemes modeled.

Several programs taught sub-lexical units of words that were larger than graphemes. For example, S.P.I.R.E. teaches "-old" and "-ost", but these are redundant with the GPC for o - /long o/. Therefore, if the curricula previously covered those, then they were not counted again. This also brings up an important point about coverage. This investigation recorded the first use, not the total coverage (time spent across the curricula) teaching a particular GPC.

Inter-rater reliability coding. A doctoral student majoring in special education served as a second coder and inter-rater reliability was calculated. While the authors initially felt that the need for an additional coder should be obsolete because the GPCs should be *explicitly* listed in the teacher edition or scope and sequence and, therefore; easy to interpret, we quickly found out this was not the case. For example, the teacher's editions often explicitly taught sub-lexical "chunks" of words that were larger than GPCs. Extra care was warranted in these situations so that the appropriate GPCs could be parsed and made sure they were not redundant with earlier taught GPCS.

Another issue that frequently arose was the fact that the online scope and sequence that was used in lieu of purchasing a new set of teacher's editions did not explicitly state which GPCs were being covered. For example, in Reading Street's (Pearson, 2013) online scope and sequence (page 3, for Grade 1, unit 2, week 3) it states "consonant digraphs wh, ch, tch" however it does not state which *sounds* are covered. For example, the grapheme 'wh' can make the sound /H/ as in 'whole' or /W/ as in 'what'. However, the scope and sequence does not state which of the sounds they are teaching. There are other instances of this in the scope and sequence as well

(Unit 2, week 1, ’th’ is taught but it is unclear if they are referring to the voiced or unvoiced /TH/. For the aforementioned reasons, the author decided a second coder was necessary. Furthermore, all the data is available upon request from the author.

Analysis

Calculating inter-rater reliability. Inter-rater reliability for scope (i.e. which GPCs were covered) was assessed by calculating both percent agreement and an unweighted Cohen’s kappa, the latter accounts for chance agreement. Percent agreement was calculated by dividing total agreements by total cases. Kappa was calculated using the ‘kappa2’ function from the package ‘irr’ (Gamer, 2019) in R version 3.5.3 (R Core Team, 2019).

Analyzing differences in scope and sequence. To determine differences in alignment between each curricula’s GPCs and the CCSSs list of GPCs, percentages were calculated. CCSSs GPCs were also compared to a larger corpus of GPCs in the English language as delineated by Berndt et al. (1987).

Differences in terms of scope were analyzed using percentages (descriptive analysis instead of inferential statistics) since the entire population (of the listed first-grade curricula) were analyzed.

Spearman (rank-order) correlation coefficients were calculated for all 66 pairwise comparisons between the reading curricula’s order of GPC presentation. These were calculated in R version 3.5.3 (R Core Development Team, 2019) using the package ‘stats’. A correlation plot indicating the significant correlations was also created in the r package ‘stats’. A Bonferroni correction of p <.008 was used to adjust the alpha level for multiple comparisons.

The similarities and differences in terms of both scope and sequence can be seen in Figure 4. Figure 4 lists the CCSSs as well as GPCs and shows several key pieces of information:

(1) whether a particular GPC was covered across reading curricula, (2) the order is was covered (indicated by the number in the cells where a higher number indicates a later introduction in the sequence) and (3) differences in sequence across reading curricula (dark blue represents early exposure in a given curriculum whereas dark red represents later exposure in the curriculum). The CCSSs do not intentionally order the list of GPCs (i.e. they do not outline a specific sequence of introduction). Therefore, GPCs were grouped by the author into the following categories: consonants, short vowels, long vowels, digraphs, dipthongs, and other (if the GPC did not neatly fit into one of the aforementioned categories).

Results

Inter-rater reliability. Results of inter-rater reliability can be seen in Table 10. Percent agreement was high and ranged from 92% (Explode the Code, Road to Reading) to 98% (Horizons). Cohen's kappa values were also high and ranged from .62 to .94 (Cohen's kappa values ≤ 0 indicate no agreement and 0.01–0.20 as none to slight, 0.21–0.40 as fair, 0.41– 0.60 as moderate, 0.61–0.80 as substantial, and 0.81–1.00 as almost perfect agreement). All disagreements were discussed and resolved via consensus coding.

Scope. Table 11 lists the total number of GPCs covered as well as the percent of the CCSSs GPCs covered and the percent of a more comprehensive list (Berndt et al., 1987) of GPCs in the English language. The number of GPCs covered on the Berndt et al. (1987) list ranged from 36 (Explode the Code) to 102 (Open Court, 2018). The number of CCSSs GPCs covered ranged from 32 (Explode the Code) to 89 (Open Court).

None of the reading curricula examined in this paper covered all of the CCSSs GPCs. The following CCSSs were not explicitly taught in any of the reading curricula: s_SH ("**s**ure"), s_ZH ("mea**s**ure"), t_CH ("fu**t**ure"), z_ZH ("a**z**ure"), a_AA ("w**a**ter"), a_AO ("sw**a**p"), ey_EY

Table 10
Inter-rater reliability statistics.

	Levelled Literacy Intervention	Explode the Code	S.P.I.R.E. Levels 1-5	Phonics from A to Z	Horizons	Reading Street (2008)	Road to Reading	Open Court	Reading Street (2013)	Treasures	Journeys	Wonders	Reading Mastery
Percent Agreement	96%	92%	95%	95%	99%	95%	92%	95%	96%	90%	94%	93%	98%
Kappa (Z value & P-value)	.87 (16.7, 0)	.62 (12.2, 0)	.82 (15.8, 0)	.86 (16.5, 0)	.94 (18.1, 0)	.87 (16.6, 0)	.78 (14.9, 0)	.88 (16.9, 0)	.88 (16.8, 0)	.66 (12.8, 0)	.81 (15.6, 0)	.77 (14.8, 0)	.90 (17.2, 0)

Table 11

Reading curricula scope of GPCs covered

	Reading Curricula											
	Core basal programs						Supplemental programs					
	*Wonders	Treasures	Journeys	*Open Court	Reading Street 2013	LLI	Horizons	Road to Reading	Reading Mastery	Explode the Code	S.P.I.R.E. Levels 1--5	Phonics A to Z
Total # of GPCs covered	83	83	84	102	92	62	52	82	44	36	60	85
Percent of Berndt et al. (1987) GPCs covered	22%	22%	23%	28%	25%	17%	14%	22%	12%	10%	16%	23%
Total # of CCSSs GPCs covered	76	67	78	89	88	58	48	74	43	32	54	75
Percent of CCSSs GPCs covered	59%	52%	61%	70%	69%	45%	38%	58%	34%	25%	42%	59%

Note. LLI = Levelled Literacy Intervention. *asterisk denotes information from a published scope and sequence, rather than the teacher's editions (when they were not available via interlibrary loan and purchasing a new set was cost prohibitive.

("th**ey**"), gh_F ("tou**gh**"), gh_G ("Pittsbur**gh**"), lf_F ("ca**lf**"), lk_K ("fo**lk**"), ps_S ("**ps**ychic"), sc_S ("**sc**ience"), c_SH ("pre**c**ious"), **si**_SH ("mission"), **ti**_SH, ("no**ti**on"), ei_EY ("v**ei**n"), ei_IY ("**ei**ther"), eigh_EY ("**eigh**t"), eu_Y.UW ("f**eu**d"), i_Y ("on**i**on").

Large differences in the scope of GPCs covered could be explained in terms of whether the curricula is considered core or supplemental, with supplemental programs covering less GPCs than core basal programs. The core basal reading programs (Treasures, Journeys, Wonders, Reading Street, Open Court, and LLI) generally covered more of the CCSSs GPCs than the intervention/supplemental programs, which is arguably, to be expected, as the latter are mainly to supplement a core curriculum. Several of these supplemental reading programs (e.g., Horizons, Reading Mastery) are designed as interventions, rather than core instruction, and children are often placed into them depending on their level. Indeed, all of the supplemental programs are designed such that the teacher is given ample leeway when determining where to start, based on student need. In this sense, it is assumed that different children will have received different GPC instruction, individualized to their unique needs. This is in contrast to the core/basal programs in which students usually start and the beginning of the curriculum and move through at the same pace.

There were no clear patterns in terms of time of publication (pre-CCSSs adoption versus post). However, Wonders, the most recently published reading curricula (2020 copyright year), marked a change in the previous core/basal programs in that its scope and sequence began explicitly teaching GPCs in kindergarten, some of which were not explicitly re-taught in first grade. Rather, it was assumed that a first-grade child would have had prior exposure to certain GPCs. This is in contrast to core programs published in earlier years such as Treasures and

Journeys, which, even if they covered some GPCs in kindergarten, were all explicitly re-taught in first grade.

Furthermore, no clear relationship is seen between percentage of GPCs covered and evidence of effectiveness as determined by the What Works Clearinghouse. Specifically, both Open Court and Levelled Literacy Intervention (LLI) are listed as having positive or potentially positive effects on at least one component of reading, yet while Open Court covered 70% of the CCSSs, LLI covered less than half of them (46%). That said, the determination of evidence of effectiveness by the What Works Clearinghouse depends on several factors, not least of which is the presence of research studies using a certain reading curriculum. It could simply be the case that there were not enough (or any) research studies available to examine the effectiveness. Interestingly, there were GPCs covered by several reading curricula that were not listed in the CCSSs. For example, 9 of the reading curricula explicitly taught the GPC 'x' - /KS/ (as in the word "axe"), but this was not listed by the CCSSs. Similar results are found for other popular GPCs: several reading programs explicitly teach 'c' - /S/ and 'a' and 'e' as the schwa sound /UH/, but these were not on the list of GPCs covered by the CCSSs.

Sequence. There were several large and statistically significant correlations between reading curricula in terms of the order that CCSSs GPCs were introduced (see Table 12 for correlations and Figure 3 for the correlation plot). In general, there were large, statistically significant correlations among the core/basal publishers: Wonders, Reading Street, Treasures, Open Court, Journeys, and Levelled Literacy Intervention. Road to Reading, S.P.I.R.E., and Phonics A to Z, while all more supplemental programs, still had large and statistically significant correlations with the core/basal publishers, suggesting they introduce GPCs in a similar sequence.

Table 12

Spearman correlation values for the sequence of GPC introduction across publishers.

Reading Curricula	Reading Street	Treasures	Open Court 2018	Journeys	LLI	Horizons	Road to Reading	Explode the Code	S.P.I.R.E. Levels 1-5	Phonics A to Z	Reading Mastery
Wonders	0.57*	0.58*	0.57*	0.65*	0.55*	0.08	0.57*	0.52	0.60	0.77*	0.23
Reading Street		0.95*	0.97*	0.90*	0.77*	-0.08	0.87*	0.65	0.95*	0.70*	0.35
Treasures			0.98*	0.98*	0.68*	-0.05	0.92*	0.73	0.93*	0.72*	0.52
Open Court 2018				0.95*	0.72*	-0.10	0.88*	0.68	0.97*	0.73*	0.50
Journeys					0.67*	-0.07	0.93*	0.72	0.92*	0.68*	0.50
LLI						-0.03	0.83*	0.68	0.82*	0.45*	0.17
Horizons							-0.10	0.33	-0.18	0.20	0.58
Road to Reading								0.77	0.90*	0.52*	0.35
Explode the Code									0.60	0.68	0.62
S.P.I.R.E. Levels										0.63*	0.37
Phonics A to Z											0.57*
Reading Mastery											

Note. * denotes correlations significant at $p < .008$ Bonferroni correction.

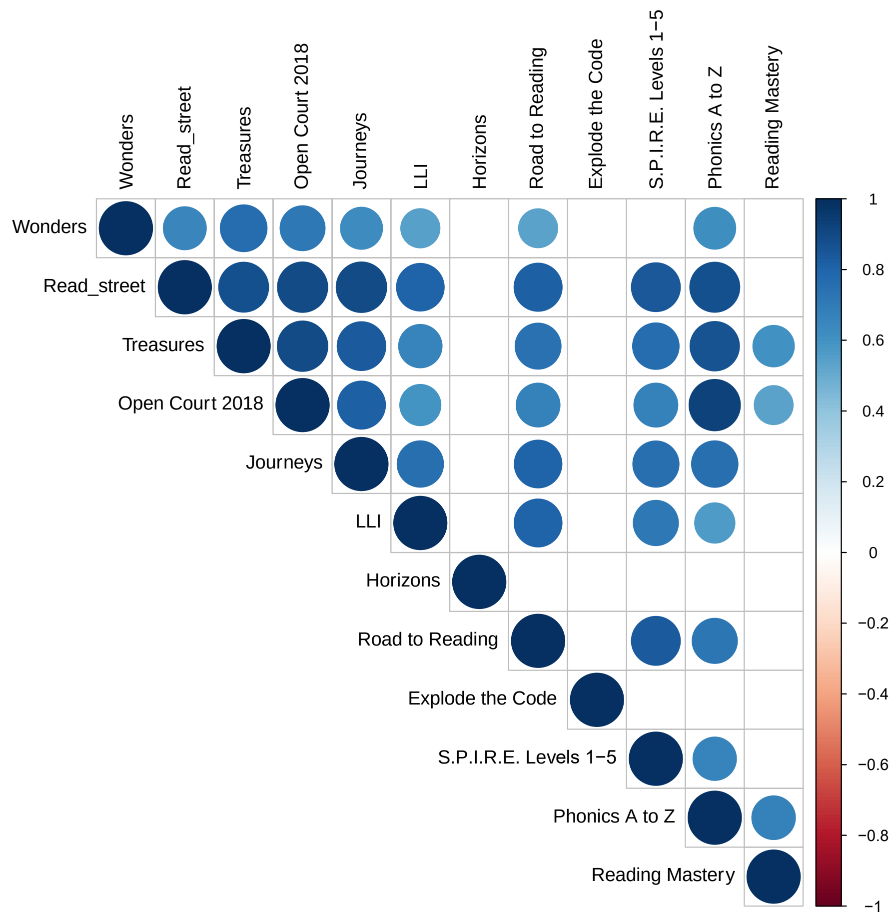
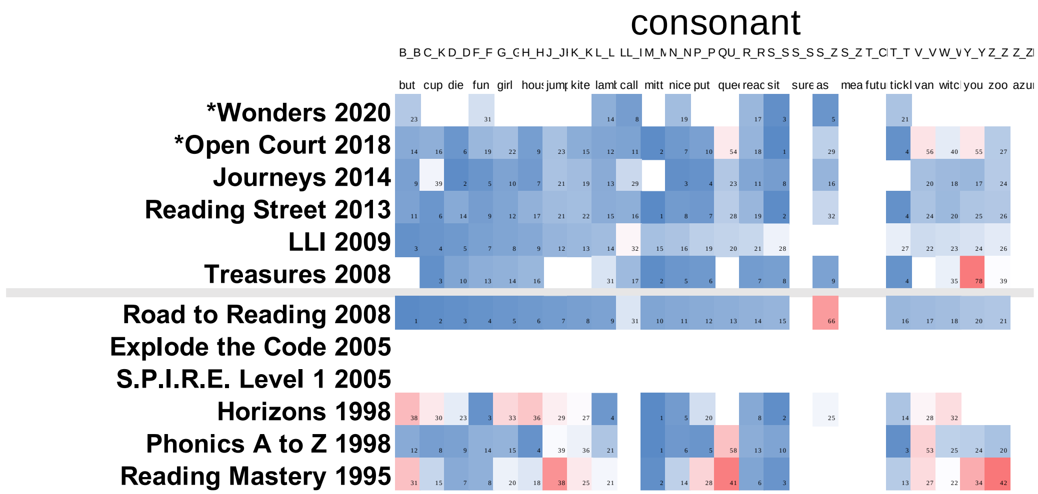
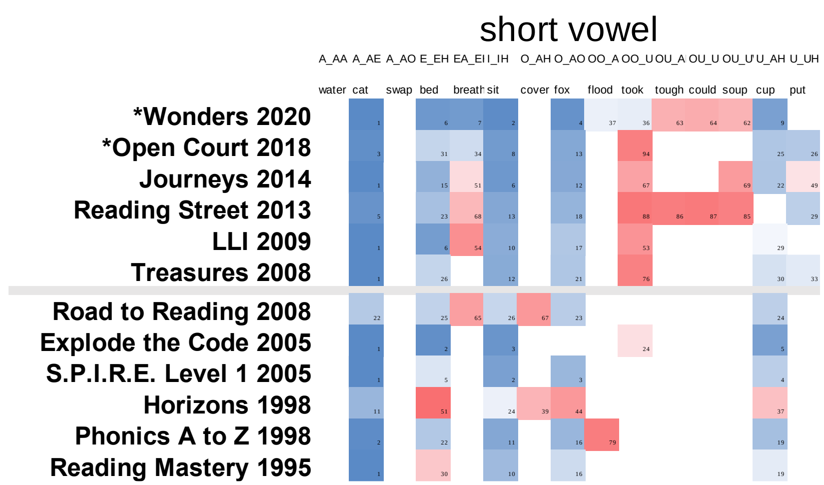


Figure 3. Significant Spearman correlations (at p < .008 correction) between sequence of GPC introduction across publishers.

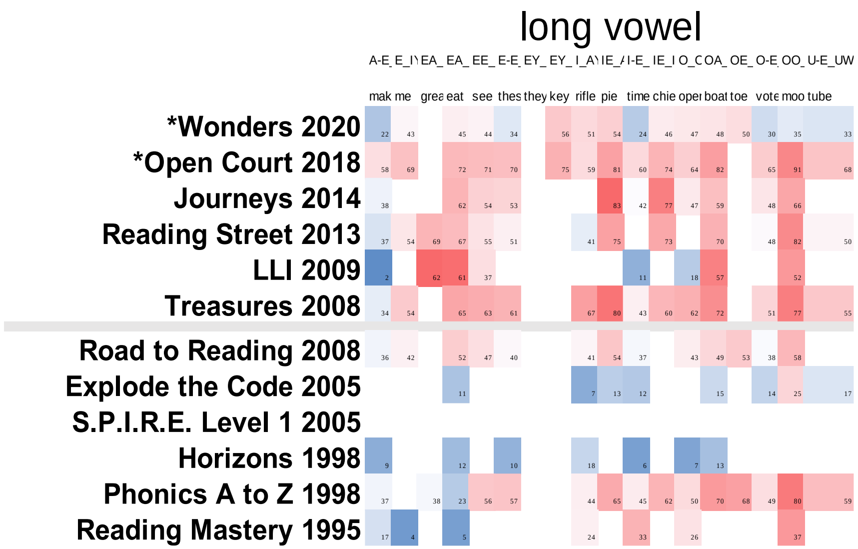
(a)


consonant
*Wonders 2020
*Open Court 2018
Journeys 2014
Reading Street 2013
LLI 2009
Treasures 2008
Road to Reading 2008
Explode the Code 2005
S.P.I.R.E. Level 1 2005
Horizons 1998
Phonics A to Z 1998
Reading Mastery 1995

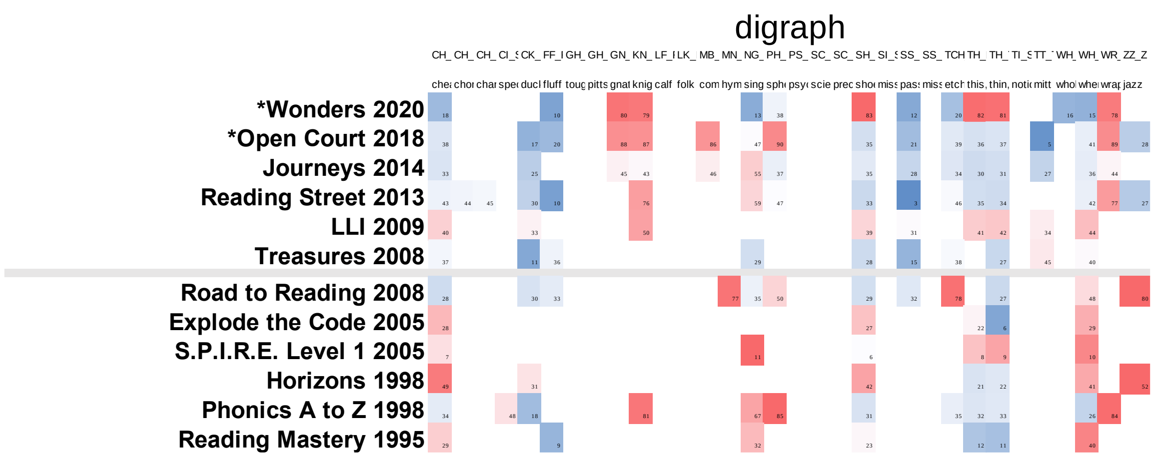
(b)


short vowel
A_AA A_AE A_AO E_EH EA_EH I_IH O_AH O_AO OO_A OO_U OU_A OU_U OU_U U_AH U_UH
water cat swap bed breath sit cover fox flood took tough could soup cup put
*Wonders 2020
*Open Court 2018
Journeys 2014
Reading Street 2013
LLI 2009
Treasures 2008
Road to Reading 2008
Explode the Code 2005
S.P.I.R.E. Level 1 2005
Horizons 1998
Phonics A to Z 1998
Reading Mastery 1995

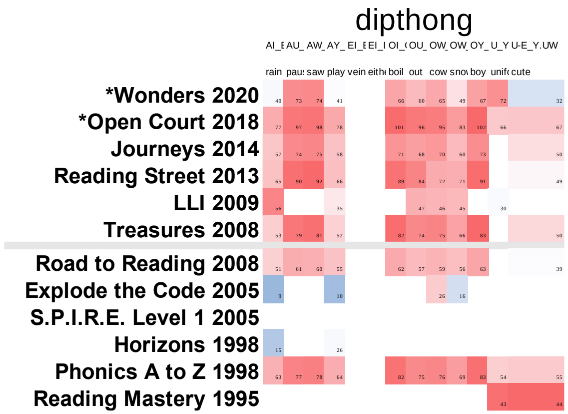
(c)


long vowel
*Wonders 2020
*Open Court 2018
Journeys 2014
Reading Street 2013
LLI 2009
Treasures 2008
Road to Reading 2008
Explode the Code 2005
S.P.I.R.E. Level 1 2005
Horizons 1998
Phonics A to Z 1998
Reading Mastery 1995

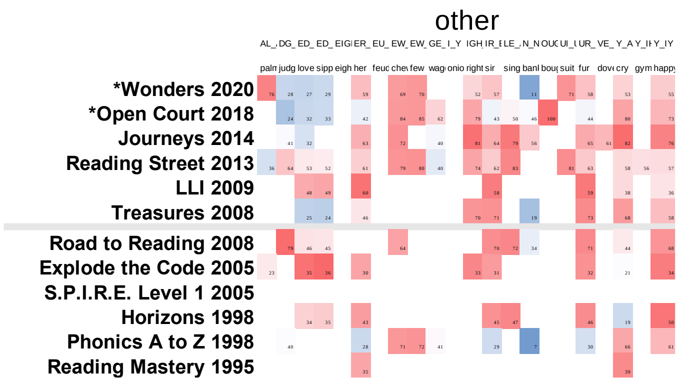
(d)


digraph
*Wonders 2020
*Open Court 2018
Journeys 2014
Reading Street 2013
LLI 2009
Treasures 2008
Road to Reading 2008
Explode the Code 2005
S.P.I.R.E. Level 1 2005
Horizons 1998
Phonics A to Z 1998
Reading Mastery 1995

(e)



(f)



(g) all

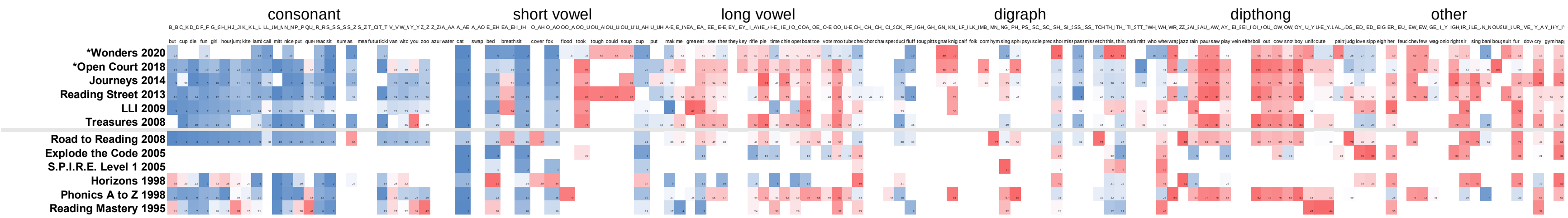


Figure 4. Scope and sequence of CCSS GPCs across different reading curricula.

Whereas Figure 3 gives an overall number of similarity (Spearman correlation coefficient), Figure 4 (graphs a through f) highlights specific differences in sequence across the different reading curricula. The blue-colored cells denote that a GPC was introduced early in the sequence whereas red indicates that particular GPC was introduced toward the end of the sequence. The saturation of the colors corresponds to the numbered order of presentation that is also displayed in the individual cells. From looking at graphs (a) through (f) in Figure 4, general patterns emerge. For example, generally speaking, most of the curricula introduce consonants and certain short vowels early in their respective sequences. That said, some short vowels such as oo_UH as in “took” or ou_AH as in “tough” were introduced much later in the sequence. Long vowels are typically introduced later as evidenced by the mostly red cells in graph (c). However, Horizons, Explode the Code, and Reading Mastery start introducing long vowels earlier than the other curricula. There was a lot of variation in the introduction of digraphs (two letters that make one sound) as evidenced by the equal numbers of red and blue in graph (c). Double consonant digraphs such as ss_S (“pass”) and ff_F (“fluff”) were introduced early whereas digraphs where two different letters such as kn_N (“knight”) and ph_F (“sphere”) were introduced much later (if at all).

Discussion

Summary & practical implications. The results of this investigation show that there were substantial differences in terms of the number of CCSS GPCs covered (scope) as well as the order they were covered (sequence) across the reading curricula examined. Large differences in scope could be explained by whether the curricula is designed to be implemented as core or supplemental, with core basal reading curricula covering more than the supplemental programs.

In terms of sequence, generally, most curricula covered consonants and short vowels early, saving long vowels, digraphs, and dipthongs for later in the sequence. That said, there was quite a lot of variation in terms of presentation of GPCs.

In terms of practical significance, the results are perhaps of most practical importance for publishers that are constructing their curricula to adhere to the CCSSs GPCs and to school districts selecting which reading curricula covers the most (or certain, specific) GPCs. Furthermore, these results could shed light on variation in children's decoding skill (and inform which GPCs to teach next). For example, consider the case when a child switches curricula (perhaps by switching schools) during their first year of reading instruction.

The results can also help practitioners select supplemental curricula that is aligned (similar scope and sequence) with a particular student's core/basal instruction. There is initial evidence that students who receive supplemental instruction that reinforces phonic concepts being learned in their core instruction made greater gains than students who received supplemental instruction in new, but related, material (Wonder-McDowell, Reutzel, & Smith, 2011).

Limitations & future directions. Since the focus of this investigation was on the scope, more so than the sequence of GPCs (we wanted to know, if, by the end of first grade the same GPCs were covered) it did not determine the *percentage* of time that was devoted to the teaching of a particular GPC. This is important, as curricula differed in whether they repeated or reviewed instruction after it was presented. For example, LLI introduced GPCs and then spent several subsequent lessons reviewing the material.

Another future direction could examine the match between the teacher edition and the student material, such as the texts that the student is given to practice. This investigation simply looked at the teacher edition, not the corresponding text to determine if a GPC was covered.

Yet another future study would be to examine the errors and omissions in GPC reporting, specifically in the scope and sequences. Given that these curricula area several thousand dollars, it is not unreasonable to ask for a comprehensive scope and sequence before making a purchase. However, several of the scope and sequences were incomplete, confusing, or contained incorrect information. Recording discrepancies such as these would be important so that the curriculum developers do not unintentionally propagate incorrect information.

Conclusion

Nearly 30% of U.S. children are not reading at a proficient level (U.S. Department of Education, 2017), suggesting that there is substantial room for improving the reading skills of young children. Over thirty years of research has shown that systematic, explicit phonics instruction improves children's reading skills, yet national reading scores have remained relatively stable. Perhaps examining sources of variation in reading curricula at a more granular level (the level of GPCs) could help explain sources of variation in reading achievement. Several guides exist to help teachers and practitioners choose research-based reading curricula yet there is no existing comparison of reading curricula scope and sequence in terms of grapheme-phoneme correspondences. This paper systematically analyzed the scope and sequence of explicit GPC instruction across several core and supplemental reading curricula. These results show variation in terms of both scope and sequence, with none of the reading programs covering even 75% of the CCSSs GPCs. These results have the potential to inform several areas:

practitioners choosing curricula, publishers designing curricula, or researchers trying to explain individual differences in decoding skill due to instructional exposure to GPCs.

CHAPTER III

Quantifying the Alignment Between First-Grade Reading Instruction and Reading Curriculum Based Measurement Passages on the Grapheme-Phoneme Correspondences listed by the Common Core State Standards

Introduction

It has been three decades since explicit phonics instruction was put forth as a research-based practice, yet national reading scores remain largely unchanged in thirty years: nearly 60% of 4th graders do not read at a proficient level (U.S. Department of Education, 2017). More recently, initiatives such as multi-tiered systems of support (MTSS), response to intervention (RTI), and the new Common Core State Standards (CCSS) have been widely adopted despite clear research, or guidelines, on how all these pieces should fit together (i.e. be implemented). With regards to reading instruction, the interplay of these entities can be perplexing. For example, in an RTI model (researchers state) teachers should be using high-quality reading curricula (created by educational publishers to match national standards and vetted by state education agencies and WWC), and screening all children, using (educational publisher created) progress monitoring probes that have been verified by (predominantly researcher-based groups) such as the National Center for Intensive Intervention. While researcher vetting of products and practices is certainly beneficial to stakeholders, the oversight often ends with a list of approved products. There is little guidance as to whether research-based curricula can be used with all the types of research-based assessment. For example, can any basal reading program be used with any R-CBM? WWC examines evidence of effectiveness for reading programs, but the NCLII examined validity and reliability of R-CBMs. It appears that there is not a group that examines both reading curricula and R-CBMs, despite both components being necessary in an RTI system. This is unfortunate given that high-stakes decisions are being made without data on one of the most basic questions: are we testing what we are teaching? Or, put another way: is there alignment between instruction and assessment? Therefore, this chapter (Aim 3) quantifies the match between the scope of the CCSSs GPCs found in the basal/core reading curricula (the results of Aim 2) and the first-grade ORF R-CBM passages (from Aim 1). See Figure 5 for an overview of how the aims fit together.

The subsequent paragraphs provide an overview on the inter-relationship of Response to Intervention, Progress Monitoring, and a specific type of progress monitoring: Curriculum-Based Measurement. An overview of the research on CBM is presented, highlighting an important gap in the literature: the extent that CBM passages are aligned with instruction. The introduction culminates with the rationale that alignment is a critical yet underexplored factor in oral reading fluency CBMs.

Progress monitoring. Progress monitoring refers to the process of repeated measurement of student academic performance. It is one of the four essential components of Response to Intervention according to the National Center on Response to Intervention. Progress Monitoring can be used for several purposes: to assess students' academic performance over time, to quantify student rates of improvement or responsiveness to instruction, and to evaluate instructional effectiveness (National Center on Response to Intervention, 2013). All children can benefit from progress monitoring, not just those who are receiving services and it can be implemented for individual students or an entire class. Data collection is critical to the validity of the process, and the National Center on Response to Intervention states that assessment should

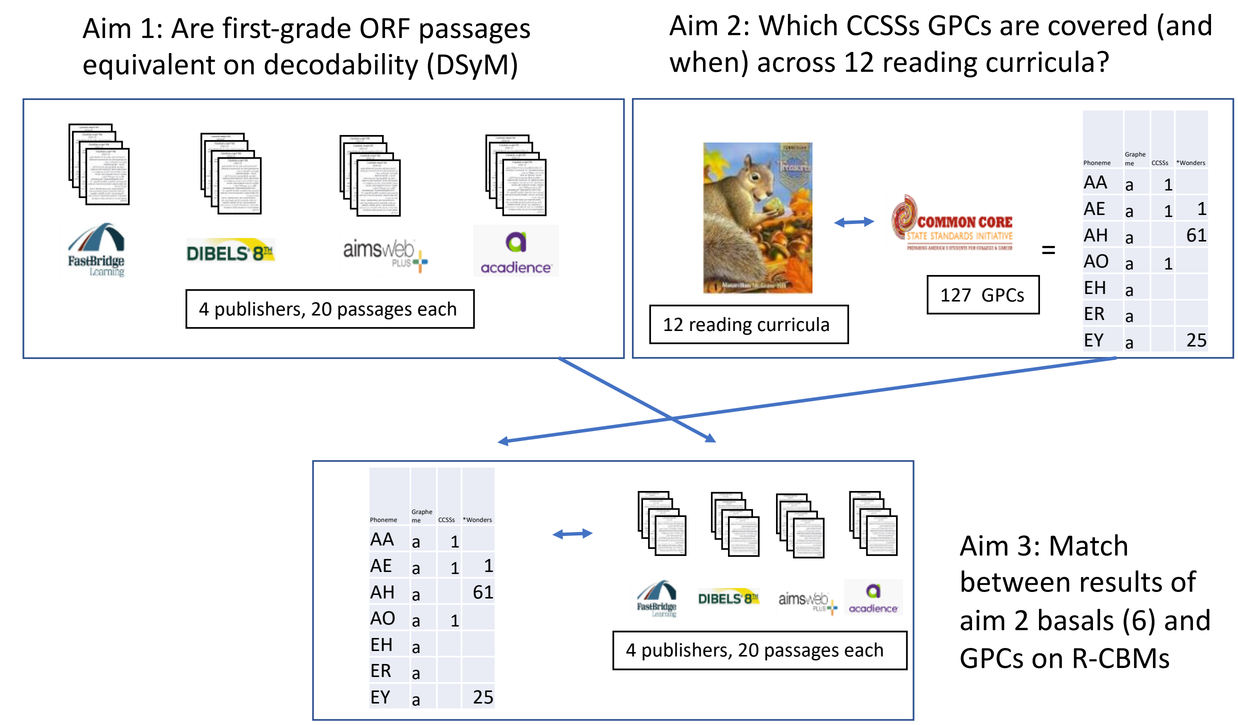


Figure 5. Overview of how Aim 3 relates to Aims 1 & 2.

be performed weekly, if possible, and at the very least, monthly (National Center on Response to Intervention, 2013).

Research-based progress monitoring: Reading Curriculum-Based Measurement. Of the many tools used for progress monitoring, curriculum-based measurement is one of the most researched frameworks. Over 200 empirical studies on CBM have been published in peer-reviewed journals and the findings are consistent: CBM is a valid and reliable framework for "assessing the development of competence in reading..." (Fuchs & Fuchs, 2001 pg. 1). CBM was more effective than informal teacher observations because the process was standardized and easy to implement (Fuchs, Deno, & Mirkin, 1984). In addition to the standardized process for administration and scoring, the passages themselves, were also standardized.

A key development in the history of R-CBM was the finding that oral word reading passages sampled outside of a particular student's curriculum were still valid (e.g., Fuchs & Deno, 1994). This led to the development of grade-appropriate sets of tests that were carefully crafted by publishers to function as alternate forms. Standardized sets of passages that could be purchased by a school saved teachers time and standardized R-CBM passages such as DIBELS, FastBridge, etc. have become the norm in schools. Indeed, CBM has become a "widely used" and "indispensable" form assessment in schools that embrace Multi-Tiered Systems of Support (Bieber & Choi, 2004; O'Keeffe, Bundock, Kladis, Yan, & Nelson, 2017).

In terms of reading, the choice of which progress monitoring assessment is determined by the skill level of the child. Generally, word reading and passage reading tests are appropriate for early elementary grades, while comprehension tests like maze become more important in the upper elementary grades. Passage reading fluency R-CBM is perhaps the most researched of all the CBM available. In passage reading fluency CBM students are asked to read a short passage of text out loud. The number of words read correctly in one minute (WCPM) is the dependent variable. WCPM has been shown to a be both valid, reliable, and sensitive to change (e.g., Fuchs, Fuchs, Hosp & Jenkins, 2001). While WCPM has been shown to be a valid measure of general reading ability (Fuchs et al., 2001), it is also known that WCPM can be influenced by different text factors (e.g., Barth, Tolar, Fletcher, & Francis, 2014).

One factor known to contribute to text difficulty (as measured by student ORF performance) is decoding complexity, or 'decodability' of a text (Mesmer et al., 2012). Decoding complexity of text is especially important to measure and control for in early grade texts (Mesmer et al., 2012), as it is arguably the main reading skill that children are acquiring at this stage (Ehri, 1995; National Reading Panel, 2000). There is substantial evidence to support the idea that emerging readers engage in slow, deliberate processing of sub-lexical (aka 'intra-word') variables, such as grapheme-phoneme correspondences, rather than coarser text variables. Adams (1994), Ehri (1995), and Perfetti (1992) all acknowledge that beginning readers differ from skilled readers in their sensitivity to letter information. This

The Common Core State Standards (CCSSS; Common Core State Standards Initiative, 2010 Appendix A) for English Language Arts list over a 100 specific GPCs that should be covered in first grade and have been adopted in over 45 states (http://www.ascd.org/common-core-state-standards/common-core-state-standards-adoption-map.aspx), influencing the design of reading curricula. Yet, it is unclear to what extent there is overlap in in the content of explicit phonics instruction, and the formative assessment passages on the grapheme-phoneme correspondences (GPCs) listed by the CCSSs. Since the publishers of core reading curricula (Pearson, Houghton Mifflin Harcourt, etc.) are different than the publishers of Reading

Curriculum Based Measurement, or; R-CBM (DIBELS; University of Oregon; FastBridge, etc.), one might expect some variability in coverage.

Alignment between assessment (R-CBMs) and instruction (basal reading programs): An unanswered question. Progress monitoring is just one part of the RTI model. The Institute for Education Sciences Practice Guide for Response to Intervention (RTI: Berkeley, Bender, Peaster, & Saunders, 2009, Gersten et al., 2008) recommends universal screening, high quality tier one core reading program and formative progress monitoring. That said, Tindal (2013, pg. 13) states that systemic research on RTI "is more conceptual than actual." One area where systematic research is lacking is the alignment between two of the RTI components listed above: the tier one core reading program and formative assessment in the form of R-CBM.

Previous research on alignment in reading has focused on several different areas. For example, the alignment between research-based practices and curricula that incorporate them is delineated in guides put out by the University of Oregon (http://oregonreadingfirst.-uoregon.edu/inst_curr _review.html) and the Florida Center for Reading Research (https://fcrr.org/resources/). Alignment has also been examined between core and supplemental reading instruction on student outcomes for students who were receiving both (Wonder-McDowell, Reutzel & Smith, 2011). Other research has examined the alignment between teacher editions and student text on explicit phonics instruction (Stein, Johnson, & Gutlohn, 1999; Toyama & Hiebert, 2019) and the percent of decodable text (Foorman, Francis, Davidson, Harm, & Griffin, 2004). Less researched is the alignment between assessment and instruction in reading, and no research (that the author is aware of) examines the alignment between instruction and assessment at the grapheme-phoneme level.

Alignment between instruction and assessment is important for making accurate data-based decisions such as determining whether a child is responding to instruction. Indeed, a key tenet of R-CBM is that each test (or passage), "…assess all the different skills covered in the annual curriculum. CBM samples the many skills in the annual curriculum in such a way that each weekly test is an alternate form (with different test items, but of equivalent difficulty) (Fuchs & Fuchs, 2001).

Despite abundant research on R-CBM the psychometric properties of R-CBM passages have recently come under scrutiny. Specifically, some researchers have found large amounts of passage variance, undermining the notion that passages actually serve as alternate forms. A full discussion of passage equivalency is beyond the scope of this paper, but the reader is referred to chapter 1 (aim 1 of this dissertation). Rather, the focus of this aim is on the notion that each test, or passage, should sample the full spectrum of the annual curriculum.

Rationale

While the R-CBM framework of assessment provides a simple indicator of general reading ability (Fuchs, 2017), one might question if looking at such a granular level as GPCs is warranted. There are several points to make in response to this argument. First, decoding (i.e. learning the phonemes that map onto the graphemes) is the primary goal of first grade reading instruction. This is evident in both theoretical models of reading development (Adams, 1994; Ehri, 1995; National Reading Panel, 2000; Perfetti, 1992) as well as by looking at the content in the major basal reading curricula (see Aim 2). Second, the widely adopted CCSSs list specific GPCs, whereas policy in the past did not. Third, examining differences at such a granular level could help inform individual differences in response to instruction. Finally, given that nearly 60% of fourth graders are still struggling to read proficiently (U.S. Department of Education, 2017), it could be argued that nothing is too granular when it comes to resistant readers. It has been nearly 30 years since systematic, explicit phonics instruction came onto the scene, yet national reading scores still have not improved. Therefore, one could argue it is time to start exploring other sources of variability in student performance.

Research Questions

The following research questions were investigated: 1) What percent of the 127 CCSSs GPCs are introduced at least once in the individual benchmark and progress monitoring passages across the four different R-CBM publishers? 2) What is the overall alignment between first-grade core reading curricula (Treasures, Journeys, Reading Street, Open Court, & Wonders) and standardized R-CBM (FastBridge, DIBELS 8th edition, Acadience, & AIMSwebPlus) passages on CCSSs GPCs?

Methods

Selection and acquisition of R-CBMs. The NCLII Tools Chart for Progress Monitoring (National Center on Intensive Intervention, 2018) was used to select four commonly used R-CBMs with strong evidence of reliability and validity: AIMSwebPlus, Acadience, DIBELS 8th edition, and FastBridge. The ORF passages for first-grade were acquired either through direct purchase or free download via online websites (in the case where materials are provided free of charge).

Scoring of R-CBM passages. All passages were scored using the Decoding System Measure's scoring system which is currently an online web application. Scoring involved copying and pasting the text from the document into the web browser. Output from the web application yields individual counts of GPCs per a given piece of text.

Scope of reading curricula. The methods for selecting, acquiring, and determining the scope and sequence of reading curricula are described in chapter 2 (Aim 2) of this dissertation. For the research questions specific to this chapter, only the scope of the six core basal reading programs was used: Journeys, Treasures, Wonders, Reading Street, Open Court, and Levelled Literacy Intervention. Only these six were used because they are purposely created to be finished in one year (first-grade) and have clear starting and stopping points, whereas supplemental programs have more flexibility in terms of implementation.

Analysis

In order to answer the first research question, the number (not count) of different CCSSs GPCs that occurred in a given passage was counted. This number was then divided by the total number of CCSSs (127) to yield a percentage match for each individual progress monitoring and benchmark passage across all four R-CBM publishers.

In order to answer the second research question a summary measure that represents the match between a given R-CBM passage set and reading curricula was calculated by determining the percentage of occurrence of a particular GPC across passages (see Table 17). This involved three steps. First, the number of passages that had at least one occurrence of a particular GPC was counted. This was done for all of the CCSSs GPCs that were taught in a given reading curricula (i.e. the scope of the reading curricula). Next, this was divided by the total number of possible passages (20 for progress monitoring, or 3 or 6 for benchmark passages). Finally, the number (count) of GPCs with 100% coverage (indicating at least one count per passages) was summed and divided by the total number of GPCs that were taught. Put another way, this analysis answers the question: what percent of taught GPCs show up at least once per passage?

For example, if Treasures taught the GPC a_AA, and we were examining the match of Treasures with Acadience, then the number of passages that had at least one occurrence of a_AA was counted (let's say 10 passages), then 10 was divided by 20 to yield 50% coverage for a_AA. This would be repeated for each CCSS GPC that Treasures taught. Finally, the count of the GPCs with at least one occurrence per passage would be divided by the total number of GPCs that were taught.

Results

What percent of the 127 CCSSs GPCs are introduced at least once in the individual benchmark and progress monitoring passages across the four different R-CBM publishers?

The percent match between a given passage and the CCSSs GPCs range from 32% (FastBridge progress monitoring probe # 18) to 52% (several different passages had 52% match, see Table 13). To yield an overall number that attempts to quantify the match between the CCSSs GPCs and the R-CBM publishers, the individual passage percentages were averaged. These averages ranged from 39% (FastBridge) to 48% (Acadience).

It is possible, however, that individual passages might not have a large percentage of the CCSSs GPCs represented, but that across the entire set of 20 passages there was representation. To investigate this, the 'Total' column (Table 13) was calculated. This column is also a percentage and was calculated by obtaining the sum of the GPCs that had at least 1 occurrence

Table 13

The percent match between CCSSs and R-CBM GPCs.

	BM 1	**BM 2**	**BM 3**	**BM 4**	**BM 5**	**BM 6**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**Total**	**Ave**
DIBELS 8	45	46	42				41	47	45	41	50	48	46	39	42	52	47	42	41	47	45	41	49	45	41	45	**75**	**45**
Acadience	51	48	49	49	42	52	52	50	52	49	46	45	48	52	41	51	51	50	52	47	47	47	46	40	50	48	**82**	**48**
FastBridge	39	40	39				40	38	37	38	37	38	43	38	41	36	41	41	39	38	41	40	46	32	38	37	**59**	**39**
AIMSwebPlus	45	44	46	43	46	47	43	45	45	53	41	45	45	42	41	45	44	46	45	42	44	46	44	46	48	48	**76**	**45**

Note. Total refers to the sum of the GPCs that had at least 1 occurrence across all benchmark and progress monitoring passages divided by 127 (the total number of possible CCSSs GPCs

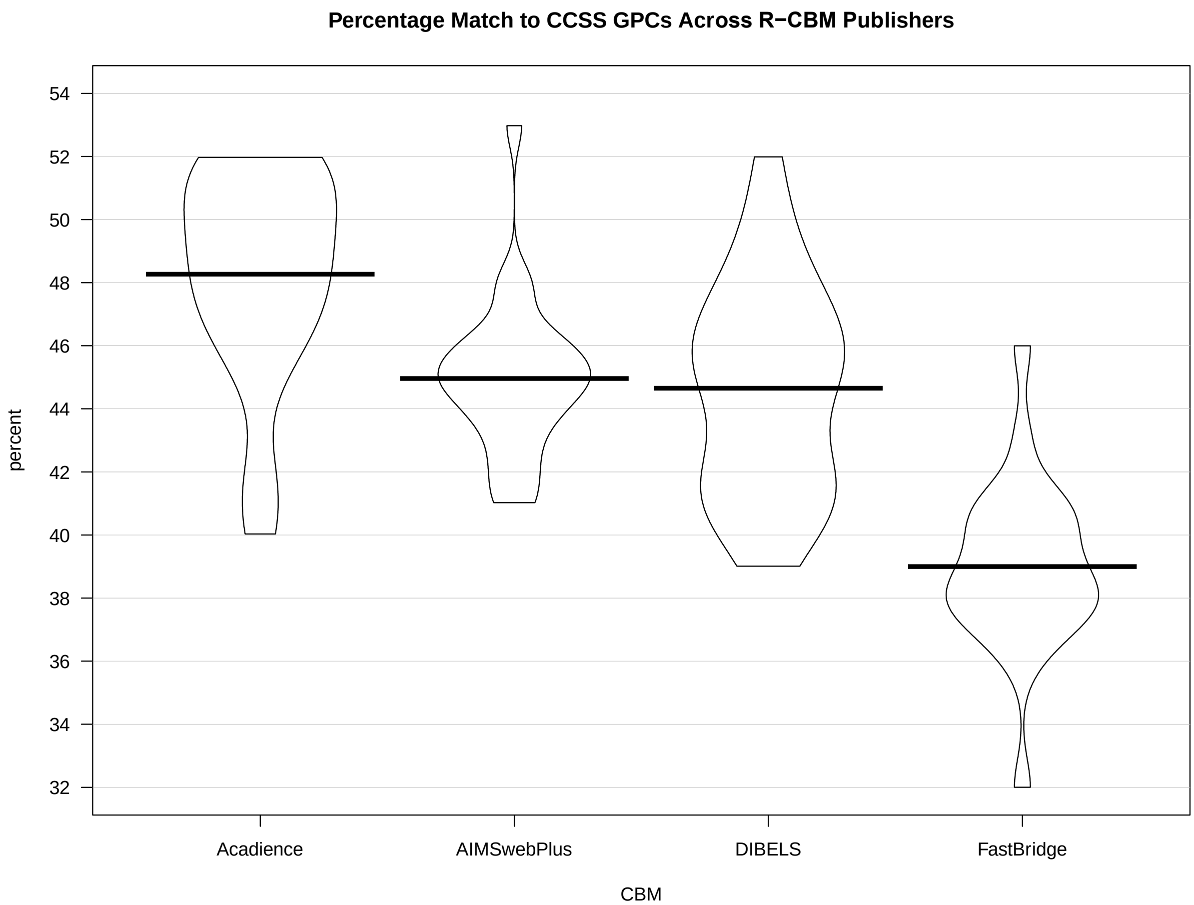


Figure 6. Boxplots of the percent match (number of the 127 CCSSs GPCs represented at least once) across publishers.

across *all* benchmark and progress monitoring passages divided by 127 (the total number of possible CCSSs GPCs). FastBridge has the least at 59% while Acadience had the most CCSSs GPCs represented at 82%. To see data on individual CCSSs GPC counts across publishers see Tables 14 to 19.

What is the overall alignment between first-grade core reading curricula (Treasures, Journeys, Reading Street, Open Court, & Wonders) and standardized R-CBM (FastBridge, DIBELS 8th edition, Acadience, & AIMSwebPlus) passages on CCSSs GPCs? The percent match (see Table 20) between R-CBMs and reading curricula for progress monitoring passages range from 34% (AIMSwebPlus: Open Court) to 64% (Acadience: Wonders). This means that when teaching GPCs using Open Court and assessing children using AIMSwebPlus, only 34% of the GPCs showed up at least once in each passage (i.e. providing an opportunity to respond and document growth). However, if teaching with Wonders and assessing with Acadience, then 64% of the taught GPCs occur at least once per passage.

For benchmark passages, the percentages range from 24% (DIBELS 8: Open Court, DIBELS 8: Reading Street, and FastBridge: Open Court), to 46% (Acadience: Levelled Literacy, and AIMSwebPlus: Levelled Literacy Intervention).

However, the percent match should not be viewed in isolation. Rather, it is important to examine the base rates of coverage in CCSSs GPCs. Examining the base rate of coverage is important because there might be high alignment between a reading curricula and a set of passages, but if the reading curricula only covered 5 (out of 127) CCSSs GPCs then no one would argue that that reading curricula is comprehensive.

Table 14

Totals, standard deviations, and percentages of GPCs across R-CBM passages and publishers for CCSSs GPC consonants.

		Acadience						FastBridge						AIMSweb						DIBELS 8th					
		Totals		**SD**		**%**		**Totals**		**SD**		**%**		**Totals**		**SD**		**%**		**Totals**		**SD**		**%**	
GPC	Example	PM	BM	PM	BM	PM	BM	PM	BM	PM	BM	PM	BM	PM	BM	PM	BM	PM	BM	PM	BM	PM	BM	PM	BM
B_B	but	268	86	5	10	2%	2%	143	26	3	1	2%	2%	365	94	10	8	3%	3%	303	22	6	2	3%	2%
C_K	cup	193	88	5	13	2%	2%	101	16	3	3	1%	1%	208	50	6	6	2%	1%	150	10	5	1	1%	1%
D_D	die	603	213	11	9	5%	6%	405	74	5	15	5%	6%	745	204	11	10	6%	5%	493	86	7	2	5%	6%
F_F	fun	233	76	5	7	2%	2%	148	22	4	5	2%	2%	185	59	5	3	2%	2%	200	32	4	2	2%	2%
G_G	girl	178	48	6	3	1%	1%	174	24	4	3	2%	2%	233	74	7	7	2%	2%	147	16	5	1	1%	1%
H_HH	house	304	69	11	8	3%	2%	283	39	7	4	4%	3%	405	159	8	14	3%	4%	217	56	8	6	2%	4%
J_JH	jump	83	16	5	3	1%	0%	41	10	3	5	1%	1%	54	39	4	4	0%	1%	17	17	1	9	0%	1%
K_K	kite	189	87	7	6	2%	2%	142	29	5	8	2%	2%	178	50	3	7	1%	1%	168	24	4	5	2%	2%
L_L	lamb	419	105	8	7	3%	3%	190	32	4	5	2%	3%	343	91	5	8	3%	2%	397	44	7	3	4%	3%
LL_L	call	100	36	3	4	1%	1%	53	9	3	3	1%	1%	118	63	4	6	1%	2%	87	18	2	4	1%	1%
M_M	mitt	410	133	10	9	3%	4%	254	40	5	9	3%	3%	528	123	9	9	4%	3%	378	54	7	3	4%	4%
N_N	nice	829	273	11	9	7%	7%	495	70	5	5	6%	6%	856	243	9	8	7%	7%	701	92	7	5	7%	7%
P_P	put	329	74	7	6	3%	2%	172	31	5	10	2%	2%	303	106	8	5	2%	3%	199	21	5	6	2%	2%
Q_K	queen	5	5	0	1	0%	0%	0	0	0	0	0%	0%	4	0	1	0	0%	0%	2	1	0	1	0%	0%
R_R	reach	791	284	9	10	7%	8%	393	76	8	5	5%	6%	633	199	9	12	5%	5%	658	69	10	7	6%	5%
S_S	sit	633	160	11	6	5%	4%	354	51	5	12	4%	4%	557	165	8	10	5%	4%	490	56	8	4	5%	4%
S_SH	sure	1	2	0	1	0%	0%	4	0	0	0	0%	0%	4	2	0	1	0%	0%	6	0	1	0	0%	0%
S_Z	as	345	91	7	7	3%	2%	240	39	4	1	3%	3%	389	102	6	5	3%	3%	287	42	8	2	3%	3%
S_ZH	measure	0	0	0	0	0%	0%	0	0	0	0	0%	0%	0	0	0	0	0%	0%	0	0	0	0	0%	0%
T_CH	future	14	1	1	0	0%	0%	0	0	0	0	0%	0%	0	0	0	0	0%	0%	0	0	0	0	0%	0%
T_T	tickle	1141	349	11	12	9%	9%	842	152	8	10	11%	12%	1049	301	11	14	9%	8%	920	116	9	2	9%	8%
V_V	van	106	44	3	4	1%	1%	73	11	2	4	1%	1%	110	39	4	5	1%	1%	108	11	3	3	1%	1%
W_W	witch	252	82	7	5	2%	2%	269	21	6	3	3%	2%	304	92	5	4	2%	2%	267	39	6	7	3%	3%
Z_Z	zoo	7	2	1	1	0%	0%	0	0	0	0	0%	0%	17	1	2	0	0%	0%	12	0	1	0	0%	0%
Z_ZH	azure	0	0	0	0	0%	0%	0	0	0	0	0%	0%	0	0	0	0	0%	0%	0	0	0	0	0%	0%

Note. PM = progress monitoring; BM = benchmark passage. 0% indicates a percentage range between 0 and 1% (the percentage was so small that it rounded to 0%)

Table 15

Totals, standard deviations, and percentages of GPCs across R-CBM passages and publishers for CCSSs GPC short vowel sounds.

		Acadience						FastBridge						AIMSweb						DIBELS 8th						
		Totals		**SD**		**%**		**Totals**		**SD**		**%**		**Totals**		**SD**		**%**		**Totals**		**SD**		**%**		
GPC	Example	PM	BM	PM	BM	PM	BM	PM	BM	PM	BM	PM	BM	PM	BM	PM	BM	PM	BM	PM	BM	PM	BM	PM	BM	Type
EA_EH	breath	34	4	3	1	0%	0%	7	3	1	1	0%	0%	7	1	1	0	0%	0%	32	2	3	1	0%	0%	short vowel
I_IH	sit	566	200	8	9	5%	5%	504	75	7	6	6%	6%	650	187	8	10	5%	5%	431	84	4	2	4%	6%	short vowel
O_AH	cover	132	50	3	3	1%	1%	70	10	2	2	1%	1%	132	30	3	3	1%	1%	112	17	3	2	1%	1%	short vowel
OO_AH	flood	0	0	0	0	0%	0%	0	0	0	0	0%	0%	0	0	0	0	0%	0%	1	0	0	0	0%	0%	short vowel
OO_UH	took	33	8	1	1	0%	0%	40	4	2	1	1%	0%	48	11	2	1	0%	0%	39	5	2	2	0%	0%	short vowel
OU_AH	tough	3	2	0	1	0%	0%	0	0	0	0	0%	0%	2	0	0	0	0%	0%	7	0	1	0	0%	0%	short vowel
OU_UH	could	28	10	1	1	0%	0%	40	6	2	2	1%	0%	12	16	1	4	0%	0%	13	12	1	4	0%	1%	short vowel
OU_UW	soup	50	3	3	1	0%	0%	0	2	0	1	0%	0%	43	14	2	3	0%	0%	37	0	3	0	0%	0%	short vowel
U_AH	cup	132	41	5	4	1%	1%	135	26	4	3	2%	2%	198	72	6	7	2%	2%	166	19	4	4	2%	1%	short vowel
U_UH	put	22	6	1	1	0%	0%	21	0	1	0	0%	0%	20	0	1	0	0%	0%	12	2	1	1	0%	0%	short vowel
O_AO	fox	90	22	2	3	1%	1%	43	5	2	1	1%	0%	98	28	4	4	1%	1%	78	3	3	2	1%	0%	short vowel

Note. PM = progress monitoring; BM = benchmark passage. 0% indicates a percentage range between 0 and 1% (the percentage was so small that it rounded to 0%)

Table 16

Totals, standard deviations, and percentages of GPCs across R-CBM passages and publishers for CCSSs GPC long vowel sounds.

		Acadience						FastBridge						AIMSweb						DIBELS 8th					
		Totals		**SD**		**%**		**Totals**		**SD**		**%**		**Totals**		**SD**		**%**		**Totals**		**SD**		**%**	
GPC	Example	PM	BM	PM	BM	PM	BM	PM	BM	PM	BM	PM	BM	PM	BM	PM	BM	PM	BM	PM	BM	PM	BM	PM	BM
A-E_EY	make	135	33	5	3	1%	1%	72	17	3	6	1%	1%	80	27	3	2	1%	1%	79	8	3	2	1%	1%
E_IY	me	156	47	5	4	1%	1%	191	24	5	5	2%	2%	256	68	8	7	2%	2%	151	28	7	8	1%	2%
EA_EY	great	5	0	1	0	0%	0%	12	2	1	1	0%	0%	1	0	0	0	0%	0%	1	0	0	0	0%	0%
EA_IY	eat	66	11	2	2	1%	0%	4	1	1	1	0%	0%	35	13	2	3	0%	0%	55	5	3	2	1%	0%
EE_IY	see	81	49	3	8	1%	1%	47	6	2	3	1%	0%	109	19	5	3	1%	1%	114	6	6	2	1%	0%
E-E_IY	these	9	4	1	1	0%	0%	19	0	3	0	0%	0%	4	9	1	3	0%	0%	4	0	0	0	0%	0%
EY_EY	they	45	19	2	2	0%	1%	45	8	4	2	1%	1%	62	30	4	6	1%	1%	88	16	5	6	1%	1%
EY_IY	key	0	3	0	1	0%	0%	0	0	0	0	0%	0%	3	0	0	0	0%	0%	1	4	0	2	0%	0%
I_AY	rifle	79	25	3	2	1%	1%	137	2	9	1	2%	0%	142	33	7	3	1%	1%	132	6	6	1	1%	0%
IE_AY	pie	5	1	1	0	0%	0%	0	0	0	0	0%	0%	2	4	0	1	0%	0%	1	1	0	1	0%	0%
I-E_AY	time	122	22	4	1	1%	1%	62	10	2	3	1%	1%	80	27	2	5	1%	1%	118	10	5	2	1%	1%
IE_IY	chief	8	0	1	0	0%	0%	0	0	0	0	0%	0%	1	2	0	1	0%	0%	0	0	0	0	0%	0%
O_OW	open	92	52	3	7	1%	1%	84	12	2	3	1%	1%	97	48	3	5	1%	1%	123	15	3	4	1%	1%
OA_OW	boat	12	26	2	7	0%	1%	0	0	0	0	0%	0%	8	2	1	1	0%	0%	7	0	1	0	0%	0%
OE_OW	toe	0	0	0	0	0%	0%	0	0	0	0	0%	0%	0	7	0	3	0%	0%	0	0	0	0	0%	0%
O-E_OW	vote	34	10	3	3	0%	0%	21	6	1	1	0%	0%	34	11	2	2	0%	0%	11	3	1	1	0%	0%
OO_UW	moo	54	12	3	1	0%	0%	42	10	2	1	1%	1%	47	4	2	1	0%	0%	59	4	2	1	1%	0%
UE_UW	blue	2	1	0	0	0%	0%	0	0	0	0	0%	0%	4	10	1	4	0%	0%	1	0	0	0	0%	0%
U-E_UW	tube	3	0	0	0	0%	0%	3	0	1	0	0%	0%	0	0	0	0	0%	0%	1	0	0	0	0%	0%

Note. PM = progress monitoring; BM = benchmark passage. 0% indicates a percentage range between 0 and 1% (the percentage was so small that it rounded to 0%)

Table 17

Totals, standard deviations, and percentages of GPCs across R-CBM passages and publishers for CCSSs GPC digraphs.

		Acadience						FastBridge						AIMSweb						DIBELS 8th					
		Totals		**SD**		**%**		**Totals**		**SD**		**%**		**Totals**		**SD**		**%**		**Totals**		**SD**		**%**	
GPC	Example	PM	BM	PM	BM	PM	BM	PM	BM	PM	BM	PM	BM	PM	BM	PM	BM	PM	BM	PM	BM	PM	BM	PM	BM
GN_N	gnat	1	1	0	0	0%	0%	0	0	0	0	0%	0%	0	0	0	0	0%	0%	0	0	0	0	0%	0%
KN_N	knight	7	1	1	0	0%	0%	1	0	0	0	0%	0%	17	3	1	1	0%	0%	12	2	1	1	0%	0%
MB_M	comb	1	3	0	1	0%	0%	0	0	0	0	0%	0%	3	1	0	0	0%	0%	2	1	0	1	0%	0%
MN_M	hymn	0	0	0	0	0%	0%	0	0	0	0	0%	0%	0	0	0	0	0%	0%	0	0	0	0	0%	0%
TT_T	mitt	23	3	2	1	0%	0%	8	6	1	2	0%	0%	14	3	1	1	0%	0%	17	1	1	1	0%	0%
ZZ_Z	jazz	7	0	2	0	0%	0%	0	0	0	0	0%	0%	2	0	0	0	0%	0%	0	0	0	0	0%	0%
CH_CH	cheap	69	11	2	1	1%	0%	24	1	3	1	0%	0%	22	10	1	1	0%	0%	24	2	1	1	0%	0%
CH_K	chorus	20	0	3	0	0%	0%	9	0	1	0	0%	0%	1	1	0	0	0%	0%	4	1	1	1	0%	0%
CH_SH	charade	2	0	0	0	0%	0%	0	0	0	0	0%	0%	0	0	0	0	0%	0%	0	0	0	0	0%	0%
CK_K	duck	38	20	2	5	0%	1%	56	13	3	5	1%	1%	41	20	1	3	0%	1%	25	11	2	4	0%	1%
CI_SH	special	0	0	0	0	0%	0%	0	0	0	0	0%	0%	0	0	0	0	0%	0%	0	0	0	0	0%	0%
LF_F	calf	1	0	0	0	0%	0%	0	0	0	0	0%	0%	0	0	0	0	0%	0%	0	0	0	0	0%	0%
LK_K	folk	15	1	2	0	0%	0%	0	0	0	0	0%	0%	20	0	1	0	0%	0%	10	2	1	1	0%	0%
NG_NG	sing	100	24	3	3	1%	1%	47	3	2	1	1%	0%	58	20	2	2	0%	1%	96	8	3	3	1%	1%
PH_F	sphere	1	1	0	0	0%	0%	0	0	0	0	0%	0%	1	0	0	0	0%	0%	0	0	0	0	0%	0%
PS_S	psychic	0	0	0	0	0%	0%	0	0	0	0	0%	0%	0	0	0	0	0%	0%	0	0	0	0	0%	0%
SC_S	science	0	0	0	0	0%	0%	0	0	0	0	0%	0%	0	0	0	0	0%	0%	0	0	0	0	0%	0%
SC_SH	precious	0	0	0	0	0%	0%	0	0	0	0	0%	0%	0	0	0	0	0%	0%	0	0	0	0	0%	0%
SH_SH	shoe	118	32	5	4	1%	1%	78	15	4	3	1%	1%	98	42	5	6	1%	1%	77	9	5	4	1%	1%
SI_SH	mission	0	0	0	0	0%	0%	0	0	0	0	0%	0%	0	0	0	0	0%	0%	0	0	0	0	0%	0%
SS_S	pass	20	5	1	1	0%	0%	30	0	4	0	0%	0%	27	1	3	0	0%	0%	22	0	1	0	0%	0%
SS_SH	mission	0	0	0	0	0%	0%	0	0	0	0	0%	0%	0	0	0	0	0%	0%	0	0	0	0	0%	0%
TCH_CH	etch	17	1	2	0	0%	0%	4	2	1	1	0%	0%	2	6	0	2	0%	0%	7	2	1	1	0%	0%
TH_DH	this, breathe, either	557	198	6	11	5%	5%	304	49	6	10	4%	4%	536	182	9	8	4%	5%	450	75	8	11	4%	5%
TH_TH	thin, breath, ether	54	15	2	2	0%	0%	42	4	2	1	1%	0%	29	16	1	1	0%	0%	60	5	5	1	1%	0%
TI_SH	notion	4	0	1	0	0%	0%	0	0	0	0	0%	0%	0	0	0	0	0%	0%	1	0	0	0	0%	0%
WH_HH	whole	1	1	0	0	0%	0%	0	0	0	0	0%	0%	3	1	0	0	0%	0%	1	1	0	1	0%	0%
WH_W	where	55	11	1	1	0%	0%	16	1	1	1	0%	0%	54	11	2	1	0%	0%	41	4	1	1	0%	0%
WR_R	wrap	4	0	0	0	0%	0%	1	0	0	0	0%	0%	1	0	0	0	0%	0%	0	0	0	0	0%	0%

Note. PM = progress monitoring; BM = benchmark passage. 0% indicates a percentage range between 0 and 1% (the percentage was so small that it rounded to 0%)

Table 18

Totals, standard deviations, and percentages of GPCs across R-CBM passages and publishers for CCSSs GPC dipthongs.

		Acadience						FastBridge						AIMSweb						DIBELS 8th					
		Totals		**SD**		**%**		**Totals**		**SD**		**%**		**Totals**		**SD**		**%**		**Totals**		**SD**		**%**	
GPC	Example	PM	BM	PM	BM	PM	BM	PM	BM	PM	BM	PM	BM	PM	BM	PM	BM	PM	BM	PM	BM	PM	BM	PM	BM
AI_EY	rain	29	6	2	2	0%	0%	0	0	0	0	0%	0%	13	2	1	1	0%	0%	5	2	0	1	0%	0%
AU_AA	pause	0	0	0	0	0%	0%	0	0	0	0	0%	0%	0	0	0	0	0%	0%	0	0	0	0	0%	0%
AW_AA	saw	0	0	0	0	0%	0%	0	0	0	0	0%	0%	0	0	0	0	0%	0%	0	0	0	0	0%	0%
AY_EY	play	66	15	3	2	1%	0%	40	12	1	3	1%	1%	73	25	2	4	1%	1%	66	12	4	3	1%	1%
EI_EY	vein	0	0	0	0	0%	0%	0	0	0	0	0%	0%	0	0	0	0	0%	0%	0	0	0	0	0%	0%
EI_IY	either	0	0	0	0	0%	0%	0	0	0	0	0%	0%	0	0	0	0	0%	0%	0	0	0	0	0%	0%
OI_OY	boil	3	0	0	0	0%	0%	0	0	0	0	0%	0%	1	0	0	0	0%	0%	5	0	1	0	0%	0%
OU_AW	out	65	16	3	1	1%	0%	41	11	2	4	1%	1%	43	12	2	1	0%	0%	53	8	3	3	1%	1%
OW_AW	cow	30	9	2	2	0%	0%	8	1	1	1	0%	0%	26	5	2	1	0%	0%	45	5	4	1	0%	0%
OW_OW	snow	48	8	3	2	0%	0%	7	0	1	0	0%	0%	51	9	3	1	0%	0%	35	5	2	2	0%	0%
OY_OY	boy	4	3	0	1	0%	0%	0	0	0	0	0%	0%	5	4	1	2	0%	0%	7	0	1	0	0%	0%
U_Y.UW	uniform	5	0	0	0	0%	0%	0	0	0	0	0%	0%	2	0	0	0	0%	0%	0	0	0	0	0%	0%
U-E_Y.UW	cute	16	3	1	1	0%	0%	5	0	1	0	0%	0%	6	0	1	0	0%	0%	16	1	1	1	0%	0%

Note. PM = progress monitoring; BM = benchmark passage. 0% indicates a percentage range between 0 and 1% (the percentage was so small that it rounded to 0%)

Table 19

Totals, standard deviations, and percentages of GPCs across R-CBM passages and publishers for CCSSs GPC in the ‘other’ category.

		Acadience						FastBridge						AIMSweb						DIBELS 8th					
		Totals		**SD**		**%**		**Totals**		**SD**		**%**		**Totals**		**SD**		**%**		**Totals**		**SD**		**%**	
GPC	Example	PM	BM	PM	BM	PM	BM	PM	BM	PM	BM	PM	BM	PM	BM	PM	BM	PM	BM	PM	BM	PM	BM	PM	BM
EIGH_EY	eight	2	1	0	0	0%	0%	0	0	0	0	0%	0%	1	0	0	0	0%	0%	0	0	0	0	0%	0%
ER_ER	her	10	6	1	1	0%	0%	11	0	1	0	0%	0%	13	0	1	0	0%	0%	18	3	2	1	0%	0%
EU_Y.UW	feud	0	0	0	0	0%	0%	0	0	0	0	0%	0%	0	0	0	0	0%	0%	0	0	0	0	0%	0%
EW_UW	chew	20	6	2	1	0%	0%	5	0	1	0	0%	0%	17	7	1	2	0%	0%	6	2	1	1	0%	0%
EW_Y.UW	few	2	0	0	0	0%	0%	0	0	0	0	0%	0%	0	0	0	0	0%	0%	2	0	0	0	0%	0%
GE_JH	wage	2	1	0	0	0%	0%	0	0	0	0	0%	0%	3	0	0	0	0%	0%	1	0	0	0	0%	0%
IGH_AY	right	25	3	1	1	0%	0%	7	3	1	2	0%	0%	9	3	1	1	0%	0%	24	6	1	3	0%	0%
IR_ER	sir	6	1	1	0	0%	0%	0	0	0	0	0%	0%	5	0	1	0	0%	0%	6	0	1	0	0%	0%
LE_AH.L	single	51	6	3	1	0%	0%	7	7	1	2	0%	1%	12	1	1	0	0%	0%	19	1	1	1	0%	0%
N_NG	bank	16	11	1	2	0%	0%	31	7	2	4	0%	1%	19	6	1	1	0%	0%	18	0	2	0	0%	0%
OUGH_AO	bought	5	1	1	0	0%	0%	8	1	1	1	0%	0%	6	0	1	0	0%	0%	2	2	0	1	0%	0%
UR_ER	fur	4	3	0	1	0%	0%	1	0	0	0	0%	0%	4	0	0	0	0%	0%	6	0	1	0	0%	0%
VE_V	dove	0	0	0	0	0%	0%	0	0	0	0	0%	0%	0	0	0	0	0%	0%	0	0	0	0	0%	0%
Y_AY	cry	28	2	3	1	0%	0%	46	0	3	0	1%	0%	43	4	3	1	0%	0%	74	8	4	3	1%	1%
Y_IH	gym	1	0	0	0	0%	0%	0	0	0	0	0%	0%	0	0	0	0	0%	0%	0	0	0	0	0%	0%
Y_IY	happy	101	42	3	6	1%	1%	44	5	2	2	1%	0%	72	46	2	7	1%	1%	85	18	3	2	1%	1%
Y_Y	you	80	7	6	1	1%	0%	4	4	1	1	0%	0%	62	25	3	3	1%	1%	76	4	6	2	1%	0%
I_Y	onion	0	0	0	0	0%	0%	0	0	0	0	0%	0%	0	0	0	0	0%	0%	1	0	0	0	0%	0%
UI_UW	suit	2	0	0	0	0%	0%	0	0	0	0	0%	0%	0	0	0	0	0%	0%	0	0	0	0	0%	0%

Note. PM = progress monitoring; BM = benchmark passage. 0% indicates a percentage range between 0 and 1% (the percentage was so small that it rounded to 0%)

Table 20

Percent of Curricula-covered GPCs that had at least 1 occurrence in each passage (split by benchmark and progress monitoring passages).

R-CBM (and match with CCSSs GPCs in parentheses, regardless of match with R-CBM- taken from Table 11)	Reading Curricula (and match with CCSSs GPCs in parentheses, regardless of match with R-CBM- taken from Table 11)	Benchmark Passages	Progress Monitoring
Acadience (48%)	Open Court (70%)	39%	28%
	Reading Street (69%)	38%	28%
	Journeys (61%)	42%	32%
	Wonders (59%)	64%	42%
	Treasures (52%)	48%	34%
	Levelled Literacy (46%)	58%	46%
AIMSwebPlus (45%)	Open Court (70%)	34%	31%
	Reading Street (69%)	37%	30%
	Journeys (61%)	42%	32%
	Wonders (59%)	55%	47%
	Treasures (52%)	45%	37%
	Levelled Literacy (46%)	54%	46%
DIBELS 8th (45%)	Open Court (70%)	38%	24%
	Reading Street (69%)	38%	24%
	Journeys (61%)	39%	29%
	Wonders (59%)	63%	38%
	Treasures (52%)	46%	31%
	Levelled Literacy (46%)	56%	39%
FastBridge (39%)	Open Court (70%)	36%	24%
	Reading Street (69%)	38%	25%
	Journeys (61%)	40%	26%
	Wonders (59%)	59%	34%
	Treasures (52%)	48%	28%
	Levelled Literacy (46%)	56%	37%

Therefore, in Table 20, the percentage next to the R-CBM in the left-most column represents the average percent of CCSSs GPCs covered across the passage sets and is taken from Table 13. The percentage in the second column next to the reading curricula represents the percentage of CCSSs GPCs covered in the given reading curricula and was taken from Table 11 in chapter 2. That said, the overall percentages do not indicate whether there was a match on the *specific* GPCs being taught and assessed. Therefore, these percentages, which represent the match between the two are in the third (for progress monitoring) and fourth (for benchmark) columns.

Discussion

Summary. This chapter presents analyses that examine the match between R-CBMs and reading curricula on the CCSSs GPCs. Two research questions were answered. First, the results showed that Acadience oral reading fluency passages covered the most CCSSs GPCs, while FastBridge covered the least. Second, the alignment between R-CBMs and basal reading curricula was investigated and the results show that there were high matches between certain reading curricula and certain R-CBMs, but that these results need to be interpreted with caution and cannot be viewed in isolation. This is discussed further in the next section.

Practical implications. The results show that while there was variation in alignment between both R-CBMs and CCSSs GPCs (research question 1) and between alignment on R-CBMs and reading curricula (research question 2), there was not a clear-cut 'winning combination' of R-CBM and reading curricula. Instead, the answer is the question of which are most aligned is: it depends. Specifically, it depends on the needs of the purchasing district and end users. For example, if a certain school district values coverage (larger scope) of GPCs over alignment (perhaps they do not use R-CBMs for informing instruction) then they would want to look for reading curricula with high percentages in the second column of Table 20. If, however, a school district already has a reading curricula and is looking to purchase a set of R-CBMs, then then they would want to examine the third and fourth columns of Table 20 to see which R-CBM yields the greatest match.

Limitations. These analyses are the first to explore the alignment between oral reading fluency R-CBMs and first grade basal reading curricula, and as such, contain several limitations. First, alignment can be operationalized in several different ways. Here, it was defined as the number of CCSSs GPCs that were taught in a given reading curricula and appeared at least once across the set of passages. Choosing one occurrence (i.e. one opportunity per passage) for a child to demonstrate their knowledge was arbitrary, but intuitive in the sense that most people would agree that zero opportunities to respond is not enough. That said, some research has shown that at least three opportunities to respond (Kern & Hosp, 2018) are necessary to minimize the chance that the student was guessing. In terms of GPCs, this would mean that each GPC would need to be presented at least 3 times per passage, across all passages, if R-CBMs are going to be used for informing instruction. To address this limitation, the analysis was repeated using 3 GPCs as a minimum (see Table 21). The percent matches are lower across the board (18% to 37% for benchmark passages, and 13% to 26% for progress monitoring passages), suggesting that there are several GPCs that are not represented at least 3 times across the set of passages. This suggests that first-grade oral reading fluency passages, in their current state, are not designed with sufficient opportunities to respond per passage to minimize chance accuracy.

Table 21

Percent of Curricula-covered GPCs that had at least 3 occurrences in each passage (split by benchmark and progress monitoring passages).

	Reading Curricula	Benchmark Passages	Progress Monitoring
Acadience (48%)	Open Court (70%)	24%	19%
	Reading Street (69%)	25%	21%
	Journeys (61%)	26%	21%
	Wonders (59%)	20%	16%
	Treasures (52%)	30%	24%
	Levelled Literacy (46%)	37%	30%
AIMSwebPlus (45%)	Open Court (70%)	24%	20%
	Reading Street (69%)	25%	22%
	Journeys (61%)	25%	23%
	Wonders (59%)	21%	17%
	Treasures (52%)	31%	25%
	Levelled Literacy (46%)	37%	33%
DIBELS 8th (45%)	Open Court (70%)	23%	16%
	Reading Street (69%)	24%	17%
	Journeys (61%)	25%	18%
	Wonders (59%)	18%	17%
	Treasures (52%)	27%	21%
	Levelled Literacy (46%)	35%	26%
FastBridge (39%)	Open Court (70%)	20%	15%
	Reading Street (69%)	22%	16%
	Journeys (61%)	22%	16%
	Wonders (59%)	20%	13%
	Treasures (52%)	27%	19%
	Levelled Literacy (46%)	32%	23%

With regard to the ideal number of opportunities to respond per passage, there is another important question that should be mentioned: should all GPCs be represented equally? For example, b_B (as in 'book') is more common than bt_T as in 'debt', yet should each have at least three occurrences in each passage (assuming both are taught in the reading curricula)?

Even if one was able to determine an 'ideal' number of opportunities for each GPC per passage, a next question would be: is it even possible for an author to combine words with the requisite GPCs to create 20 (26 including benchmark passages) passages that are cohesive and interesting enough for a first-grader to read? This is an important practical question that would need to be investigated.

Finally, there is an underlying assumption throughout these analyses that more CCSSs GPC coverage is better, when in fact, there is little research to support this notion. For example, is covering all the CCSSs GPCs in first grade better than covering some in first grade and some in second grade and being able to devote more instructional time with each? This is an important, yet unanswered question that future research could address.

Conclusion

Despite all the gains in reading instruction research in the last three decades, there has been little improvement in children's reading scores on national testing (U.S. Department of Education, 2017). The motivating force behind the research questions presented in this dissertation was to examine if, one day, R-CBMs could be used to inform the specific GPCs on which a given child needs extra instruction. With this overarching goal in mind, several questions had to be answered first: how are passages currently constructed? (Chapter 1), How do reading curricula currently cover GPCs? (Chapter 2) Are there equal opportunities to respond (CCSS GPCs) across individual passages, and what are the differences across publishers?

(Chapter 3) and; How do reading curricula currently align with R-CBMs? (Chapter 3). The answers to these questions, presented throughout these three chapters provide insight into where changes can be made to improve R-CBMs so that they can continue to be a revolutionary form of assessment while providing specific, psychometrically sound data to help individualize instruction.

REFERENCES

Adams, M. J. (1990). *Beginning to read: Thinking and learning about print.* Cambridge, MA: MIT Press.

Albano, A. D., & Rodriguez, M. C. (2011). Statistical equating with measures of oral reading fluency. *Journal of School Psychology*, *50*, 43-59.

Anderson, R. C., Hiebert, E. H., Scott, J. A., & Wilkinson, I. A. G. (1985). *Becoming a nation of readers: The report of the Commission on Reading.* Washington, DC: National Academy of Education, Commission on Education and Public Policy.

Andrews, D.F., Bickel, P.J., Hampel, F.R., Huber, P.J., Rogers, W.H., & Tukey, J.W. (1972) Robust Estimates of Location: Survey and Advances. Princeton, New Jersey: Princeton University Press.

Ardoin, S. P., & Christ, T. J. (2009). Curriculum-based measurement of oral reading: Standard errors associated with progress monitoring outcomes from DIBELS, AIMSweb, and an experimental passage set. *School Psychology Review*, *38*, 266-283.

Ardoin, S. P., Christ, T. J., Morena, L. S., Cormier, D. C., & Klingbeil, D. A. (2013). A systematic review and summarization of the recommendations and research surrounding curriculum-based measurement of oral reading fluency (CBM-R) decision rules. Journal of school psychology, *51*, 1-18.

Ardoin, S. P., Suldo, S. M., Witt, J., Aldrich, S., & McDonald, E. (2005). Accuracy of Readability Estimates' Predictions of CBM Performance. *School Psychology Quarterly*, *20*, 1-22.

Berkeley, S., Bender, W. N., Gregg Peaster, L., & Saunders, L. (2009). Implementation of response to intervention: A snapshot of progress. *Journal of Learning Disabilities*, *42*, 85-95.

Bieber, G., & Choi, H. (2004). The use of curriculum based measurement by school psychologists. *Journal of Psychological Practice*, *10*, 25–36.

Bailin, A., & Grafstein, A. (2001). The linguistic assumptions underlying readability formulae: A critique. *Language & Communication*, *21*, 285-301.

Baker, S. K., Fien, H., & Baker, D. L. (2010). Robust reading instruction in the early grades: Conceptual and practical issues in the integration and evaluation of Tier 1 and Tier 2 instructional supports. *Focus on Exceptional Children*, *42*, 1-20.

Barth, A. E., Tolar, T. D., Fletcher, J. M., & Francis, D. (2014). The effects of student and text characteristics on the oral reading fluency of middle-grade students. *Journal of Educational Psychology*, *106*, 162-180.

Bruce, B., & Rubin, A. (1988). Readability formulas: Matching tool and task. In A. Davison & G. M. Green (Eds.), *Linguistic complexity and text comprehension: Readability issues reconsidered* (pp. 5-22). Hillsdale, NJ, US: Lawrence Erlbaum Associates, Inc.

Christ, T.J., and Colleagues (2015). Formative Assessment System for Teachers: Technical Manual Version 2.0, Minneapolis, MN: Author and FastBridge Learning (www.fastbridge.org)

Common Core State Standards Initiative. (2010). Common Core State Standards for English language arts & literacy in history/social studies, science, and technical subjects. Washington, DC: CCSSO & National Governors Association.

Compton, D. L., Appleton, A. C., & Hosp, M. K. (2004). Exploring the relationship between text-leveling systems and reading accuracy and fluency in second-grade students who are average and poor decoders. *Learning Disabilities Research & Practice*, *19*, 176-184.

Cummings, K. D., Park, Y., & Bauer Schaper, H. A. (2013). Form effects on DIBELS Next oral reading fluency progress-monitoring passages. *Assessment for Effective Intervention*, *38*, 91-104.

Chall, J. S. (1996). Learning to read: The great debate. New York: McGraw-Hill. (Original work published 1967)

Common Core State Standards Initiative (2010). Common Core State Standards for English language arts & literacy in history/social studies, science, and technical subjects. Washington, DC: CCSSO & National Governors Association.

Cunningham, J. W., Spadorcia, S. A., Erickson, K. A., Koppenhaver, D. A., Sturm, J. M., & Yoder, D. E. (2005). Investigating the instructional supportiveness of leveled texts. *Reading Research Quarterly, 40,* 410-427.

Danielson, L., & Rosenquist, C. (2014). Introduction to the TEC special issue on data-based individualization. *Teaching Exceptional Children*, *46*, 6-12.

Davison, A., & Kantor, R. N. (1982). On the failure of readability formulas to define readable texts: A case study from adaptations. *Reading Research Quarterly, 17,* 187-209.

Deno, S. L. (2003). Developments in curriculum-based measurement. *The Journal of Special Education*, *37*, 184-192.

Dunn, E. K., & Eckert, T. L. (2002). Curriculum-based measurement in reading: A comparison of similar versus challenging material. *School Psychology Quarterly*, *17*, 24-46.

Ehri, L. C. (1995). Phases of development in learning to read words by sight. *Journal of research in reading*, *18*, 116-125.

Ehri, L. C., Nunes, S. R., Stahl, S. A., & Willows, D. M. (2001). Systematic phonics instruction helps students learn to read: Evidence from the National Reading Panel's meta-analysis. *Review of Educational Research*, *71*, 393-447.

Flynn, L. J., Hosp, J. L., Hosp, M. K., & Robbins, K. P. (2011). Word Recognition Error Analysis: Comparing Isolated Word List and Oral Passage Reading. *Assessment for Effective Intervention*, *36*, 167-178.

Foorman, B. R., Smith, K. G., & Kosanovich, M. L. (2017). Rubric for evaluating reading/language arts instructional materials for kindergarten to grade 5 (REL 2017-219). Washington, DC: US Department of Education, Institute of Education Sciences. National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Southeast. Retrieved from https://ies. ed. gov/ncee/edlabs.

Francis, D. J., Santi, K. L., Barr, C., Fletcher, J. M., Varisco, A., & Foorman, B. R. (2008). Form effects on the estimation of students' oral reading fluency using DIBELS. *Journal of School Psychology*, *46*, 315-342.

Fuchs, L. S. (2017). Curriculum-Based Measurement as the Emerging Alternative: Three Decades Later. *Learning Disabilities Research & Practice*, *32*, 5-7.

Fuchs, L. S., & Deno, S. L. (1994). Must instructionally useful performance assessment be based in the curriculum? *Exceptional Children*, *61*, 15-24.

Fuchs, L. S., Deno, S. L., & Mirkin, P. K. (1984). The effects of frequent curriculum-based measurement and evaluation on pedagogy, student achievement, and student awareness of learning. *American Educational Research Journal*, *21*, 449-460.

Fuchs, L. S., & Fuchs, D. (2001). What Is Scientifically-Based Research on Progress Monitoring? *National Center on Student Progress Monitoring*, 1-9.

Fuchs, D., Fuchs, L. S., & Compton, D. L. (2012). Smart RTI: A next-generation approach to multilevel prevention. *Exceptional Children*, *78*, 263-279.

Fuchs, L. S., Fuchs, D., & Hamlett, C. L. (2007). Using curriculum-based measurement to inform reading instruction. *Reading and Writing*, *20*, 553-567.

Fuchs, L. S., Fuchs, D., Hosp, M. K., & Hamlett, C. L. (2003). The potential for diagnostic analysis within curriculum-based measurement. *Assessment for Effective Intervention*, *28*, 13-22.

Fuchs, L. S., Fuchs, D., Hosp, M. K., & Jenkins, J. R. (2001). Oral reading fluency as an indicator of reading competence: A theoretical, empirical, and historical analysis. *Scientific Studies of Reading*, *5*, 239-256.

Gamer, M. (2012). Package "irr": Various Coefficients of Interrater Reliability and Agreement. https://www.r-project.org. Accessed August 2019.

Good, R., Kaminski, R., Dewey, E., Wallin, J., Powell-Smith, K., and Latimer, R., (2013). DIBELS Next Technical Manual.

Good, R. H., Kaminski, R. A., Smith, S., Simmons, D., Kame'enui, E., & Wallin, J. (2003). Reviewing outcomes: Using DIBELS to evaluate a school's core curriculum and system of additional intervention in kindergarten. *Reading in the classroom: Systems for observing teaching and learning. Baltimore: Paul H. Brookes*.

Graney, S. B., & Shinn, M. R. (2005). Effects of reading curriculum-based measurement (R-CBM) teacher feedback in general education classrooms. *School Psychology Review*, *34*, 184-201.

Griffiths, A. J., VanDerHeyden, A. M., Skokut, M., & Lilles, E. (2009). Progress monitoring in oral reading fluency within the context of RTI. *School Psychology Quarterly*, *24*, 13-23.

Hasbrouck, J., & Tindal, G. A. (2006). Oral reading fluency norms: A valuable assessment tool for reading teachers. *The Reading Teacher*, *59*, 636-644.

Hiebert, E. H., & Mesmer, H. A. E. (2013). Upping the ante of text complexity in the common core state standards: examining its potential impact on young readers. *Educational Researcher*, *42*, 44–51.

Hiebert, E. H., Pearson, P. D. (2010). *An examination of current text difficulty indices with early reading texts* (Reading Research Report No. 10-01). Santa Cruz, CA: TextProject, Inc.

Hintze, J. M., & Christ, T. J. (2004). An examination of variability as a function of passage variance in CBM progress monitoring. *School Psychology Review*, *33*, 204-217.

Hintze, J. M., Shapiro, E. S., & Lutz, J. G. (1994). The effects of curriculum on the sensitivity of curriculum-based measurement in reading. *The Journal of Special Education*, *28*, 188-202.

Kern, M. L., & Friedman, H. S. (2009). Early educational milestones as predictors of lifelong academic achievement, midlife adjustment, and longevity. *Journal of Applied Developmental Psychology*, *30*, 419-430.

Kern, A. & Hosp, M. (2018) The status of decoding tests in instructional decision-making. *Assessment for Effective Intervention*, *44*, 32-44.

Kolen, M. J., & Brennan, R. L. (2004). Test equating, scaling, and linking.

McGlinchey, M. T., & Hixson, M. D. (2004). Using curriculum-based measurement to predict performance on state assessments in reading. *School Psychology Review*, *33*, 193-203.

Mesmer, H. A., Cunningham, J. W., & Hiebert, E. H. (2012). Toward a theoretical model of text complexity for the early grades: learning from the past, anticipating the future. *Reading Research Quarterly*, *47*, 235–258.

Mesmer, H. A. E., & Griffith, P. L. (2005). Everybody's selling it—But just what is explicit, systematic phonics instruction? *The Reading Teacher*, *59*, 366-376.

Miura Wayman, M., Wallace, T., Wiley, H. I., Tichá, R., & Espin, C. A. (2007). Literature synthesis on curriculum-based measurement in reading. *The Journal of Special Education*, *41*, 85-120.

National Center on Response to Intervention (January 2013). Progress Monitoring Brief #1 Common Progress Monitoring Omissions: Planning and Practice. Washington, DC: U.S. Department of Education, Office of Special Education Programs, National Center on Response to Intervention.

National Center on Intensive Intervention. (2018). Academic Progress Monitoring Tools Chart for Reading. https://charts.intensiveintervention.org/chart/progress-monitoring

National Reading Panel (US). (2000). *National Reading Panel: Teaching Children to Read: an Evidence-based Assessment of the Scientific Research Literature on Reading and Its Implications for Reading Instruction*. National Institute of Child Health and Human Development, National Institutes of Health.

NCS Pearson, Inc. (2012). *AIMSweb Technical Manual.* Bloomington, MN: NCS Pearson, Inc.

O'Keeffe, B. V., Bundock, K., Kladis, K. L., Yan, R., & Nelson, K. (2017). Variability in DIBELS Next progress monitoring measures for students at risk for reading difficulties. *Remedial and Special Education, 38*, 272-283.

Petscher, Y., & Kim, Y. S. (2011). The utility and accuracy of oral reading fluency score types in predicting reading comprehension. *Journal of School Psychology*, *49*, 107-129.

Phillips, Nathanial. (2017). *yarrr: The Pirate's guide to R (Version 0.1.5)*.

Poncy, B. C., Skinner, C. H., & Axtell, P. K. (2005). An investigation of the reliability and standard error of measurement of words read correctly per minute using curriculum-based measurement. *Journal of Psychoeducational Assessment*, *23*, 326-338.

Powell-Smith, K. A., Good, R. H., & Atkins, T. (2010). DIBELS Next Oral Reading Fluency Readability Study (Technical Report No. 7). Eugene, OR: Dynamic Measurement Group.

R Core Team (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

Saha, N., DelTufo, S., Bailey, S., & Cutting, L.E. (in preparation). Initial validation evidence for the Decoding System Measure (DSyM): A measure of decoding difficulty.

Santi, K. L., Barr, C., Khalaf, S., & Francis, D. J. (2016). Different approaches to equating oral reading fluency passages. In *The fluency construct* (pp. 223-265). Springer, New York, NY.

Scholz, F. W., & Stephens, M. A. (1987). K-sample Anderson–Darling tests. *Journal of the American Statistical Association*, *82*, 918-924.

Shapiro, E. S. (2011). *Academic skills problems: Direct assessment and intervention*. Guilford Press.

Shapiro, E. S. (2013). Commentary on progress monitoring with CBM-R and decision making: problems found and looking for solutions. *Journal of School Psychology, 51*, 59-66.

Snow, C. E., Burns, S., & Griffin, P. (Eds.). (1998). *Preventing reading difficulties in young children.* Washington, DC: National Academy Press.

Stoolmiller, M., Biancarosa, G., & Fien, H. (2013). Measurement properties of DIBELS oral reading fluency in grade 2: Implications for equating studies. *Assessment for Effective Intervention*, *38*, 76-90.

Tindal, G. (2013). Curriculum-based measurement: A brief history of nearly everything from the 1970s to the present. ISRN Education, 2013.

Toyama, Y., Hiebert, E. H., & Pearson, P. D. (2017). An analysis of the text complexity of leveled passages in four popular classroom reading assessments. *Educational Assessment*, *22*, 139-170.

U.S. Department of Education, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2017 Reading Assessment.

Wonder-McDowell, C., Reutzel, D. R., & Smith, J. A. (2011). Does instructional alignment matter? Effects on struggling second graders' reading achievement. *The Elementary School Journal*, *112*, 259-279.