Title
Rethinking the Role of Decodable Texts in Early Literacy Instruction

Author
Frey, Rick Chan

Publication Date
2012

Peer reviewed|Thesis/dissertation

Rethinking the Role of Decodable Texts in Early Literacy Instruction

By

Rick Chan Frey

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Education

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor David Pearson, Chair

Professor Anne Cunningham

Professor Allison Gopnik

Fall 2012

Abstract

Rethinking the Role of Decodable Texts

in Early Literacy Instruction

by

Rick Chan Frey

Doctor of Philosophy in Education

University of California, Berkeley

Professor David Pearson, Chair

Decodable books based on previous classroom instruction are the most frequently used texts for 1st grade reading instruction in public schools, yet no empirical studies exist demonstrating their efficacy or their benefits for beginning readers. This study attempts to address this gap in the research literature by analyzing the reading behaviors of a group of 1st grade students reading the decodable texts included as part of the 1st grade reading curriculum in a large public, urban school district.

Students read the 24 texts sampled in this study over the course of five months and detailed analysis of their reading behavior demonstrated the differential effects of these texts on high-performing, mid-performing and low-performing readers. High-performing readers began the year reading beyond the level of the decodable texts and interacted with the decodable texts minimally. Mid-performing students clearly benefited from reading the decodable texts even though their accuracy and fluency scores indicated substantial difficulties with the texts. For struggling beginning readers, however, the decodable texts were too difficult and the consistently low accuracy and low fluency scores were coupled with a range of problematic reading behaviors that demonstrated the disproportionate difficulty struggling readers had with the decodable texts.

Table of Contents

Acknowledgements

This has been a long process and one I am very grateful to have finally completed. I would like to thank David Pearson for his exceptionally helpful discussions about my ideas, his suggestions regarding the design and analysis of this study and his careful reading of my dissertation as well as for his leadership of our reading group and to those in the group who contributed to this research over the years. Special thanks to Anne Cunningham, coming into the process half-way finished and her willingness to navigate those challenges and to Allison Gopnik for her willingness to cross over from Psychology to Education and be a part of this team.

Thanks to all of the staff in the education department at Cal, especially Ilka Williams, who was constantly available for questions and for her proactive communication that helped keep me on course. And final thanks to my wife Mae, daughter Ruby and many friends who listened to me talk about decodable books way more than they ever wanted to.

Introduction

This dissertation project focuses on the texts used to support beginning readers. While questions surrounding the role of phonics in classroom instruction have largely been settled, at least in terms of instructional emphasis, there is little research that has attempted to isolate the effect that specific types of texts have on supporting early reading development (Allington, 2005; Adams, 2009; Hiebert, Mesmer & Cunningham, 2010). In the most commonly cited study of the effects of text on early reading acquisition, Juel & Roper-Schnieder (1985) found that student reading development was significantly affected by the types of texts children used to learn to read—specifically that students who learned to read using texts with a high percentage of easy to decode words developed a stronger habit of decoding than did students who learned to read using texts with a high percentage of irregularly spelled high frequency words (the, come, etc.). Even though subsequent researchers have pointed out shortcomings or limitations of this seminal study (Mesmer, 2010; Hiebert et al, 2010a; Jenkins, Peyton, Sanders, & Vandasy 2004), this basic principle of coupling decoding texts with systematic phonics instruction has been accepted as foundational for thinking about text design (Adams, 2009) and turned into educational policy mandates by states like California and Texas (Mesmer, 2010).

Complicating the question of the role of different types of text on early reading development is the recent rise to prominence of an arguably new type of text. While texts based on the phonetic regularity of English orthography have been used for years (see Pearson, 2002 for a history), states such as California and Texas define decodable texts as those requiring a high percentage of the words in a given book be made up of letters and spelling patterns that have been previously taught in the classroom, with the majority of the remaining text consisting of previously taught sight words—a design strategy referred to as lesson-to-text-matching (LTTM) (Mesmer, 2001). Previous design constraints used to moderate text difficulty such as vocabulary control, sentence length, word repetition and word frequency are only marginally utilized in decodable texts (Hiebert & Mesmer, 2006). The theoretical rationale underlying the use of LTTM decodable texts starts from the assumption of research-based, high-quality classroom instruction in phonics. In order to help beginning readers establish the habit of using phonics information to solve unknown words they encounter while reading connected texts, LTTM decodable texts allow students opportunities to practice decoding using phonics information directly related to their most recent classroom lessons. These texts are used to a small degree in kindergarten and then form the bulk of student reading material for the first two thirds of 1st grade. They also appear as support and review materials up through 3rd grade (Roit, 2008). For many struggling readers, these texts represent a significant and large component of their early text exposure.

These two factors—the lack of a wide research base looking at the specific effects of different types of text on reading development and the almost universal adoption of the new genre of LTTM decodable texts—have created a significant gap in the research literature. In short we have what appears to be a theoretically research-based practice (reading texts that allow one to practice the phonics patterns taught in classroom) with little or no empirical data to support this widespread practice. While much has been theorized about how decodable texts are supposed to work (Adams, 2009), and their use is sanctioned as a component of research-based instructional practices as mandated by No Child Left Behind (Hiebert, 2010a), to date no studies have been published demonstrating either students' actual reading performance or the longer term consequences (positive or negative) of using the daily, instructional texts based on the LTTM principle. In order to address this research gap, the goal of this dissertation project is to

collect and analyze data on 1st grade student reading performance using the decodable texts included as part of daily, classroom instruction over a five month period. In analyzing these data, I hope to present an initial picture of student reading behavior using decodable texts as well as an analysis of specific text features that might explain, or are at least potentially related to, patterns of observed reading behaviors.

A Brief History of Text Use in Early Literacy Instruction

The issue of selecting or designing texts for early literacy development has a long history and can be traced to the beginning of written language. Phoenician scholars organized lists of words by their initial sound to support novice scribes in their acquisition of the syllabary. Throughout history, everywhere that reading has been taught, decisions have been made about how best to teach reading and the type of text that best supports that instruction (see Pearson, 2002 for an overview). While it is a bit of a simplification, three primary categories help describe the nature of instructional texts used in the United States over the last 50 years, and variants of these three categories have been the mainstay of educational strategies over the last 400 years.

The "look-say" model of reading instruction dominated US public education from the early twentieth century through the mid 1960's. Huey (1908) had studied the reading behavior of university students using a tachistoscope (a device that displayed an image for a brief amount of time) where he discovered that entire words were identified with as brief exposure times as individual letters. Working from this theory of words as the most significant unit of written language, the look-say instructional model taught students to visually recognize the most frequently used words and after having developed a core set of sight words, students were taught phonics analytically. The infamous Dick and Jane texts that were matched with this model of instruction contained numerous repetitions of key sight words, simple sentence structures and highly constrained vocabulary. This model of instruction was used in roughly 90% of the classrooms in the United States during the 1960's (Pearson, 2002), up through the publication of Chall's (1967) *Learning to Read: The Great Debate*.

Flesch's (1957) publication of *Why Johnny Can't Read* strongly critiqued the look-say model of early reading instruction and set the stage for the United States Office of Education to fund a comparative analysis of different methods of early reading instruction, commonly referred to as, "The First Grade Studies." Chall's (1967) publication of *Learning to Read: The Great Debate* argued that most alternatives to the look-say method of reading instruction were an improvement, and she recommended a refocusing on early phonics instruction accompanied by texts that would allow students the opportunity to practice decoding. Based in part on these recommendations, the basal series used for early reading instruction turned away from a focus on high frequency words to incorporate the concept of decodability as a primary design criterion. Texts were often split into two categories; traditional fiction was used for independent reading and classroom discussion while phonetically regular texts were used to support skill development. Bloomfield & Barnhart's, "Let's read. A Linguistic Approach" (e.g. Dan can fan Nan) became a fairly standard model for basal content during the late 60's and 70's (Pearson, 2002).

A more recent model of text design (or in this case, text selection), most typically remembered from its re-emergence in the 1990's in California is called whole language and is based primarily on earlier work by Smith & Goodman (1971). In describing early literacy development, they argued, "The child is already programmed to learn to read. He needs written language that is both interesting and comprehensible, and teachers who understand language-

learning and who appreciate his competence as a language-learner" (p. 180). Basal programs used for this model of instruction jettisoned earlier look-say readers filled with high frequency words and linguistic readers filled with phonetically regular texts. Publishers replaced the simplified and abridged versions of classic children's literature with complete versions, ignoring issues such as vocabulary control, repetition and decodability. After a roughly six year reign as the dominant framework for teaching reading, from the late 1980s through the mid 1990s, whole language instruction was replaced in California schools with a phonics based model of instruction, instantiated in the state-adopted Open Court reading series (Saunders, 1999); it was accompanied by a new genre of decodable texts to support decoding practice and aid in the development of automatic word recognition. The primary distinguishing feature of these new decodable texts was the design strategy of aligning the phonics content and sight words included in individual books with previous classroom instruction—a design strategy called lesson-to-text-matching (LTTM)

Literature Review

Even with the breadth and depth of research on early literacy development conducted over the last hundred years, there is little to no research isolating the specific issue of the texts used for beginning reading instruction (Allington, 2005; Hiebert, 2009; Mesmer, 2010). To date, only a handful of published studies focus on the specific features of text as a potential factor in shaping or explaining student reading outcomes. While this might seem surprising in light of the persistence of controversies surrounding early literacy instruction, Menon & Hiebert (2005) describe the dilemma researchers face, saying, "Rather than attempting to separate the relative influence of texts versus instruction, the reading text is viewed here as a tool that *mediates* instruction" (p. 12). In a similar way, according to findings from the National Reading Panel (NICHD, 2000) numerous studies have been conducted looking at the relative strength and/or weakness of a wide range of instructional programs for beginning readers, but in none of those studies were the effects of texts isolated from the broader impact of the instructional program. In the cited studies, phonics based instruction and decodable texts were designed as mutually dependent components in the approaches that seemed to work best with beginning readers, thus no conclusions about the specific effectiveness of texts could be drawn. The summary booklet of the National Reading Panel report, "Put Reading First" (Armbruster, Lehr, Osborn, Adler, & NIL, 2001), listed as a key finding of scientific based research that texts used to support systematic phonics instruction should include a high number of words students can decode using previous classroom instruction. Yet the actual text of the NRP never makes this assertion and the specific section of the full report addressing the use of decodable texts for beginning readers concludes, "Surprisingly, very little research has attempted to determine whether the use of decodable books in systematic phonics programs has any influence on the progress that some or all children make in learning to read" (NICHD, 2000, p. 2-137).

To further complicate the issue of isolating the effects of specific texts on early reading development, it is difficult and costly to conduct the large scale experimental research needed to disentangle method from text. Even though public schools are typically willing to support the research community, few school districts could allow a study that placed a substantial number of students in the experimental conditions required to separate out the effects of instructional method and text if those conditions were clearly likely to produce poorer learning outcomes for the students involved (e.g. pairing systematic phonics instruction with whole language texts). Given the more recent trend for publishers to design an entire set of texts specifically matching

their instructional program instead of incorporating trade books, it is extremely rare to find a naturally occurring situation where text and method are disentangled; e.g., where two classrooms or schools use the same phonics-based instructional materials, yet one class uses decodable texts while the other uses authentic literature.

In light of these challenges, only three studies (Juel & Roper-Schnieder, 1985; Jenkins et al, 2004; Menon & Hiebert, 2005) have looked at the broad ranging effects of different types of texts on beginning reading development over extended periods of time. Other researchers have focused on snapshots of student reading behavior, analyzing reading data and textual features of passages from commonly used assessments (Hiebert et al, 2010a; Compton et al, 2004). A final group has experimentally examined the effects of small sets of supplemental materials on beginning readers (Cunningham, 2006; Mesmer, 2010). While each of these studies contributes to the developing picture of how different types of text affect early reading development, none of them provide a broader window into the actual reading behaviors of students using their daily, classroom, instructional texts and the effects those texts have on their reading development.

This literature review will first present a detailed description of lesson-to-text-matching (LTTM) decodable texts and the theoretical rationale behind their usage. With this description in place, the recent literature on the effects of different types of text on early reading development will be analyzed to provide both a general overview of the field as well as specific findings and issues relevant to the case of LTTM decodable texts. Finally, additional studies examining the issues of text difficulty rating and the impact of difficult texts on beginning readers will be addressed.

Lesson-to-Text-Matching Decodable Texts

The specific issue of designing texts to support classroom-based phonics instruction is powerfully and subtly laid out by Chall (1967). In a series of questions at the conclusion of her analysis of current basal series, she asked why children were not being given texts that allowed them to practice the specific phonics skills they had been taught in the classroom. In relation to the strategy of vocabulary control to support beginning readers, she agreed that vocabulary control is critical, but wondered, “Couldn’t some of the control result from a consideration of the phonics elements previously taught?” (p. 261). These two elements—students practicing the phonics skills they’ve learned previously in the classroom and controlling the material in texts to match this principle—form the foundation of LTTM decodable texts. While Chall (1967) had praise for a few basal series that attempted to incorporate these principles (e.g. Lippincott), these recommendations took years to work their way to the forefront of design strategies of early reading materials.

Chall (1967) laid out the general framework that undergirds LTTM decodable texts, but the details of the degree to which texts should match instruction were left unresolved. In describing the evolution of LTTM decodable texts, Hiebert et al (2010a) references the work by Beck (Beck & McCaslin, 1978; Beck & Block, 1979) who analyzed the degree to which material from current reading programs that was laid out in teachers’ guides matched the resources students were given for practice. Beck and her colleagues used the phrase, “potential for accuracy” to describe the degree to which students’ reading of text had been scaffolded by previous classroom instruction. In a response to Allington’s (1997) argument that there was no empirical evidence supporting the use of decodable texts for beginning readers, Beck (1997) argued that the evidence did in fact support texts that significantly matched previous classroom instruction and even went as far as to argue, “It would seem that about 70 to 80 percent

decodable would be reliable enough for children to refine their knowledge of the spelling-to-speech mapping system, while 30 or 50 percent is not enough" (p. 17).

This final critical element, a specification by a prominent researcher of a degree to which texts needed to match classroom instruction in order to optimally support student reading development, was taken by California and Texas as the grounds for requiring state standards for decodability for Kindergarten through 2nd grade texts. The specific language would typically require that a high percentage (75% for California) of words be decodable on the basis of elements (letter-sound correspondences and phonogram patterns) in the phonics lessons that had been previously taught in class. The majority of the remaining 25% of words were to consist of previously taught, high frequency "sight words" and only 5% of the words in any given text did not have to match either of these two requirements. Over the last few years, states have been in a legislative "arms race" to see who can enact the most stringent mandates regarding early literacy instruction and decodable texts, as documented in this press release from the National Right to Read Foundation:

> The State Board of Education of Texas approved new 1st grade Reading standards with the strongest ever real teeth mandating phonics instruction. Those rules say these textbooks must teach 70 letter-sound correspondences (LSCs) in 1st grade; that they must document five student opportunities to read phonetically-regular words containing those 70 LSCs, after all LSCs in these words have been taught; and that all student opportunities to read shall count toward Texas' 80% decodability minimum—not just those passages that publishers so designate as counting toward that figure. Through these Texas rules we are finally looking at true pro-phonics 1st grade Reading reform across America (Frey, National Right to Read Foundation, 2011)

The primary theoretical rationale for the use of LTTM decodable texts is: to support the development of the decoding-based strategies for identifying unknown words, to provide beginning readers with materials they can read accurately and fluently and to support the development of automatic word recognition through repeated successful readings (Adams, 2009; Armbruster et al, 2001; Roit, 2008). While researchers and proponents of the use of LTTM decodable texts have mostly focused on the texts' role in supporting the development of decoding-based strategies for solving unknown words, other researchers, as well as the curriculum authors, have foregrounded the role decodable texts play in supporting the development of fluent word recognition skills (Ehri, 2005; Share, 1999; Adams, 2009; Roit, 2008).

Adams (2009) discusses the fundamental purpose of decodable texts in allowing beginning readers to practice the phonics elements they were taught during classroom lessons but goes on to explain that repeated opportunities to successfully decode words encountered in text leads to fluent word recognition. "Gradually, through repeated encounters, the representation of the word and its parts become so richly and strongly interconnected that the word is recognized virtually at a glance" (p. 33). Adams describes repeated successful chances to decode as the critical factor affecting the development of automatic word recognition and argues that phonics instruction, phonemic awareness and even prior decoding sophistication only serve to support successful decoding and that successful decoding during the reading of connected text is the key to the development of fluent reading. Data from Share (1999), however, demonstrate an important counterpoint, showing that students who made errors during their initial attempts to decode target words tended to misremember those words in line with their incorrect

pronunciations. Torgesen (2002) echoes this concern, arguing that errors during decoding slow the growth of automatic word learning. There is strong support across the research community for the idea that successful reading of connected text is a critical factor supporting the development of automatic word recognition. These concerns about the impact of inaccurate reading on word recognition development will be addressed again, later in the literature review as well as in the data analysis and discussion sections.

Critics of LTTM decodable texts. Critics of LTTM decodable texts have focused on the lack of empirical research supporting specific mandates for percentages of decodability (Allington, 2005) and the ways in which the focus on lesson-to-text-matching as the single criterion determining readability ignores previous research identifying a much wider range of features affecting text difficulty and readability (Hoffman et al, 2002; Hiebert et al, 2010b). In analyzing the decodability mandates of both California and Texas, Allington (2005) argued that while the basic premise of aligning reading materials with classroom instruction makes obvious sense, the typical design of LTTM decodable texts and their usage in classrooms goes far beyond that simple premise:

> I believe that it makes no sense to teach children any aspect of decoding without providing them the opportunity to practice that aspect while reading connected text. But there is no evidence that creating the artificial but highly decodable texts that have pigs doing jigs is necessary to foster effective decoding proficiency (p. 465).

Allington (2005) cites the National Reading Panel report which found, "very little research has attempted to determine whether the use of decodable books in systematic phonics programs has any influence on the progress some or all children make in learning to read" (NRP, 2000; p. 2-134). Allington's basic argument is that one-size-fits-all reading programs that present undifferentiated classroom instruction and have struggling readers reading the same books as advanced readers disproportionally disadvantage struggling beginning readers—the group who are most dependent on public schools for instruction in learning to read. Hiebert (2009) describes data analysis demonstrating that while successful early readers learn the most common English sight words by the end of 1st grade, struggling beginning readers don't master the same words until the end of 2nd grade or later. Additionally, the authors described the tendency of LTTM decodable texts to utilize rarely repeated, low-frequency words that serve as examples of target spelling patterns (e.g., woe for the –oe family). While these words qualify as decodable according to LTTM criteria, Hiebert & Fischer (2007) showed that many of these words caused difficulty for struggling beginning readers because of their relatively low word frequency in standard usage.

As is often the case with any crucial design specifications, the devil is in the details. The general principle of LTTM decodable texts, that texts designed for beginning readers incorporate previous classroom instruction, is a simple premise that even strident critics of LTTM decodable texts wouldn't argue with. But when it comes to specific percentages of words that must meet LTTM decodability standards, state mandates, and the sufficiency of lesson-to-text-matching as the single criterion for design (Mesmer, 2010), the research community is much more divided. Proponents argue that LTTM decodable texts play a critical but short-term role in the development of beginning readers, yet the research community is far from united on this position. In order to better understand the ways in which LTTM decodable texts affect beginning readers, research looking at the impact of different types of texts on early reading development will be assessed.

Effects of Text Types on Beginning Reading Development

Of the published literature that examines the effect of different types of text on beginning readers, the work of Juel & Roper-Schnieder (1985) is by far the most commonly cited study. The authors located a setting in which they were able to conduct a naturally-occurring quasi-experiment: a single phonics-based classroom instructional program was used across a school district for the entire school year. In roughly half the schools in the district, the phonics program was paired with a set of decodable texts, in the other half; the program was paired with a set of texts that focused on high frequency words. Each set of books was divided into three levels, pre-primer, primer, and first reader and the words of each set of texts were analyzed in terms of features that might impact their recognition by students (e.g. decodability, length, number of syllables, frequency, etc.). At three points during the school year, upon completion of the books for a specific level, students were assessed using the Bryant Test of Basic Decoding Skills (Bryant, 1975). Additionally, students were shown a list of the core content words for the texts they had just finished reading and asked to read the words aloud. The final assessment in May included a combined list of words from all three levels of texts. At the end of the year, each group of students took the Iowa Test of Basic Reading Skills and each group was assessed with a word list representing core content words that had been included in the other group's basal reading series but not their own.

An analysis of the pre-primer level of texts for both sets showed that decodable texts were rated as significantly easier to decode than the high frequency word texts. An analysis of regression data from the word recognition assessment highlighted word-level factors that were most powerful in explaining the variance in scores between the two groups. These findings suggested the types of strategies students were using to solve unknown words. For students reading the pre-primer decodable texts, word decodability predicted the greatest degree of variance in word recognition scores, suggesting these students primarily used decoding-based strategies to identify unknown words. For students reading the pre-primer high frequency word texts, the number of letters in a word and the number of syllables predicted the greatest degree of variance, suggesting these students were primarily attending to visual features of words to aid in identification. The authors interpreted these results, saying, "The Economy (i.e. decodable) pre-primers also appear to induce an earlier, and more lasting use, of a phonological strategy [to identify words] based on letter-sound correspondences than in Houghton-Mifflin (i.e. high frequency word pre-primers)" (p. 150).

For students reading decodable texts, end of the year assessments showed that word decodability continued as the strongest predictor of accuracy during word recognition assessments, suggesting these students were inclined to use decoding strategies to solve unknown words. For students who were instructed using the texts with a high percentage of irregularly spelled sight words, number of repetitions was the strongest predictor of accuracy on end of the year word recognition assessments, suggesting that their primary strategy for solving unknown words was visual analysis. Additionally, when given a list of previously unseen words from the basal series the other group of students had practiced with, students who had practiced with decodable texts were able to recognize significantly more words than students who had practiced with the less-decodable texts. These findings led the authors to conclude, "Results of the current study suggest that the types of word which appear in beginning reading texts may well exert a more powerful influence in shaping children's word identification strategies than the method of reading instruction" (p. 151).

While this study has been discussed at length by a variety of authors (see Adams, 2009; Hiebert et al 2010a; Mesmer, 2010), there are a number of important issues that warrant revisiting, especially in light of current discussions concerning decodability and the effects of different types of text on beginning reading development. First, the students in the study were carefully selected to avoid both low-performing students and high-performing students. Students below the 40th percentile on the Metropolitan Readiness Test were not included as participants and according to the authors, "the children were non-readers upon entering first grade as determined by teacher and experimenter screening on reading word lists and informal reading inventories" (Juel & Roper-Schnieder, 1985, p. 138). All of the students were placed in mid-level reading groups in their respective classrooms. The decision to exclude both low-performing and high-performing students reduces the variability in the learning task faced by study participants. However, it severely limits the ability to make generalized assertions as to the effects of different types of texts on early reading development for students of different ability level.

Second, the books in the basal series used by students in the decodable texts condition of the study were phonetically regular, but they were not aligned with the phonics instruction used in the classroom—i.e. there was no lesson-to-text-matching (LTTM). The beneficial effect of the texts in supporting development of decoding-based strategies for solving unknown words was entirely independent of a structured relationship between the classroom phonics lessons and the specific content of the books that students read during the course of the school year. This leaves the most cited study looking at the effects of different types of text in supporting early reading development with nothing at all to say in regard to the issue of LTTM. Given the equivalent gains of both groups of students over the course of the year across a variety of measures, the study actually demonstrates significant learning gains without the use of LTTM decodable texts, and equivalent gains for the group using texts specifically focused on phonetically irregular, high frequency words.

Third, not only did the decodable texts in Juel & Roper-Schnieder (1985) not match classroom instruction, but by the middle of the school year, the texts focused on high frequency words were as phonetically regular if not more so than the decodable texts. While this might sound paradoxical, this highlights one of the less well known but consistent findings across studies looking at the nature of early texts. Both sets of texts were divided into three difficulty levels (pre-primer, primer and first reader) and decodability was computed for the core content words in each set of texts by difficulty level. According to the measure used by Juel & Roper-Schnieder, lower decodability ratings indicated words that were easier to decode (see table 1). As was to be expected, words in the pre-primer level of the high frequency texts were more difficult to decode than words in the pre-primer decodable texts. However, the primer and first reader high frequency texts contained words with lower or equal decodability ratings (lower ratings means easier to decode) than the decodable texts. These data led the authors to conclude, "The content words in Economy pre-primers (i.e. decodable pre-primers) tend to … follow more regular letter-sound patterns … However, words in primer and first reader appear more similar between series and, if anything, Economy has slightly longer, less regular words" (p. 141). These data regarding equivalent decodability ratings for the two sets of texts significantly complicates the authors' conclusion that the types of words appearing in a text affect the development of student word identification strategies.

Fourth, and most importantly, the authors summarize their study by arguing, "the text children are exposed to early in first grade may differentially shape their word identification strategies" (p. 134). The problem with this conclusion, however, comes from an analysis of the

regression data when both groups of students had completed reading the primer level texts in February, a point roughly half way through the school year. For students using the high

Table 1

Text decodability ratings by set and by difficulty level for Juel & Roper-Schnieder (1985)

	Decodable Texts	High Frequency Texts
pre-primer	1.2 (.5)	1.8 (.8)
primer	2.0 (.6)	1.8 (.8)
first reader	2.0 (.7)	2.0 (.8)

Note: For decodability, 1 = predictable, easy; 2 = predictable, hard; 3 = predictable, irregular

frequency texts, decodability explained the greatest degree of variance in their word recognition scores, nearly six times higher than the second closest factor. It also explained a greater degree of variance for high frequency text users than it did for students using the decodable texts. The amount of variance explained by decodability increased greatly for both groups from the pre-primer texts to the primer texts (from .079 to .118 for decodable texts students and from .038 to .147 for high frequency text students) seemingly indicating that both groups were increasing in their capacity to use decoding-based strategies to solve unknown words. This finding is seemingly at odds with the authors' contention that the decodable pre-primers were the critical factor, establishing decoding as the primary word identification strategy for students using decodable texts. If the authors' analysis is correct, i.e., that the percentage of variance predicted by word decodability is an indication of students using decoding-based strategies to solve unknown words, then students in the high frequency word texts condition developed a stronger habit of using decoding-based strategies to solve unknown words halfway through the school year than did students in the decodable texts condition, even though they began the year with texts that were less easily decoded. Noting this finding, the authors commented, "It appears that a phonological strategy predominates in the middle of first grade" (p. 144). Yet there is no explanation for how the development of this focus on phonological information might have come about for the students reading the high frequency word texts.

The authors cite the pre-primer and end-of-year combined assessment word recognition data as most clearly representing their argument that students having read decodable texts developed stronger decoding-based strategies of word identification and that students reading the high frequency word texts developed more visually oriented strategies of word identification. While this is an entirely reasonable conclusion, an alternative interpretation of the same data would suggest that the overall story is quite complicated and does not support the authors' primary contention. Since one of the purposes of early texts are to support the development of students' automatic word recognition, it could also be argued that the end of year data for students using the high frequency word texts demonstrates Ehri's (2005) model of word learning. Students reading the high frequency word texts began the year with fairly simple visual analysis in the pre-primers, progressed to more thorough decoding with the primers, and by the end of the year, the students' repeated success in decoding core content words led them to automatically recognize the majority of the words they had seen throughout the year. This seems a much more plausible explanation for why students who had been using decoding as their primary strategy back in February would seemingly "switch" primary strategies for identifying words. The authors themselves could find no explanation for why students who had started as visual processors and then successfully moved to decoding would simply let go of decoding and go

back to a less useful strategy. The progression from visual analysis to decoding-based strategies to automatic word recognition is the goal of early reading development and this learning progression fits the data much more powerfully and simply.

Additionally, one might ask an important related question, why were students reading the decodable texts not more affected by the number of repetitions? At what point should students move from decoding-based strategies for solving words they've seen repeatedly over the course of a year to automatically recognizing the words? The regression data for word recognition scores for structure words (pronouns, prepositions, quantifiers, helping verbs, etc.) from all three text levels shows that for students having read the decodable texts, even for words the students had seen an average of 115 times each, decodability (.072) was a close second in terms of predicting the variance in word recognition scores compared to number of repetitions (.084). For students having read the high frequency word texts, even though the average number of repetitions for their structure words (100) was lower, the number of repetitions explained 21.8% of the variance in their word recognition scores while decodability explained only 1.3%. These data could be seen as suggesting that students having read the decodable texts over-relied on decoding-based strategies for solving words and that their ongoing reliance on phonological analysis for recognizing words they had encountered repeatedly is one of the critical findings of this study.

There are too many interesting findings and important questions raised by the data from Juel & Roper-Schnieder (1985) to simply ignore their primary conclusions. However, because of the limited range of students involved, the lack of clear distinctiveness in the texts involved, the limited degree to which word-level factors predict the variance in word recognition assessments, the small amount of variance predicted by key word-level factors, the availability of alternative interpretations for many of their data and their lack of any significant finding on any of the measures from the end of the year standardized assessment, care should be used in drawing conclusions about the reading behaviors of these students and the implications for the design of materials for beginning readers in general.

Other Long Term Studies Analyzing the Role of Texts in Early Reading Development

Juel & Roper-Schnieder's (1985) discovery of converging ratings of decodability as texts increase in difficulty seems to play a significant role in Jenkins, Peyton, Sanders & Vadasy's (2004) contrary findings that different types of text had no significant effect on a wide range of reading outcomes in an intervention study for struggling 1st grade readers. Students were randomly assigned either to a control group, a more decodable text group or a less decodable text group. For the first 30 lessons, 85% of the words in the more decodable texts were classified as decodable (i.e. could be solved using phonics information that had been previously taught in the classroom), whereas only 11% of the words in the less decodable texts could be solved using the same information. Students in both experimental conditions significantly outperformed control group students, but there were no significant differences between students who had been reading more decodable texts as compared to students who had been reading less decodable texts across any of the study measures, challenging earlier findings from Juel & Roper-Schnieder (1985) and arguments from Beck (1997) that students require highly decodable texts to support the development of decoding-based strategies for solving unknown words.

While the authors highlight the decodability differences between the texts used in the beginning of the intervention, by the end of the study, the less-decodable texts were scored as 68.5% decodable as compared to 80.4% decodable for the texts in the more decodable condition.

This raises the obvious question, exactly what percentage of decodable words do students need in order to develop a habit of decoding and when do they need it? Additionally, decodability was measured by comparing the phonics information required to solve a specific word with previous classroom instruction. But many of the words in the less decodable text series were one syllable, phonetically regular words, yet because they included even a single spelling pattern that had not been previously taught in the classroom at the exact point the words were encountered in text, they did not count as decodable according to the LTTM criteria. Adams (2009) had argued prior to the design of the LTTM Open Court decodable texts that the basic consonants should be used in the texts from the start, a position that was "outlawed" by the State of California in favor of the lesson-to-text-matching principle that governs text design. But based on these results, there is no evidence that LTTM texts benefit beginning readers in any significant way as compared to less decodable texts that do not meet LTTM standards.

Juel & Roper-Schnieder (1985) argued that texts might exert a greater influence on beginning readers than the method of instruction. The specific intervention design in the Jenkins et al (2004) study seems to provide evidence against that conclusion. During text reading, when students encountered a word they were unable to read, tutors "prompted them [students] to use previously taught phonics skills" (p. 62), having them sound out words and providing any missing phonetic information the students required. If the primary goal of text reading is to support the development of decoding-based strategies in solving unknown words, one-to-one tutoring that reminds students to practice decoding each time they encounter a word they can't solve would have to be seen as an instructional model that could readily lead to helping beginning readers develop a habit of decoding, especially given that almost no words in any set of books for beginning readers are entirely undecodable. Ironically, given the apparent interaction of instructional model potentially mediating the effect of the different types of text, it is impossible to isolate the effectiveness (or lack thereof) for either type of text. However, Jenkins et al's (2004) hypothesis was simply that LTTM decodable texts provide no measurable benefit for struggling beginning readers—a hypothesis that was strongly supported by their results.

In the last of the longer term studies, Menon & Hiebert (2005) implemented a quasi-experimental intervention design where two groups of 1st grade students read specially selected little books while two control groups read from the literature basal series being used in their classrooms. The authors described the significant differences between the two sets of text in detail, focusing on word decodability, number of word repetitions and sentence structure. Because of the small number of students that came from four different classrooms, the findings are characterized by the authors as, "an initial investigation of a model of text that is designed to support independent word-solving skills in beginning readers" (p. 32). Their clear finding, across a wide variety of measures, was that moderately decodable texts supported a greater development of word-solving skills for first graders of all reading abilities in comparison to texts from the literature basal series. The authors also noted, "These results do not suggest, however, that beginning readers require texts where all words fit particular patterns or where each unique word is repeated a particular number of times" (p. 33).

Hiebert and Menon's study also indirectly addresses the issue of all students using the same texts, regardless of reading ability. In discussing potential shortcomings of their design, the authors acknowledged that the degree to which the moderately decodable texts used in the study could be matched to student reading levels might have played a role in the gains made by students using those texts. They point out, however, that the design of the current literature basal

series assumed that undifferentiated materials could support struggling readers, and that their findings argue against that assumption. This question about the impact of undifferentiated texts being used for readers of all levels is a critical question that will be addressed in the design and analysis of results of this research project.

Effects of Word-Level Features on Reading Performance

Moving from the three longer term studies looking at the effects of different types of texts on early reading development, the next set of studies represent a focus on analyzing word level features of texts that affect reading performance and development. Compton et al (2004) analyzed 2nd grade student performance reading a set of passages used for curriculum based measurement. The passages were analyzed according to a variety of readability formulas as well as analyzed for a range of word-level and text-level features such as word length, sentence length, decodability, word frequency, etc. Their large and varied sample (n = 248) allowed them to differentiate their findings specifically for readers of different achievement levels. For the entire population, the percentage of high frequency words per sentence significantly predicted accuracy whereas passage decodability (as indexed by the percentage of decodable words, not lesson-to-text-matching criteria) had no significantly measured impact on reading accuracy. Both passage decodability and the percentage of high frequency words predicted reading fluency. When the analysis was limited to low-achieving students, the percentage of high frequency words influenced both accuracy and fluency but decodability had no significant impact.

Hiebert, Stewart & Uzicanin (2010) studied the impact of word features on 1st and 2nd grade reading performance of the DIBELS oral reading fluency assessment passages. Data from the oral reading fluency sub-test for 14,064 students were chosen randomly from a database of 3.6 million. Words in the passages were coded on seven dimensions including decodability, word frequency and imageability. Their non-experimental data analysis project utilized a series of regressions to determine the degree to which varying word-level factors affected individual word recognition performance for readers of different ability levels. Only word frequency, word length and imageability were shown to significantly predict accuracy. Most notably, however, was that none of the three measures of decodability significantly predicted reading accuracy for students of any reading level.

In attempting to clarify and frame this finding, the authors acknowledge that the decodability rating system was not based on lesson-to-text-matching since the DIBELS assessment is given nationally and no LTTM criteria could be generated for different instructional programs. Additionally, the authors wondered if the decoding-level challenges of the DIBELS assessment were not as rigorous as might be the case with other types of texts. But no matter what the reason, across five different texts of roughly 100 words each, at levels ranging from mid-first year through end of second grade, no decodability-related measure significantly predicted the accuracy of student word reading for students of any reading level. The importance of this finding cannot be overstated in relation to the arguments regarding the supposed efficacy of LTTM decodable texts. The authors go so far as to speculate about the possibility that the nature of current LTTM decodable texts might contribute to the very problem they are designed to avoid. They argue that because LTTM decodable texts include a high percentage of decodable low-frequency words that are rarely repeated across texts, the only words the students see repeatedly are the most common, high frequency words. Of the words read accurately by at least 75% of the lowest performing readers, all but one were among the 100 most common words in the English language.

Experimental Studies Looking at the Effect of Text on Beginning Readers

Mesmer (2010) includes a detailed review of the limited literature on the effects of different types of text on reading development in an attempt to address what she describes as the muddled evidentiary picture regarding the relative benefits of leveled and decodable texts in supporting accuracy and fluency. Mesmer worked with 74 1st grade students, having them read an appropriately leveled decodable text as well as an appropriately leveled qualitatively-leveled text at four different points during a school year. She analyzed student reading data for accuracy and fluency hoping to clarify the effect of different types of texts on student reading performance. Her results, however, were mixed and served to further confound the debate. Data from the first set of students showed they were significantly more accurate reading decodable texts vs. the qualitatively-leveled texts. Data from the second set of students, however, showed the opposite outcome. For the first set of students, practice had no significant effect on their reading accuracy scores whereas for the second set of students, practice had a significant effect on reading accuracy. Additionally, both sets of students read the decodable texts significantly less fluently than they read qualitatively leveled-texts—both during their initial readings and across all subsequent retesting over the course of the year. Mesmer also found a significant text by practice interaction, indicating that students reading the leveled texts were getting significantly more fluent with practice than were students who read the decodable texts.

In order to determine an appropriate decodable text for students in the decodable texts condition, texts were selected that only included phonics information for which students demonstrated mastery, based on assessments given by the researchers. While there are clear theoretical grounds for determining an appropriate decodable text by this means, it is a situation rarely encountered by typical struggling readers (Torgesen, 2002; Hiebert, 2009) and a concern that Mesmer herself identifies. The instructional practice of lesson-to-text-matching is built on the assumption that instructed means learned. Texts are classified as decodable not based on a student's demonstrated mastery of the requisite phonics material but because that material has been included in previous classroom instruction, whether or not the student was absent or paying attention, let alone that they demonstrated mastery. Thus Mesmer's students were very likely much better qualified to read the specific decodable texts used in the study than a typical struggling reader working in a typical commercial program would be.

Even with the supportive definition of appropriately decodable text, however, all four decodable text readings from the first set of students had accuracy scores below 94% and three of the four readings in the second set of students were below 94% accuracy. Thus even when students read a selection of decodable texts specifically matched to their actual phonics knowledge, they still struggled to read the texts accurately and fluently. The lack of clear, significant benefit for students reading decodable texts on measures of reading accuracy and fluency again demonstrates the lack of empirical support concerning the efficacy of LTTM decodable texts for beginning readers.

The work of Cunningham (2006), while not conducted to examine the effect of text on reading behavior, provides some serendipitous insights on the question. Cunningham set out to test the effect of textual coherence on student word learning. Her research challenged the argument that when students have alternative avenues available for identifying unknown words (such as contextual clues) they will be less likely to attend to the orthographic features of words, which will decrease the likelihood that they will form a visual representation of a word's spelling, which, in turn, will make it less likely that they will learn the word orthographically.

She presented students with a set of 16 target words, selected because they have a high likelihood being known orally but a low likelihood of being recognized automatically in print. Students read the words embedded in a set of eight carefully written passages, designed for easy readability and minimal predictability. Students read half of the passages as they were written and the second half with word order in the passage scrambled. The videotaped results were scored for accuracy and fluency, and a variety of post-test measures were used to assess word learning outcomes. Per her hypothesis, Cunningham (2006) showed that there was no significant difference across any of the word learning measures for students in the context supported condition vs. the scrambled order condition. She concluded that even though students utilized the non-visual clues made available by context in order to help identify unknown words (they were significantly more accurate in the context vs. the scrambled conditions), the contextual supports did not negatively impact their orthographic learning.

The interesting finding in relation to the concern of this paper is the strong correlation ($r = .66$, $p<.001$) between accuracy and word learning. Independent of condition, if a student decoded the word correctly each time during the practice sessions, they were much more likely to identify it accurately during the post-test. A reanalysis of her original data show that with each mistake a participant made during practice, the likelihood of reading the word accurately on the post-test (i.e. having learned the word) decreased. In situations where students made even a single error in their opportunities to decode a target word, those words were significantly less likely to be identified on the word identification post-test. In order to make as level of a playing field as possible in testing the validity of her primary hypothesis, Cunningham compared word learning outcomes only on words where the participants had been equally successful in the scrambled condition as they were in the context condition. But because participants were significantly more successful in the context condition, they learned more total words in the context condition than they did in the scrambled order condition. Cunningham (2006) provides empirical evidence supporting the theoretical arguments of Torgesen (2002) and Share (1999), that errors during initial attempts to decode words encountered in text undermine a student's ability to learn words while reading text.

Summary of the Literature Review

Decodable texts built around the lesson-to-text-matching design criteria are still mandated in both California and Texas and used broadly across the United States public educational system as a whole. LTTM texts require that a high percentage of the words in any books used for beginning readers consist primarily of words made up of spelling patterns that have been previously taught during classroom instruction, or, previously taught sight words. The goal of LTTM texts is to support beginning readers in using decoding-based strategies to identify unknown words while reading connected text and to support the development of automatic word recognition. While the research supporting the role of phonics instruction is well-established, there is little available research demonstrating the specific effects of different types of texts on beginning readers. Of the three studies looking at the long term effects of texts on beginning readers, Juel and Roper-Schnieder (1985) concluded that students with early and consistent exposure to highly decodable texts were more likely to use decoding-based strategies to solve unknown words than students who were exposed to less decodable texts. However, when their results are examined more closely, they present a mixed picture of the exact effects of decodable texts on beginning readers and their study says nothing about LTTM decodable texts. Jenkins et al (2004), in an intervention study comparing outcomes across a wide variety of measures found

no significant effects due to students using either more or less-decodable texts during the intervention. Hiebert and Menon (2005) showed that students reading moderately decodable texts with no lesson-to-text-matching significantly outperformed students reading a literature basal series across a variety of measures.

In a pair of studies looking at text features affecting beginning readers, both Compton et al (2004) and Hiebert et al (2010a) found word frequency to be critical in predicting reading accuracy while in neither study was decodability a significant predictor of reading accuracy. In two experimental studies looking at the role of text in early reading, Mesmer (2010) found inconsistent results between studies when she compared reading accuracy for LTTM decodable texts versus qualitatively leveled texts; by contrast, in both studies, qualitatively leveled texts better supported the development of fluent reading. Cunningham (2006) demonstrated a strong correlation between reading accuracy and word learning; in particular, she found that errors during text reading led to reduced word learning outcomes.

The clear conclusions from this review of the research literature is that there is little empirical research analyzing the effects of different types of texts on early reading development and the few studies that do evaluate those effects reveal that at this point there is no evidence supporting the use of LTTM decodable texts for beginning readers. More specifically, to date, there is no evidence demonstrating a critical period where LTTM decodable texts are especially useful, no evidence linking their extended use to measurable gains on any standardized tests, and no evidence from either intervention studies or correlational studies that LTTM decodable texts are especially effective as compared to alternatives. Yet amidst requirements from No Child Left Behind that classroom instruction be based on scientific research, the two largest states in the country, Texas and California, mandate that LTTM decodable texts be used for beginning reading instruction.

This non-experimental study attempts to address some of the gaps in the research literature by describing the reading behavior of 45 1st grade students over a five month period while reading their daily, instructional LTTM decodable texts. The goal of this study is to shed light on student reading behavior across a substantial period of time during a critical stage of their reading development—early to mid grade 1, when students are required to read LTTM decodable texts as their primary practice material. Through analysis of data from this key swath of development, I hope to evaluate the effectiveness of a widely used example of implementing the lesson-to-text-matching criterion for designing decodable texts.

Methods

Participants

The participants come from two 1st grade classes at a single elementary school in Oakland during the 2009-2010 school year. The 47 first grade students (26 girls and 21 boys) at Lockwood Elementary School (pseudonym) are part of a student population of 253 students, a mix of 40% African American, 28% Caucasian, 17% mixed race or non-responding, 8% Asian and 6% Latino. Of the 253 total students (K- 5th grade), 20% participated in the free and reduced lunch program and just 2% of the students were English language learners. The school's academic performance index (API) for 2009 was 864, placing it in the 80th percentile of California schools. The school was selected because of its proximity, its willingness to allow access, its use of the Open Court Reading program and its mandated use of the DIBELS assessment that could be used as an external index for assessing student reading level.

Students came from two reasonably similar 1st grade classes taught by experienced teachers (14 and 5 years of experience respectively). Both teachers implemented the Open Court reading program with similar fidelity; apropos of the study, both required students to read their decodable book of the day more than 90% of the time. Both teachers followed proscribed lesson plans for introducing new books to students and had roughly similar strategies for whole class shared reading and independent reading, with help provided if necessary. Both teachers also allowed for a small number of their advanced readers (n = 11 between both classes) to skip the whole class book instruction; instead they read the book independently at their desk and then moved to free independent reading. Both teachers read frequently to their classes, using standards from children's literature, and both teachers had substantial classroom libraries that students were able to access during free choice times.

All students from both classes initially participated in the data collection. Two students (one from each class) had been previously identified as learning disabled and both were receiving special intervention services. In order to avoid having two hugely disparate sets of scores skew the data, scores for both students were dropped from the analysis. Statistical analysis comparing the beginning of the year DIBELS scores for the two classes showed no significant differences, so teacher was dropped as an independent variable and students were treated as a single group.

Materials

The books in question come from SRA/McGraw Hill's Open Court Reading program. The program includes 118 decodable texts for (nearly) daily reading; they are sequentially ordered to match the lessons on various phonics elements. The decodable texts are used from the beginning of the school year until roughly 2/3 of the way through the first grade year, at which point literature anthologies become the primary instructional materials for classroom and independent reading. All of the texts read by students are part of a set of texts that have been designed according to California's mandates regarding decodable texts and are part of a reading curriculum that has been certified by California as using scientifically based instructional practices. Each text is designed to include at least 75% decodable words based on previously taught phonics elements as the decodability standard. Of the remaining words, 15% - 20% are previously taught sight words, yielding an overall LTTM rate of close to 95%. Texts from the Open Court Reading program were selected because of the program's widespread usage both in California and across the nation. All of the current core reading programs in 2010 (Scott Foresman's *Reading Street*, MacMillan/McGraw Hill's *Treasures*, Harcourt's *Storytown*, and SRA's *Imagine it!*) have sets of decodable texts modeled after those of Open Court, 2000 (Hiebert, 2010).

Accuracy data were primarily analyzed on the first 50 words of each of the 24 books sampled in this study (n = 1181, one of the books only had 32 and another had only 49 words, thus, 1181 total words). 381 of the 1181 words (32.4%) were sight words (as classified by Open Court), 783 (66.3%) were decodable words and 17 (1.4%) were story words. Of the 783 decodable words, 110 (14%) were decodable proper names (e.g. Kim). These percentages do not at first glance seem to match the state mandates. The most likely explanation has to do with how sight words and decodable words are counted. The word "and" was introduced in decodable book 1, before the letter sounds for 'n' and 'd' had been taught, so each instance of the word "and" would count as a sight word. By book 14, however, each of the letter sounds in "and" had been taught, thus each instance of the word could be counted as decodable from then on. Given

that many of the 128 sight words taught as part of Open Court 1st grade reading are decodable based on previous phonics instruction most likely explains the discrepancy.

Data Collection

In order to assess the performance of students with typical, daily-use, LTTM decodable texts, I collected oral reading records from individual students in two 1st grade classrooms in an urban school district in Northern California during the 2010 – 2011 school year. Roughly twice per week I met with students from each class in a one-on-one setting and had them read their most recently "instructed" decodable text. Each of the texts had been read by the students in two settings, first in a whole-group, guided reading setting with their classroom teacher. In these settings, teachers provided support focusing on identifying the specific phonics information that had been taught in class that was designed into the text and identifying words that might be problematic. Afterward, each student also read the book independently at his or her desk prior to reading with me. Additionally, each student had been given a copy of the book to take home and read independently as homework before they read with me. Data for any student who had been absent during the initial classroom reading of a text were not recorded. Due to a variety of factors including field trips, assemblies, holidays, testing and occasional logistical difficulties, the sequence of books from which data were collected does not follow the exact pattern of presentation across the lessons. Data were collected from student readings of 24 books, starting near the end of November and ending in mid April. The strategy for data collection was designed to create a large enough sample to give a clear picture of student reading performance across the program, to identify patterns in student performance over time and to guard against confounds that might arise from using any individual book as the basis of inferences about student performance. The specific schedule for data collection was structured around the ongoing classroom schedules of the teachers.

On a day scheduled for reading assessment, I would show up during the morning block of literacy instruction, sit at a table in the back of the classroom, and call students one at a time to come and read with me. I would show them the decodable text, ask them if they remembered reading it the day before and then have them read it out loud to me. I kept a detailed record of the student's reading, using a modified running record system to record mistakes; keeping detailed accounts of their substitutions, omissions, pauses, self-corrections and requests for help. In situations where I was unable to identify what a specific student had said, the response was marked simply as an error. Each record was scored for accuracy and fluency (words correct per minute).

In keeping with the model of reading assessment used in both DIBELS and most curriculum embedded assessments, students were asked to read the text independently. If they were unable to read a word correctly after three seconds, I told them the word and marked it as an error. Students who generated errors but continued reading were not corrected. Students were given two minutes to read as much of the text as they could. For students who finished the text before the allotted time, their actual reading time was recorded and used for fluency calculations. For students who were unable to finish, the number of words read by the end of the two minutes was recorded and used to create a fluency score. From earlier research I had found stopping a student in the middle of a book to be a somewhat frustrating experience for children. So on many occasions, I allowed students to continue reading the book and continued to mark their responses. The responses from these "extra" reads were kept separate from the primary data used

for accuracy and fluency calculations, but I did analyze them in a search for patterns of text features that seemed to explain, or at least correlate with, student reading errors.

In order to provide an external reference measure of student reading level and specific reading skills, student performance data on the beginning of year, mid-year and end of year DIBELS assessments were collected.

Results

The results section is broken into two sub-sections—results related to student reading performance and results related to an analysis of text features and their impact on student reading performance. The fundamental goal of using lesson-to-text-matching (LTTM) decodable texts is to allow students opportunities to practice what they have been taught previously in the classroom, and through successful, repeated readings of texts, to support the development of fluent decoding and automatic word recognition. Thus the data will be analyzed to assess the validity of this rationale—do students, especially struggling readers, increase in fluency and develop their ability to recognize words automatically when they read LTTM decodable texts?

Additionally, the principle of lesson-to-text-matching (mostly in response to state mandates of requiring specific percentages of words in a text that must meet these decodability standards) does not take into account potential differences in the difficulty of equally decodable words. Thus *cat* and *trudged* are both decodable words for the sake of LTTM decodability standards, but they represent very different challenges in terms of word frequency, length and how often they are seen in text. The data will be analyzed to assess the efficacy of the assumption that decodability trumps other aspects of word difficulty.

Student Reading Performance

According to data from the DIBELS (Good, Wallin, Simmons, Kame'enui, & Kaminski, 2002) assessment given to all students in the district at the beginning, middle and end of the year,

Table 2

Percentage and Number of Students Falling in DIBELS Percentile ranks

DIBELS Sub-Test	25th	50th	75th	top
Beginning of year assessments				
Letter Naming Fluency	8	12	12	13
Phonemic Segmentation Fluency	2	7	15	11
Nonsense Word Fluency	4	14	10	17
Mid-Year assessments				
Phonemic Segmentation Fluency	1	9	20	15
Nonsense Word Fluency	3	7	11	24
Oral Reading Fluency	3	9	9	24

Note: Total students = 45, assessments given in September and February

the students in this study represent a higher than average population of readers. In both 1st grade classrooms, across each of the DIBELS measures taken from the beginning-of-year and mid-year assessments, there were no significant differences between students from the two different classrooms at either point, thus classroom teacher was dropped as an independent variable.

Student performance data across the DIBELS measures were distributed with a noticeable negative skew (see table 2), again reflecting the higher than average scores of this population.

Following an analytical strategy used in Hiebert et al (2010a), students were divided into terciles in order to support an analysis of student reading performance by ability level (see Table 3). Tercile-based analysis identifies three fairly distinct groups of students and their relationship to early literacy instruction. High-performing students typically come to school already knowing most of what an early literacy curriculum attempts to teach (Hiebert et al, 2010a). Mid-performing students come with sufficient background knowledge and language ability to take advantage of classroom instruction and successfully learn to read through classroom instruction (Adams, 2009; Juel & Roper-Schnieder, 1985). Low-performing students come to school with limited language and literacy skills and struggle to learn to read through classroom instruction (Hart & Risley, 1995; Neuman & Celano, 2006). Neuman & Celano (2006) argue that a variety of factors, such as family income and mother's education level, affect the specific size and distribution of each of these three groups in any given school or school district.

A composite score taken across all DIBELS measures was used to compute student ranking for determining terciles, yielding 15 students in each group. The data show that the top- and mid-tercile students are well above average nationally across most DIBELS measures for their respective terciles and that the bottom tercile students have average scores at the top or beyond the top of scores for their tercile. The average percentile ranking for students in the top tercile across all DIBELS beginning-of-year and mid-year measures was 91st percentile. For students in the middle tercile, they averaged in the 71st percentile and study participants in the bottom tercile averaged in the 46th percentile across all DIBELS measures. Roughly 75% of students scored in the "At or Above Benchmark" range according to DIBELS norms (Good et al, 2002) across all measures. 20% of students scored in the "Below Benchmark" range and less than 5% of students scored in the "Well Below Benchmark" range on any of the assessments.

Table 3

Percentile Ranks Across DIBELS Measures by Tercile

	Beginning-of-year			Mid-year			End-of-year		
Group	LNF	PSF	NWF	PSF	NWF	ORF	PSF	NWF	ORF
Top tercile	87	84	95	67	95	89	69	90	93
Middle Tercile	53	74	68	64	90	78	58	79	66
Bottom Tercile	26	53	29	76	29	44	71	43	35
Total	61	72	76	69	70	80	66	77	70

Note: LNF - letter naming fluency, PSF - phonemic segmentation fluency, NWF - nonsense word fluency, ORF - oral reading fluency

Accuracy and fluency data. The data in this dataset is non-experimental—there are no control groups, treatment groups and no experimental conditions. The data most relevant to the question of the role of LTTM decodable texts in the students' reading development come from all 45 1st grade students at the site as they read 24 different decodable texts over the course of five months of grade 1. Because of the substantial differences in reading performance across terciles, the data used for calculating both accuracy and fluency scores come from the first 50 words of each text (n = 1181, one of the 24 books had 32 words, one had 49 words). Similar to Mesmer (2010) and Compton et al (2004), the primary student performance measures in the data

set are accuracy and fluency scores. The calculation for reading accuracy involved subtracting the number of errors from the total number of words read and then dividing by the total number of words read (i.e. (wr – err)/wr). Reading fluency was a measure of the number of words read correctly in a given amount of time (in this case, two minutes), expressed as words per minute (wpm).

In relation to the specific research question of how do 1st grade students perform when reading their daily, instructional decodable texts, table 4 shows the average accuracy and fluency scores by tercile as well as average scores for all students. Students in the top tercile made on average one error every 100 words, students in the middle tercile made on average an error every 20 words while students in the lowest tercile made on average an error every seven words.

Table 4

Summary of Average Accuracy & Fluency Results

Group	Accuracy	sd	Fluency	sd	Total Errors	sd
Top-tercile (n = 15)	98.9%	1.3%	91.2	15.0	0.5	0.5
Mid-Tercile (n = 15)	94.9%	3.5%	57.7	15.1	2.6	1.8
Bottom-Tercile (n = 15)	85.9%	7.3%	31.7	10.1	7.1	2.7
Average for all students	93.2%	7.2%	60.2	28.0	3.0	3.1

Notes: Summary data are averaged from individual data from all books read by each student. Fluency data calculated on entire text read by students in allotted time, accuracy score and total error data collected from first 50 words of each text.

Fluency scores give an approximate measure of the degree to which both the sight words and words consisting of previously taught spelling patterns have become sufficiently internalized to support rapid, automatic word recognition (Adams, 2009; Ehri, 2005). Fluency scores are typically correlated with accuracy scores (Mesmer, 2010; Fuchs, Fuchs, Hosp & Jenkins, 2001) although the directionality of the relationship is a matter of ongoing research and debate.

Borrowing a framework from Betts (1946), the DIBELS assessment literature describes three classifications of texts in terms of their difficulty as measured by student performance: frustrational texts that students read at or below 94% accuracy, instructional texts that students can read between 94% and 97% accuracy and independent texts that students can read at 97%

Table 5

Percentage of Total Readings Broken Down by Betts Categories

Betts (1946) Category	Total	Top	Mid	Bottom
Independent (accuracy > 97%)	42.20%	84.30%	50.20%	8.10%
Instructional (accuracy between 94% - 97%)	15.20%	9.60%	19.00%	15.30%
Frustrational (accuracy < 94%)	43.50%	6.10%	30.80%	76.40%
* Readings between 80% - 90% accuracy	27.40%	1.50%	12.50%	31.00%
* Readings under 80% accuracy	10.40%	0.00%	3.50%	23.30%

Notes: Summary data is averaged from individual data from all books read by each student. Fluency data calculated on entire text read by students in allotted time, accuracy score and total error data collected from first 50 words of each text.

accuracy or better (University of Oregon, 2012). While Betts' (1946) framework was originally created as a tool for assessing minimal levels of reading accuracy to support comprehension, teachers and researchers have used the framework over the years as a rough guide for interactively assessing text difficulty (Lipson & Wixson, 1986; University of Oregon, 2012). Table 5 shows the percentage of student readings in which the specific accuracy score for a given book fell into each of the Betts categories. For example, of the 334 total readings by students in the lowest tercile (24 books x 15 students equals 360 total possible readings, 334 accounts for absences, etc.), 255 of those readings (76.4%) had accuracy scores of less than 94% while 27 out of the 334 readings (8.1%) had accuracy scores at 97% or above.

Blocks of reading errors. An additional index of reading accuracy involves identifying blocks of text where students made enough errors to substantially undermine their ability to continue reading the text let alone make any sense from it. While there are no absolute or transparent criteria for this measure, identifying situations in which students made four errors in a space of six words (33% accuracy or less) demonstrates unusually high local error rates and suggests situations where comprehension and sense making have broken down. As it turned out, this standard fits closely with descriptive data that were collected regarding student behavior in these situations. Treating the likelihood of making four errors in the space of six words as a binomial probability, there is a less than .001 chance of having four errors in six words (the chance of three errors in six words is less than .01) using the average accuracy rate of 93.2% for all students in the sample. Even using the error rate for the bottom tercile students (14.1%), the chance of having four errors occur in six words is significant ($p < .01$). The instances of these clusters of errors were cross-checked to make sure that leaving an 's' off the end of a word or saying 'a' for 'an' weren't included as errors.

Across all 24 books, no student in the top tercile ever crossed the 67% error threshold (i.e. made four errors in six running words) on any book in this study. The fifteen students in the mid tercile produced clusters of errors occasionally, averaging just under two/book across the group (i.e. an average of two out of fifteen students would produce a cluster of errors each book). For students in the bottom tercile, however, the fifteen students averaged over seven blocks of errors in each book. (i.e. an average of seven of the fifteen bottom tercile students would produce a cluster of errors each book). Rarely were two blocks of errors produced by a single student.

Text Features Related to Student Reading Performance

In addition to specific accuracy and fluency information, it is possible to analyze student error patters in relation to specific text features. According to the California state standards (California Board of Education, 2006), at least 75% of the words in each story are to be decodable using phonics information that has been previously taught in the classroom. Of the remaining 25% of words, 15% - 20% of those are to be previously taught sight words. The remaining 5% - 10% of words are classified as "story words", typically words necessary for the story in some way, yet containing spelling patterns that had not been previously taught. Many of the story words were high frequency words that either were never taught or hadn't been introduced yet (e.g. says, should, friend). Others were inflected forms of common words where the inflection caused an irregular pronunciation or the inflection hadn't been taught yet (e.g. feeds, loved, heard). Table 6 shows the breakdown of student errors by word category (story words, sight words, decodable words) with an additional sub-category of decodable proper names.

Table 6

Errors by Word Type Across All Books for All Students, Summed by Terciles

	Absolute #				Percentages			
Word type	Total	Top	Mid	Bottom	Total	Top	Mid	Bottom
Story words	102	5	28	69	3.3%	3.3%	3.4%	3.0%
Sight words	423	20	85	318	12.9%	13.1%	10.3%	13.9%
Decodable words	2742	128	712	1902	83.9%	83.7%	86.3%	83.1%
(Decodable proper names)	363	5	97	261	11.1%	3.3%	11.8%	11.4%
	3267	153	825	2289				

Notes: Table shows the absolute number of errors made by each tercile by word type. (e.g. Students in the bottom tercile made a total of 1641 errors on decodable words across the 24 books)

A different way of looking at the same data is to calculate the average error rate, expressed as a percentage, for each type of word. Table 7 shows the three main classes of words (with information for decodable proper names as well) and their respective error rates. Error rates give a basic description of the likelihood of a certain class of word being read incorrectly when encountered by a student.

Table 7

Error Rates by Word Type

Word type	Avg.	Top	Mid	Bottom
Story words	14.3%	2.1%	11.8%	29.0%
Sight words	2.6%	0.4%	1.6%	6.0%
20 most frequent words	1.4%	0.4%	0.9%	2.9%
Decodable words	8.3%	1.2%	6.5%	17.4%
Decodable proper names	7.8%	0.3%	6.3%	16.9%

Notes: Error rates are calculated by dividing the total number of errors for a given word type by the total number of words of that type for all students/books

Most commonly missed words. Using a simple binomial probability test and an average error rate of 6.8% (based on the average accuracy for all students across all books), individual words in a specific story that were missed by at least eight students (of the possible 45 readers) were flagged as significant at the .01 level. A variety of alternative measures were used to attempt to correct for possible skewing in this analysis, but in the end, the simple measure proved sufficiently reliable given the heightened standard for significance testing. Of the 1181 words read, 94 (8.0%) were missed by at least eight students. The 20 most frequently missed words were read incorrectly in one instance by at least 18 of the 45 students (most of the 20 most commonly missed words appeared only once across all 24 texts). See appendix 2 for a list of the most frequently missed words. Of the 20 most commonly missed words, one was a story word, none were sight words and 19 were decodable words.

Table 8

Most commonly missed words

Frequency of commonly missed words	Story	Sight	Decodable	Names
words with 18 or more errors (top 20)	1	0	19	(1)
words with 10 or more errors (top 64)	3	3	58	(3)
words with 8 or more errors (top 94)	5	5	84	(10)

Using the 20 most commonly missed words as examples, there are a number of distinguishing features of these words. While the average number of letters per word for the entire corpus is 3.7, the 20 most commonly missed words averaged 5.8 letters/word. The average number of syllables across the entire text was 1.2 syllables/word, for the 20 most commonly missed words it was 1.5 syllables/word. Not a single word of this group was repeated in any of the 24 texts analyzed here. Additionally, the 20 most commonly missed words appear on average less than 30 times in a million (Brysbaert & New, 2009) with no word appearing more than 300 times in a million.

Looking out to the larger group of the 94 most commonly missed words, the average number of letters/word is 5.5 and the average number of syllables is 1.4. In this larger group, a number of the words are repeated, especially the proper names. However, not all instances of the same word were mis-read with equal frequency. The word *asked*, for example, occurred eight times across the 24 books in this sample, yet of the 12 total errors made reading this word (by all students), nine of them occurred in relation to one specific passage. The word *huge*, however, occurred only three times, all in one story, and all three instances of the word were missed by at least eight students. 14 of the 94 most commonly missed words would be classified as easily decodable (Juel & Roper-Schnieder, 1985) and consisted of single syllable words made up of simple letter sounds (e.g. *maps, yum, band, crops*). Only one of the top 20 most commonly missed words fit these criteria (*pups*).

Problematic Passages. Combining the analysis of the most commonly missed words with the earlier analysis of blocks of errors brings up an additional finding—situations where multiple students produced blocks of errors at the same point in the same book. Given the extremely low probability of any given student producing a block of errors randomly (four mistakes in the space of six words, $p < .001$), the likelihood of multiple students producing a block of errors at the same point in the same book is statistically improbable ($p < .0001$) yet descriptively significant. The error clusters were not distributed evenly amongst the texts. In some books, not a single student made four errors in six words, in other books, up to seven students made four errors in six words on the same passage. Three different books had at least ten instances of students making at least four errors in six words. The heavy majority of these error clusters were produced by students in the bottom tercile (82.7%). Not a single student in the top tercile ever produced an error cluster and less than a fifth (17.3%) were produced by students in the middle tercile. The remaining 82.7% of error clusters were produced by students in the bottom tercile.

An additional component of the error cluster finding is a qualitative description of the nature of the errors produced by students. While the correlation is not exact, many error clusters were accompanied by one of three patterns of reading behavior produced by the students making the errors. In some instances students stopped trying to read words they didn't know and requested to be told words they didn't recognize immediately. Once they encountered a block of

text they could read independently or even individual words they recognized, they would start reading independently again, but during these blocks of errors, they would refuse to attempt to decode unfamiliar words. Another common finding was that students would attempt to decode the words of the text without monitoring to see if what they read made sense. This behavior of "reading gibberish" occurred repeatedly in situations where students had made such a significant number of errors in a small space that it appeared their grasp of the meaning and context of the story had been lost. They made clear attempts to decode text, but regardless of what they produced, whether it was an intelligible sequence of words or if the specific word they attempted to read was even a real word, they continued decoding. The last pattern that occurred repeatedly during blocks of errors was that students would stop attempting to decode the text and simply say out loud what they thought was happening in the story. It was clear from a comparative analysis of the actual text and what was "read" in these instances that there was little or no connection between the actual orthographic information on the page and what the student read out loud.

Discussion

Before discussing the findings from this study, it is important to understand that the results described here are very likely "best case scenario" results, and in that, they likely far overstate student performance using lesson-to-text-matching decodable texts. The first reason for characterizing these data as "best case" stems from the specifics of Lockwood elementary school. The school is in the top 80% of California elementary schools in terms of performance on standardized tests and was recently designated a California distinguished school. The neighborhood in which the school is situated is affluent, with average home prices well above $400,000. Only 20% of the students receive free and reduced lunch, the majority of the parents are working professionals and the majority of students participate in some type of afterschool academic enrichment program. The two first grade teachers have a combined 19 years of teaching experience and both first grade classrooms have regular, daily volunteers as well as paid academic support specialists for struggling students. The students' scores on the widely used DIBELS assessment do not indicate any unusual patterns and are representative of the higher end of each of the three terciles. All of these factors suggest a well-functioning, highly supportive academic setting that consistently contributes to the success of the students.

A second reason for the "best case" characterization stems from the instructional environment in which the books were presented. The data gathered for this analysis came from the third or fourth reading of the LTTM decodable texts that had been previously read twice in class. Students were given a book introduction, they were pre-taught target phonics content and specific sight words, and they were shown and taught any story words or other words that might be problematic in the text. The texts were read once as a class together and then read again independently in class with struggling readers getting individual attention from the teacher or classroom volunteers. The students were then given a copy of the book to take home and read again. Adams (2009) noted that the first read of a decodable text is often difficult, time consuming and laborious. Subsequent readings, she argued, would become easier, more fluent and more accurate. The SRA Open Court Administrator's Guide (Hoit, 2008) makes the same point, stating that rereading texts supports the development of fluency. Any students who were absent during the initial introduction and reading of a given book were not assessed for that specific book. Data from Nation et al (2007) support the view that student scores on a variety of word learning measures are highest when measured one day from initial instruction and fall off significantly over time. Thus the results presented here regarding student reading accuracy and

fluency reflect optimal conditions in regard to student experience with the LTTM decodable texts in question.

The final reason why these results are likely best case scenario has to do with the methodological decision to follow the DIBELS protocol during student reading—telling students a word they were unable to read after three seconds (the DIBELS assessment was used across the district as a standardized measure of student literacy achievement). Share (1999) discusses this specific methodological decision and explains his decision for not telling students words they are unable to read by themselves. He argues that telling students unknown words artificially elevates their reading outcomes by providing them with additional contextual information that they would not otherwise have. It is not, in Share's view, equivalent to the situation they would face when reading texts independently. The need or request to be told a word varied dramatically across terciles. While students in the top tercile almost never needed to be told a word, students in the middle tercile were told an average of 1.2 words/book and students in the bottom tercile were told an average of 3.8 words/book. For the readers in the bottom tercile especially, each of these tells gave extra contextual and syntactic support to students who were struggling to make sense of the text. It is highly likely that at least in some cases, this extra information enabled students to continue reading texts they might otherwise have been unable or unwilling to finish independently. This methodological decision remains one of the most controversial issues in early reading assessment—and pedagogy for that matter.

Impact of LTTM Decodable Texts on Reading Accuracy

According to SRA McGraw Hill's support materials for Open Court Reading (Roit, 2008), "reading decodable text gives students the opportunity to develop two of the three elements of fluency—accuracy and automaticity" (23). The data analyzed here, however, suggest important problems for participants in this study. Even with the optimal conditions described above, the results indicate that students in the bottom and middle terciles struggle to read the daily, instructional lesson-to-text-matching decodable texts accurately and fluently. Students in the top tercile read the books without difficulty. 84.3% of the readings from top tercile students were at Betts (1946) independent level (> 97% accuracy). Middle tercile students were reasonably successful with the books, reading 50.2% of the books at the independent level and averaging 94.9% accuracy across all 24 books. Students in the bottom tercile, however, averaged 85.9% accuracy across all 24 texts. 76.4% of the readings for students in the bottom tercile were below Betts (1946) frustration cutoff (< 94% accuracy), whereas 30.6% of the readings for middle tercile students and only 6.1% of the readings for top tercile students were at the frustration level. 76.4% of the times bottom tercile students read a decodable text, it was below 94% accuracy and 31.0% of the time it was below 90% accuracy. So it is the low achieving students who appear not to be getting the sort of practice that leads to the development of automaticity and fluency.

Analysis of the student reading indicates that students primarily made mistakes on decodable words, including a large percentage on decodable proper names. While sight words accounted for a significant number of errors, they were misread much less frequently as a class than were decodable words. Since sight words made up only 20-25% of the words in a given text (as compared to 75% for decodable words), they contributed less to overall accuracy scores. The clear take home from these results is that while top tercile students read the decodable texts easily and the middle tercile students were moderately successful with them, the bottom tercile students, the ones most dependent on classroom instruction to support their reading development,

had the greatest difficulty reading the LTTM decodable texts accurately. If the stated purpose for decodable texts is to build towards reading fluency by supporting accuracy and automatic word recognition, the texts are not supporting struggling readers in achieving those goals.

On the one hand, it is critical for the research community to consider this finding and address the implications. Data from Hiebert et al (2010) and Compton et al (2004) suggest that word decodability does not significantly predict reading accuracy. Compton et al (2004) and Mesmer (2010) found that text decodability does not lead to reading fluency. The results presented here add to the chorus of concerns over the role of LTTM decodable texts and their supposed benefits for beginning readers. The analysis of student reading data provides a detailed picture of the research debate over the efficacy of LTTM decodability mandates being played out in daily instruction. It is a critical theoretical assumption of the LTTM design framework that a high percentage of words in decodable texts meet the lesson-to-text-matching criteria, because this supposedly ensures students will be able to read the texts accurately and that repeated successful readings will lead to automaticity in decoding, automatic word recognition and fluency. Data from a number of studies have provided evidence arguing against this assumption and evidence in this study, from students using actual texts designed according to LTTM mandates, demonstrates the shortcomings of lesson-to-text-matching decodable texts.

Why Don't LTTM Decodable Texts Support Accuracy?

In order to advance the debate over the questions of the efficacy of LTTM decodable texts, it is important to look at the question of why lesson-to-text-matching decodability is not having the desired effect in supporting student reading accuracy and fluency for struggling readers. After all, the argument for using texts primarily consisting of previously taught phonics patterns and sight words is, on face value, transparent. If one were comparing texts with words difficult to decode or words that contained phonics content students had never seen before and compared them to LTTM decodable texts, clearly LTTM texts would be read more accurately and fluently. It's equally true that LTTM decodable texts present few if any problems for high-performing readers; they read the texts accurately and fluently. But the texts are designed to be used by all students, regardless of reading ability level. Indeed, some advocates argue that struggling readers are the ones most in need of the unique design feature built into LTTM decodable texts. Every day in classrooms across the country, students of different reading levels read the same LTTM decodable texts as part of their research-based curriculum. So the critical question for researchers and text designers is why don't the scaffolds built into LTTM decodable texts work as effectively for students in the bottom tercile, as they do for students whose average DIBELS composite score represents the 40^{th} to 50^{th} percentile of student nationally? While this study is clearly not an experimental analysis of that question, a post-hoc analysis of student reading data suggests four important factors that undermine the degree to which LTTM decodability standards support struggling readers.

Lesson-to-text-matching (LTTM) and the transmission model. First and most importantly, the lesson-to-text-matching theory of decodability is based on the binary question, "Has the relevant phonics material been taught previously?" Hiebert et al, 2010a, Mesmer (2010) and Allington (2005) have all pointed out that this is the ultimate example of the transmission model of instruction, where a teacher tells students what they want them to know and assumes that students are therefore able to use that information successfully—in other words, instructed means learned. As a way of testing this model, consider a thought experiment. What if the most difficult and variable phonics elements of the English language were taught first (e.g. the –ough

ending; *cough, rough, through, though*, etc.). Would it still be assumed that students would be able to successfully decode words that incorporated the –ough pattern since it had been previously instructed? What if students were taught the 100 most common phonics patterns in English in the first week, would that imply that students could read any text incorporating those phonics patterns regardless of text complexity, sentence structure and vocabulary in week two? No one is arguing that 1st grade curricula is designed in such an irresponsible manner and the material included in 1st grade reading instruction as well as the pace of instruction has been assessed repeatedly at a national level. But the state mandates for text design in California and Texas require that LTTM decodable texts be built from what has been previously taught in class and make no mention of the fact that not all students learn from classroom instruction at the same pace. While it is an age old dream of the education establishment that instruction would directly translate into learning, and while top and middle tercile students seem able to function according to this pedagogical model, for bottom tercile students, the evidence from this study demonstrates that instructed does not mean learned.

But what if the issue of instruction were removed from the equation and we only looked at students who had demonstrated mastery of the relevant phonics content—how accurately would these students be able to read a decodable text? Data from Mesmer (2010) demonstrate the struggles with accuracy even such incredibly well-prepared students have with decodable texts. Mesmer assessed participants in her study ahead of time for mastery of the specific phonics content included in the decodable texts she had them read. She describes her standard for matching participants with decodable texts as, "based on verified letter/sound knowledge as opposed to previously instructed letter/sounds" (p. 29). Yet participants in her study averaged 84% accuracy during their initial reading of the text and by the fourth reading, had achieved only 93% accuracy. For the second set of participants, they ranged from an accuracy score of 88% accuracy on their initial reading to 95% accurate by the fourth reading. Mesmer's data demonstrate that even when texts are matched specifically to students' specific phonics knowledge, they still produce error rates of a magnitude that can undermine their reading development.

Non-linear increase in decoding difficulty. Second, while seminal research looking at early literacy development describes the special role of simple, easy to decode, CVC words (i.e. words made up of a consonant – vowel – consonant, like cat or mop) in supporting beginning readers (Ehri, 1998; Adams, 1990), the difficulty of the task faced by 1st grade students increases drastically across the school year. For readers of Finnish, Spanish, Hebrew and other languages with a shallow orthography, the task of learning to read is much less daunting and far fewer students struggle to acquire basic literacy (Share, 2004). For students learning English, however, the task is much more complicated. Many students move through an initial phase in reading acquisition where they successfully apply simple letter sound relationships in solving words, as long as the words are carefully selected and follow regular spelling patterns (Share, 1995; Ehri, 2005). A few key sight words are acquired quickly and used successfully. Much of this instruction has been moved to kindergarten and it is assumed that kindergarten students have learned all their basic alphabetic sounds and a corpus of 20+ high frequency sight words. By the end of kindergarten, students are expected to consistently apply a previously learned set of regular spelling patterns to decode a carefully selected set of words or to recognize a handful of the most frequently used words in the English language (Hiebert, 2005). It is entirely reasonable that even a struggling 1st grade reader could start first grade reading a text like, "I see a dog and a cat", with 100% accuracy.

By the middle of 1st grade, however, the analytical task facing students is much more challenging. Over the course of 1st grade in Open Court Reading (Roit, 2008), students are introduced to over 100 new sight words and over 100 spelling patterns representing the 44 unique sounds that make up the English language. These spelling patterns are combined to form a wide variety of words that students encounter across the 119 texts that make up the decodable set. As decoding demands increase, struggling readers who have not mastered all of the new phonics information required to read the LTTM decodable texts read more slowly, have a weaker sense of story context and have limited free cognitive resources to "resolve decoding ambiguities" as Share (2005) describes, when they encounter an unfamiliar word and their initial attempts to decode it do not reveal an immediately likely option. The net result, as exemplified in this dataset, is that struggling readers make more errors and these errors undermine reading development. While any set of texts for beginning readers must increase in difficulty level over time, the rate at which LTTM decodable texts increase in difficulty outpaces the rate at which struggling readers develop the ability to accurately read the more difficult texts (Hiebert et al, 2010a).

Repeated inclusion of low-frequency words. A third factor affecting the ease with which students can accurately decode LTTM decodable texts has to do with the repeated inclusion of low-frequency words. Hiebert et al, (2010b) found that while students at benchmark were unaffected by word frequency on the end of year DIBELS oral reading fluency assessment, students in the strategic group (mildly underperforming) were significantly impacted by a word's frequency and students in the intensive group (significantly underperforming) were even more affected. They concluded, "Students at risk are significantly more sensitive to word frequency than the other groups" (p. 21). Pearson and colleagues in a conference presentation described data showing that simple, CVC low frequency words were recognized less quickly than CVC high frequency words (Pearson, Duong & Callahan, 2005). Share (2008) argued for the critical role of context in supporting readers to correctly recognize and identify accurate phonetic recodings (i.e. the mechanical component of decoding). Yet what happens when a struggling reader encounters a sentence like "Kim tramped on grass and trudged up hills"? Even if they were to successfully decode *trudged*, if they are unfamiliar with the word, how will they know if they have decoded it accurately?

Across the 24 LTTM decodable texts used in this study, low frequency, LTTM decodable words like *trudged*, *woe* and *glee*, appear a single time. These words were presented without any supportive context to help verify correct decoding or to determine their meaning. Surprisingly, the inclusion of low frequency words was not limited to instances required to practice target phonics content. Of the 50 words that students missed most frequently, 46 were LF words and only 12 of them were related to specific phonics content being practiced in that specific story. The other 34 most commonly missed, low frequency words were included as examples of previously taught phonics content according to LTTM mandates. One story in the sample, *Kim's Trip*, that described a woman walking in the park, included the words *trudged* and *tramped.* These words did not reflect target spelling patterns recently taught nor were they essential words for telling the story. The –dge spelling pattern had been taught previously in class and the word tramped is comprised of simple letter sound mappings (the final –ed sound had been taught previously as well). Even though they were exceptionally low frequency words most likely outside the oral vocabulary of typical 1st grade students, they were included, or at least were allowed to be included, because the met the decodability standards.

The clearest conclusion from analyzing the average word frequency of decodable words is that LTTM decodable texts do not attempt to control for word frequency. According to LTTM theory, low-frequency decodable words are fair game because the scaffold of LTTM gives students sufficient support to read words made up of previously taught spelling patterns, even if the words are rarely used in the English language and are thus much more likely to be outside the oral vocabulary of struggling 1st Grade readers. And while it might be nice to argue that this practice is an egalitarian form of raising the bar so to speak, increasing the challenge level of materials for all students, it essentially creates another gap for low performing students. There are no supports built into the texts or the teacher's manuals for how to teach new vocabulary (McKeown & Beck, 2004). There are no state mandates or requirements that 1st grade decodable texts teach vocabulary nor does SRA McGraw Hill ever clarify that a goal of their decodable texts is to introduce students to a wide range of new vocabulary words they have never heard before.

Repetition. In the text, *Rescue the Cat*, the word 'mew' appears 10 times. Needless to say, the word never appears again in any of the other 118 decodable texts used in first grade. Of the 50 most frequently repeated words across all 24 books (repeated at least five times), 12 of them are words unique to a single story (e.g. *mew, Ralph, Kim, blur*, etc.) and are never encountered again. Of the 423 unique words that make up the first 50 words of the 24 books in this study, 238 of them (56.3%) appear only one time; 66 of the words (15.6%) appear just two times. Adams (2009) argues that successful repeated decoding is the "prepotent determinant" in the development of automatic word recognition. Yet over 70% of the unique words students encountered in these books were repeated at most twice. Of the remaining roughly 30% of words, 20% - 25% of them are sight words, leaving only 5% - 10% non-sight words that are repeated at least three times across 24 books. LTTM decodable texts clearly offer repetition of target phonetic patterns, and supporting students to develop better decoding skills could lead to quicker, more successful decoding which should lead to automatic word recognition. But developing automatic word recognition is contingent upon students getting to successfully decode target words multiple times across texts. While it is clear that frequent opportunities to decode unfamiliar words provides a wealth of opportunities to practice decoding, the texts in this study provide little opportunity for students to develop accomplish Adam's (2009) ultimate goal of repeated successful decodings of new words in order to develop automatic word recognition.

Related to the issue of repetition, an unintended consequence of the LTTM mandate requiring 20% - 25% of the words in a given text to be sight words affects the language patterns that make up decodable texts. In the majority of texts used in public school, roughly 50% of words in a given text come from the 300 most frequently used words in the English language. By mandating that 75% of the words in a story be decodable, text designers are forced to use language patterns that are significantly different from other school texts. Amsterdam, Ammon, & Simons (1990) showed that students had difficulty orally repeating the unusual language patterns from the linguistic readers of an earlier era, which were constructed using principles that are similar to those that underlie today's LTTM decodable texts. Amsterdam et al concluded that students having difficulty repeating unusual language patterns orally would be much less likely to read similar patterns successfully when they encountered them in text. By requiring 75% of the words in a text to be decodable, LTTM mandates force text designers to adopt unusual language patterns not found in other texts. While top tercile students have the skills and literacy experience to flexibly navigate these unusual texts, struggling beginning readers are the least well-equipped to handle the unusual language patterns created by the LTTM mandates. Texts

like these frequently remove from the student's arsenal of resources the ability to compare what they read to the speech patterns they hear on a regular basis. This makes monitoring for sense (Does this make sense? Does this sound like something I might say or hear?) very difficult.

These four factors significantly and disproportionately affect the reading performance and learning gains of struggling beginning readers. In 84% of instances in which a student in the top tercile read a decodable text in this study, they read it at 97% accuracy or higher Betts' (1946) standard for independent reading). In 50% of the instances in which a middle tercile student read a decodable text, they read it at 97% accuracy or higher. For students in the bottom tercile, only 8% of their readings were at 97% accuracy or higher—Betts' (1946) cutoff for independent reading. More importantly for the bottom tercile students, 76% of their readings of LTTM decodable texts were at or below the frustration level (< 94% accuracy). Lesson-to-text-matching is a design criteria for matching decodability with previous instruction, but it ignores the wealth and history of research on text difficulty (Hoffman, Sailors, & Patterson, 2002). All 119 texts of the Open Court series of decodables are certified LTTM decodable. But that says very little about how difficult they are and the degree to which struggling beginning readers will be able to read them accurately and fluently.

Impact of LTTM Decodable Texts on Fluency

All study participants made significant gains in fluency over the course of the school year as measured by the DIBELS mid-year and end of year Oral Reading Fluency assessment with struggling readers making the greatest gains. Because each book was read only once for this data set, there are no data available for assessing within book improvement over time. Since the data collected are non-experimental, there is no way to know for sure the exact cause of students increase in reading fluency over the course of the year. Because the LTTM decodable texts used were part of a comprehensive curriculum, there is no way to partial out the effect of decodable texts over and against the effect of instruction in general. Even with these limitations, however, a number of theory based observations that can be made related to the fluency development of students in the study.

For students in the top tercile, there is little evidence that the books provide much if any benefit. Top tercile students started the year reading the decodable texts with 99% accuracy and ended the year the same. Eleven of the fifteen top tercile students read 90 wpm or higher on the mid-year DIBELS oral reading fluency assessment, a score that is the cutoff for 2nd grade end of the year reading success. Ten of the fifteen top tercile students were only required to read the decodables a single time and then were allowed to read a wide range of classroom texts more appropriate to their skill level and interests for the remainder of that time block. It is worth noting that while it was always permissible for top tercile students to reread or select any of the decodable texts they had read previously, not once did a single student ever choose to do so. From every indication, the top tercile students read beyond the level of the decodable texts from the beginning of the year, they spent minimal time with the decodable texts and there is little reason to believe that those single exposures played a significant role in the reading development of the top tercile students. Comparing the mid-year and end of year DIBELS Oral Reading Fluency assessments, the top tercile students who spent the least amount of time with the decodable texts were the only group to rise in percentile ranking (from 89th to 93rd) in terms of their reading fluency.

Students in the middle tercile made significant gains on fluency over the year that can clearly be attributed to the overall instructional program, thus with some portion of the credit

going to practice with the decodable texts. The LTTM texts allowed middle tercile students to practice the range of phonics elements they were taught in class as well as practice reading the sight words they had been previously taught. By the end of the 119 books included in the curriculum, all of the 15 middle tercile students had made significant and large gains across all DIBELS measures. Middle tercile students averaged 55 WPM on the mid-year Oral Reading Fluency assessment and averaged 72 WPM on the end of year assessment ($t = 5.6$, $p < .001$).

If, as I have argued, these books are so problematic for struggling readers, why did their DIBELS Oral Reading Fluency scores almost double from mid-year to end of year (from 22.9 wpm to 39.9 wpm)? First, beginning readers get better at reading with practice. The instructional program in Open Court Reading provides students with intensive systematic phonics instruction coupled with opportunities to practice decoding using texts that match instruction. Even without a perfectly designed experiment controlling for all possibilities, it is clear from results across the nation that many students improve in their reading ability using Open Court Reading and other programs based on similar instructional practices. A critical issue addressed in this paper is not whether significant gains were made, but whether or not there is evidence that gains are being constrained by the nature of the instructional texts being used in 1st Grade classrooms.

Apart from the earlier accuracy and fluency data that describe the struggles of students in the bottom tercile, participants' DIBELS data suggest a developmental trajectory that is potentially problematic for both middle and bottom tercile readers. While both groups increased significantly in their reading fluency, both groups dropped, compared to national norms, in the percentile rank that their respective fluency scores represent. Middle tercile students scores on the DIBELS Oral Reading Fluency assessment dropped in percentile rank from 78th percentile to the 66th percentile (see Table 2). For students in the bottom tercile, their DIBELS Oral Reading Fluency scores dropped in percentile rank from the 44th percentile to the 35th. Top tercile students' percentile rank rose from 89th percentile to 93rd. Students in the middle and bottom tercile, the students who spent the most time with the LTTM decodable texts and who had the most difficult time with the decodable texts made smaller gains in reading fluency (which is a composite score of fluency and accuracy) than did their peers across the nation. Top tercile students who used the LTTM decodable texts the least were the only group to gain in the percentile rank their end-of-year fluency scores represented.

In Mesmer's (2010) study comparing reading performance of students using LTTM decodable texts and qualitatively leveled texts, she found that decodable texts consistently produced smaller gains in the development of reading fluency over time than qualitatively leveled texts. In describing why students had lower fluency rates even at the end of the year (when the decodable texts should have been easier to read) Mesmer wrote, "There was something about text construction in this study that influenced fluency even in relatively easy books" (p. 35). She concluded that the lower percentage of high frequency words and the inclusion of low-frequency words both negatively impacted reading fluency in the decodable texts.

Summarizing the Effect of LTTM Decodable Texts on Reading Accuracy and Fluency

The foundational principles undergirding the widespread use and mandates of LTTM decodable texts are (a) that texts designed to match previous instruction support reading accuracy and (b) that repeated accurate decoding leads to automatic word recognition which leads to the development of fluency and established readers. This study demonstrates that for virtually all students in the bottom of the achievement distribution and for many students in the middle, they

are not repeatedly successful in decoding words in texts they've read twice before. Data from a wide variety of sources demonstrate that accuracy correlates strongly with word learning and that errors undermine word learning (Share, 2005; Cunningham, 2006; Nation et al, 2007). Regardless of the stated intentions of curriculum designers and educational researchers, LTTM decodable texts produce significantly different outcomes based on student reading ability level and as Juel & Roper-Schnieder (1985) argue, one's experiences during reading practice powerfully impact reading development—for better or for worse.

From the current dataset it is clear that top tercile students read the LTTM decodable texts at the independent level from the beginning of the year. Most top tercile students read the decodable texts only once and there is little theoretical or empirical justification for assuming that the decodable texts played a significant role in their gains over the course of the year. Middle tercile students made significant gains and clearly benefitted from the texts included in the program. Middle tercile accuracy scores still show error rates that indicate potential problems with the theory of how automatic word recognition develops and the drop in percentile rank on the DIBELS Oral Reading Fluency assessment from mid-year to the end of the year suggests that reading gains for middle tercile students are not keeping pace with national averages.

For students in the bottom tercile, the picture is much more problematic. Bottom tercile students are clearly developing as readers and increasing in their ability to decode and read text accurately and fluently. Yet the primary resource for their ongoing practice of reading connected text, the LTTM decodable texts analyzed in this study, cause a disproportionate degree of difficulty for the lowest performing readers. Accuracy scores for bottom tercile students reading LTTM decodable texts are consistently at levels where previous studies have demonstrated reduced word learning outcomes (Cunningham, 2006; Nation et al, 2007). Low fluency scores indicate an ongoing weakness in their ability to decode automatically and to recognize words automatically. The decline in percentile rank for fluency scores of students in the bottom tercile indicates that they are developing fluency more slowly than their peers across the nation. While all of the struggling students in this study made clear and substantial progress, the data suggest the LTTM decodable texts are not working according to plan. Students in the bottom tercile are the most dependent on the public school system for learning to read. Yet there is every indication based on the data presented here and on the analysis of research over the last 30 years that specific design attributes of LTTM decodable texts present challenges to the reading development of struggling students at the very least, and may play a major role in constraining and limiting the progress they might otherwise make in 1st grade.

Text Features and Student Reading Performance

The first and foremost goal of this paper is to highlight the critical finding that low-performing readers struggle with reading LTTM decodable texts accurately and that this finding is problematic given the primary purpose of decodable texts as providing opportunities for repeated successful decoding leading to automatic word recognition. With any data showing clear differences between student achievements across any measure, there can be a tendency to focus on the students as the source of the differences in outcomes. The students' DIBELS scores provide clear evidence that students vary in their knowledge of phonemic awareness and orthographic patterns and in their recognition of sight words. But there is also clear evidence from an analysis of the reading behavior of students reading the LTTM decodable texts in question that specific text features strongly impact student performance. Lipson & Wixson (1986) described an interactive model of assessing text difficulty, arguing that instead of

measuring student ability by their performance on a given text, a text's appropriateness should be measured based on students' success in reading a given text. Given that the wide range of research cited earlier discussing word learning concludes that decoding accuracy strongly correlates with word learning, an interactive model assessing text difficulty and text suitability seems highly appropriate. Using this interactive framework to analyze student reading on the 24 books in this study offers a different perspective that clarifies and elaborates the main findings of why students struggle to read LTTM decodable texts accurately.

In order to understand the role the LTTM decodable texts play in determining student reading outcomes, each book was analyzed by placing students on the rows of a spread sheet and the first 50 words of a story in each column of the same spreadsheet (see Appendix 1). To get a rough "picture" of what was happening during student reading, an "X" was entered for each student under each word they missed; cells representing words read correctly were left blank. Once all the data had been entered for all the books, the results were visually emphatic. While there was clearly a certain amount of randomness to the errors and there were discernible horizontal lines representing high error rates for the lowest performing students, what was clearly apparent were the vertical patterns of errors at specific locations in each book. In some books it would be two or three locations with 50% or more students making errors. In other books, over 80% of the students made errors at a specific point. Across almost all of the 24 books, this pattern of specific error locations was replicated. There were clearly specific words in different texts that were read inaccurately in a substantial number of students.

The features of these most commonly missed words were analyzed earlier—longer average word length, low-frequency, they appear only once or twice at most across the 24 texts and the words cannot be identified through context. Hiebert (2010a), analyzing results from 14,000 samples of students responses on the DIBELS Oral Reading Fluency assessment found that word length and word frequency significantly predicted the likelihood of errors for struggling readers, but had no significant effect for readers at benchmark. Her data showed that words identified correctly by the majority of struggling readers were almost entirely from the 100 most common words in the English language. The findings from the data collected for this dissertation match Hiebert et al's (2010a) findings closely and help clarify the simple but divisive idea that the ease of correctly identifying a word is more complicated than assuring that the relevant phonics information has been previously taught.

***Kim's Trip*.** A specific example sheds light on the effect of texts on student reading performance. In the book, *Kim's Trip* (Book 53 out of 119, introduced in early December), the text on the first page of the story reads, "For her trip, Kim packed park maps, snacks and a jacket. Kim tramped on grass and trudged up hills." This example is not especially egregious in terms of odd language structures or unusual vocabulary choices and is representative of the types of language and vocabulary choices made across the twenty four texts analyzed in this study. Twenty two of the twenty four stories contained at least one long, low frequency word, likely not in the oral vocabulary of struggling readers, that most likely could not be identified from context. This passage is designed according to LTTM decodability standards and provides students an opportunity to practice decoding using previously taught phonics content. Yet in this example, 25 of 45 students missed the word *tramped* ($p < .0001$) and 33 of 45 missed the word *trudged* ($p < .0001$). Those two words were not required as examples of specific phonics content unique to that text—the text *Kim's Trip* taught the letter *k* sound. These words were not essential to the story, and the picture on the first page does not indicate trudging or tramping (the woman in the story is walking on flat ground, looking at a map). There was nothing in the teacher's guide that

indicated these words were to be pre-taught or that these words might be unfamiliar to many students. The words are simply decodable according to LTTM standards. But beyond those two hugely problematic words, fifteen students missed the word *packed*, eleven missed *park*, nine missed *maps*. There was only one error on each of the first four words ("For the trip, Kim") so it wasn't an issue of students not being able to get started, and all of the easy connector words between the more challenging words ("and", "a", "on") were read with at most one error. In all, 122 errors were made across all 45 students in the first 20 words—an accuracy rate of 86%. 69 of those 122 errors were made by students in the bottom tercile for an accuracy rate of 75.4%. There were no story words in this section and only two of the 122 errors occurred on sight words.

The results across all 24 books show example after example of the theory of LTTM decodable texts bumping into the practical reality of early literacy development. As was argued earlier, all decodable words are not equal. Imagine for a moment, a reader who had never heard the word *trudged*—what should they do in a situation like this? If they correctly decode the word and produce "trudged", they have no idea what it means. Are they supposed to read on? Should they try, as Share (2005) suggests, to test their phonetic recoding against alternate hypotheses to see if there's a more likely option? Are they able to infer the meaning of trudged from the context? Should they assume it's a sight word they can't remember?

Eight of fifteen readers in the top tercile missed the word *trudged* in that story. For almost all of them, it was the only word they missed (two others missed tramped and three others missed the name of the Park, *Birch Park*). Top-tercile readers went on, finished the story and were only minimally impacted by having missed the word *trudged*. When asked at the end of the story what trudged meant, the majority of top tercile students said something related to walking. 13/15 middle tercile students missed the word *trudged*, 10/15 missed the word *tramped*. At the end of the book, roughly half said the word trudged had something to do with walking while the other half either didn't know or gave an unrelated answer. 12/14 of the bottom tercile students (one was absent when the book was presented in class) missed the word *trudged* and 13/14 missed the word *tramped*. Only two of the 14 bottom tercile students gave a definition for trudged related to walking.

Are there reasonable explanations that could clarify the role these words played in the text or explain why students had such difficulty with them? Even though the words were not called out in the teacher's guide, both classroom teachers identified and "pre-taught" the words *trudged* and *tramped*. The words were introduced, the teachers acted out the meaning of the words; in one class, a student came up front and demonstrated trudging and tramping. The students involved are above average for each tercile they represent, with above average scores on phonemic segmentation and nonsense word fluency, so there is no evidence that the students are the source of the difficulty. If the explanation doesn't lie with the students, could it be that the words in question are important pedagogically? Is the –dge cluster important in English and useful for early reading development? Are these words used again in other stories or are they important words in other early literature? Is this a new strategy for vocabulary development, including target words in decodable texts? Is this a form of curriculum embedded assessment to measure student decoding levels and to offer especially challenging words to test high level readers and it's an intended feature that struggling readers aren't supposed to be able to read them? There is no evidence for any of these alternate explanations, seemingly because there is no need—the words are included because they are decodable per LTTM standards.

Intentionally unsupportive texts. There is one potential advantage for using the words *tramped* and *trudged*—the words can't be solved through context and must be decoded. A number of researchers have argued that contextual support (e.g. predictable language or helpful pictures) undermine orthographic learning (Ehri & Roberts, 1979; Nation et al, 2007; Landi et al, 2006). If this is a guiding principle behind word choice in LTTM decodable texts, it could explain why low-frequency words are so frequently selected (Foorman et al, 2004; Hiebert et al, 2010a). If the words in a story make too much sense or are too easily predicted based on context, text designers might worry that students would be distracted from paying attention to the orthographic details of the words in question. In an earlier study (Frey, 2009), examples of decodable passages were identified that seemed to intentionally violate the Gricean norms of readers. In a book where a family went out for a picnic, instead of calling the device that carried their food a picnic basket, it was called a picnic sack. In a book describing a whale jumping out of the water, the whale is described as charging out of the water. By intentionally violating the expectations of readers, text designers can force students to rely on decoding based strategies for solving unknown words.

While there is no explicit endorsement of this strategy governing vocabulary selection in these or any decodable texts, text designers are faced with a predicament. If texts that offer too much syntactic or contextual support distract readers from paying attention to orthographic details, then in order to promote attention to orthographic details, texts that violate syntactic and semantic expectations and provide limited contextual support should be the most helpful. Yet texts designed with minimal contextual support would likely be more challenging for struggling early readers (Amsterdam, Ammon, & Simons, 1990; Hiebert et al, 2010a). Adams (1997) suggests that even if decoding is more challenging initially than being supported in recognizing a word through context or pictures, that once students have successfully decoded the word a number of times, the task becomes easier and easier. But what about when students aren't successful in decoding the words in a story? Adams (1997) goes on to argue, "If too many words of a text are unfamiliar, then the reading experience becomes tedious and frustrating. In these cases, children should be given help or, better yet, an easier text" (p. 432).

Catch 22 – One size fits all curriculum. A primary purpose of the texts used in this study and in all commercial curriculum using LTTM decodable texts is to allow students opportunities to practice decoding using the sight words and phonics content that students are taught in class. The LTTM texts comprising the Open Court Reading program reflect a progression in early reading development, from where kindergarten left off (with a certain amount of review) to where 1st graders need to be by the end of the year. First grade students reading *Huck Finn* or *The Odyssey* would be inappropriate, but it would be equally inappropriate for them to be reading "A cat sat on a mat" at the end of 1st grade as well. Since the books reflect the learning progression built into the classroom curriculum, all students must read the same books at the same time in order to keep pace with classroom instruction. Commercial curriculum built around LTTM decodable texts plan that every student in class, regardless of ability level, read the exact same texts on the exact same day.

Yet this practice of having every student read the same book on the same day introduces a Catch-22 into the works. As Adams (1997, 2009) argues, in order for LTTM texts to support beginning readers, students need to be able to read them successfully. Yet because of the one size fits all curriculum, top tercile students and bottom tercile students read the same book on the same day. Every student in every class across the country using Open Court Reading read *Kim's Trip* at the same time. For many students, the book was far too hard and lead to problematic error

rates. The context and pictures included didn't help students decode unfamiliar words successfully and they were unable to decode them accurately on their own and thus produced a high number of errors. These errors lead to low fluency rates, and undermined the development of orthographic learning and automatic word recognition. Additionally, researchers are fairly unanimous in arguing that overly difficult texts frustrate and demoralize the readers who are forced to read them (Adams, 1997; Allington, 2005). Over 75% of the time bottom tercile students read a decodable text, it was below 94% accuracy. If, according to Juel & Roper-Schnieder (1985), the types of texts students use in learning to read has the most significant impact on their early reading development, what is the impact of having 75% of the books you read be too difficult?

Three Patterns of Responses from Struggling Readers

Even though the primary finding identified by the graphical analysis of student reading behavior was individual words with significantly high error rates, it was also apparent that student errors frequently clustered together as well. Some of these error clusters correlated with passages in specific books where a substantial number of students produced clusters of errors on the same passage. Other clusters of errors were centered around frequently missed words. A simple Excel data analysis tool was designed to scan for clusters of errors, three errors in five words, four errors in six words, etc. 58 error clusters were produced at the four errors in six words level, 48 of them by students in the bottom tercile and the other ten by middle tercile students. Roughly half of the 58 error clusters at the four errors in six words level were related to specific passages (i.e. at least five students made four mistakes in six words on the same passage of text). 192 students produced error clusters at the three in five level, 162 of them by bottom tercile students, 28 by middle tercile students and two by students in the top tercile. Fifteen different passages across the 24 books in this study had at least five students produce three errors within five words.

The passage of text with the highest rate of error clusters comes from the book, *The Blur with Fur* (book 49, introduced in early December). The passage in questions reads, "The blur with fur dashed and crashed. It curled and twirled." The target phonics content for the book was the 'ur' vowel pattern. 16 students misread three out of five words in the passage, four students misread four out of six words. Six different students missed all four of the more challenging words to decode—*dashed, crashed, curled, twirled*. Nine more students missed three of the four more difficult words. The word *dashed* was missed by 20/45 students, the word *crashed* by 15/45 students, the word *curled* by 23/45 students and the word *twirled* by 18/45 students. While the specific error counts are problematic in terms of decoding accuracy and its effect on the development of automatic word recognition, student reading responses on these passages offer a window into types of reading behaviors produced by students when faced with texts they are unable to read successfully.

Making no attempt to decode unfamiliar words. One frequently occurring pattern in the face of multiple errors during the reading of *The Blur with Fur* passage as well as during other passages where error rates were abnormally high was that students stopped attempting to decode unfamiliar words and simply requested to be told a word they did not immediately recognize. Almost no students missed the high-frequency connecting words between the four difficult words in the passage. Both instances of the word *and* were read correctly by all students and the word *it* was missed by two bottom tercile students, and interestingly, by one middle tercile and two top tercile students who read the entire passage as separated by instances of the

word *and* (i.e. The blur with fur dashed and crashed and curled and twirled). It wasn't that students stopped reading or gave up entirely, they simply stopped attempting to read words they didn't recognize immediately.

Eight different students made no attempt to decode the word *curled* and eleven different students made no attempt to decode the word *twirled*. The pattern of refusing to attempt to decode unfamiliar words occurred repeatedly for struggling readers across the 24 books. While the methodological decision to follow the DIBELS assessment protocol could have lead students to rely on investigator assistance, only two students made no attempt to decode the first two difficult words (*dashed* and *crashed*) and students read each of the connector words without difficulty. Many of the bottom tercile students became adept at quickly identifying words they did not recognize and that appeared difficult and refused to make attempts to decode them. Adams (2009) concludes that the critical factor supporting the development of automatic word recognition is that students successfully decode words they read in text. She goes on to argue, "Whether or not students do so depends not just on whether they have learned to decode but, equally pivotal, on whether they have developed the inclination to do so when encountering new words in text" (p. 35). By repeatedly tasking struggling readers with decoding overly difficult words, LTTM decodable texts predispose struggling readers to give up on long, unfamiliar words they don't recognize. This pattern of response undermines the development of the "inclination to decode" which is the primary purpose LTTM decodable texts were created to serve.

Tolerating gibberish. A second frequently occurring pattern seen in the data occurred when students attempted to decode unfamiliar words, but with no seeming awareness of what they were saying—effectively producing gibberish. In these situations, students attempted to decode the text in question but did not monitor their decoding output to determine if it made sense, where making sense is defined as fitting typical English syntax and where it mapped onto existing knowledge and experience. In the book *The Bee and the Deer*, a section of text near the end reads, "The little deer felt timid and meek." (book 75, introduced in early February). The challenge for students was what to do with two words that were most likely not in their oral vocabulary. Both words, *timid* and *meek* are relatively easy to decode, not especially long and consist of high-frequency letter combinations used repeatedly in English. Yet looking at what students produced when attempting to read these words demonstrates a significant problem. Nine students missed the word *felt*, twenty six students missed the word *timid* and fifteen students missed the word *meek*. Of the students who attempted to decode the word *timid,* about a third produced 'timmed', another third produced 'timed' and the remaining third produced a range from 'tempted' and 'tilt' to 'tired'. For students who missed the word *meek*, it was primarily read as 'make' with alternatives of 'meet', 'meck' and 'met'. Twelve students produced a reading of the passage that went something like, "The deer felt timed and make." While the inaccurate reading is problematic in its own right, what is much more dangerous is the fact that not a single student stopped and either attempted to correct their mistakes or indicated in any what that what they had just read made no sense.

Time after time across the 24 books in this study, students made errors when attempting to solve unfamiliar words but rarely did they go back and attempt to correct their mistakes or indicate in any way that they were aware if what they read made no sense. In the book, *A Photo for Fred*, a character is described as looking like a "funny pheasant." Just under a third of students missed the word *pheasant*, but most of them said the student looked like a "fee-sant." When asked what a pheasant was, only two of the students who had missed the word knew it was a type of a bird. This tendency to "tolerate gibberish" occurred much more regularly with bottom

tercile readers and occurred frequently in relation to sections of reading with low accuracy which often correlated with sections including low-frequency words outside the student's oral vocabulary. Students in the top tercile never produced readings that could be considered gibberish and middle tercile readers did so rarely. Producing a reading of a text, where what the student says made no sense was almost exclusively a reading pattern of bottom tercile students and is a direct function of the students' inability to read the texts accurately.

A critical question in relation to this discussion is what would be the difference for a struggling reader if they accurately decoded the words timid and meek? Even if they accurately decoded the two words, they still wouldn't know what they meant. Their willingness to tolerate gibberish is not just a function of their reading ability, but equally or more significantly here, a case of encountering texts with words they don't understand and can't figure out from context. To struggling readers, the text, "The deer felt timid and meek" is gibberish whether they decode it accurately or not. The repeated use of low-frequency words outside the oral vocabulary of many struggling early readers turns the task of reading a book into a task of decoding nonsense words. Just as Juel & Roper-Schnieder (1985) argued, the type of texts that students use in learning to read has the most significant impact on their early reading development. By requiring struggling readers to read texts that repeatedly include low-frequency words not in their oral vocabulary, students are repeatedly forced to "tolerate gibberish" and develop the habit of producing gibberish and reading on. This deformation of the task of reading for the sake of LTTM decodability has significant consequences for struggling readers for whom LTTM decodable texts make up the greatest percentage of their early literacy experience.

Inventing text. In the third pattern, students rely heavily on context to identify unfamiliar words and significantly ignore phonetic information in the task of identifying unfamiliar words. The errors described in the previous section regarding the words *timid* and *meek* were clearly failed attempts to decode the words in question. The responses phonetically recoded many of the graphemes accurately, but failed to come up with an accurate final reading. The majority of the attempts prioritized attempting to decode accurately over attempting to maintain a sense of meaning or context. In the book, *Jen's Pen* (book 28, introduced in November), the farmer Ted promises his hen, Jen, to "mend Jen's pen." The word *mend* was missed by 21 of 45 students, and by 13 of 15 in the bottom tercile. Yet over half of the students making errors substituted the words *made* or *make* for the word *mend*. Both of these substitutions use the initial 'm' sound, but clearly depart from any attempt to decode the word sequentially. What they do show, however, is an attempt to maintain some sense of what is happening in the text. The students knew that the farmer had promised to fix Jen's pen, they saw a word that started with 'm' and needed something that could mean fix, so they came up with *made* and *make*. Across all 24 texts, students frequently substituted more familiar or contextually relevant alternatives for words they struggled to decode. "Coils of silver ribbon" was read as "Curls of silver ribbon". In the story "Craig Sails", Craig's and Gail's name were confused and interchanged frequently. A babbling brook is a bubbling brook, "The cat is female" turned into, "The cat is family."

In almost every instance, students attempted to decode the target words. In some instances, the alternative words were selected quickly and the student read on not noticing their mistake. In other instances, students struggled to decode, attempting multiple options, finally settling on an option that helped preserve or generate a sense of meaning in the story. All beginning readers make mistakes and often making an initial mistake in a passage disrupts the reading of subsequent words (Torgesen, 2002). Stanovich, Siegel, & Gottardo (1997) argue that struggling readers overly depend on context to help them identify unfamiliar words. Share (2008)

described situations where struggling readers had made so many errors that they didn't have a sufficient grasp of story context to generate contextually based guesses. Both of these situations seem relevant in describing the above mentioned reading behaviors of students in this study. When students have enough of the sense of the story and they confront an unfamiliar word they are either unable or unwilling to decode, they frequently generate a plausible alternative and continue reading. When a student's reading of a text has broken down or been so laborious that no sense of the meaning of the story has been established, students either passively request to be told unfamiliar words or attempt to decode them mechanically, unaware if what they say makes any sense or is even a real word.

All three of these patterns of behavior occur almost exclusively with students in the bottom tercile. It could be argued that these reading behaviors (refusing to attempt to decode, tolerating gibberish and contextual guessing) are established signs of struggling readers and that these behaviors have nothing to do with the texts in question. Yet the data from this study clearly show that bottom tercile students frequently read through sections of text and even some entire books without significant struggle; producing none of the above cited error patterns. No one would argue that struggling readers randomly produce gibberish or intermittently refuse to attempt to decode unfamiliar words. The data from this study suggest that these behaviors are responses to texts—texts with specific features that elicit these types of reading behaviors. At the simplest level, these reading behaviors are the result of attempting to read texts that are overly difficult. But at a more specific level, these behaviors are the response of struggling early readers to features of texts designed specifically to help them learn to read.

Conclusion

According to proponents, the primary goal of lesson-to-text-matching decodable texts are to provide students with opportunities to practice decoding while reading connected texts constructed primarily of spelling patterns and sight words students have been taught previously in class. Through repeated opportunities to successfully decode both familiar and unfamiliar words, students become fluent decoders and begin to recognize a growing pool of words automatically. 1st grade students spend roughly six months working with LTTM decodable texts after which it is assumed that their reading system has developed sufficiently such that students could begin to read a broad range of more conventional early literature.

In order for decodable texts to support beginning readers in this progression, beginning readers must be able to read them accurately. Decodable texts are designed to match the required learning progression built into 1st grade, not to match the reading level of the specific students tasked with reading them. Many students reading decodable texts as part of their daily classroom instruction are not capable of reading the text for that day accurately. Struggling early readers are constantly being asked to read and reread texts that are too difficult for them. This produces error prone reading and low fluency rates. It undermines the development of decoding fluency and automatic word recognition and in certain instances leads to the development of habits such as refusing to attempt to decode unfamiliar words, tolerating gibberish and ignoring phonetic information. Inaccurate, slow, laborious reading produces frustration, a lack of motivation for reading and a sense of one's self as a poor reader. The students most in need of support and acceleration in learning to read are the ones most affected by the use of materials that disproportionately disadvantage them.

The most important piece of research cited in support of the use of decodable texts (Juel & Roper-Schnieder, 1985) did not use texts where decodability was based on previous classroom

instruction and the study specifically left out low-performing and high-performing students. Thus the most commonly cited study in defense of the use of decodable texts provides no evidence for the use of LTTM decodable texts and says nothing about the effect of decodable texts for struggling or advanced readers. A wide variety of research has indicated a strong correlation between reading accurary and word learning (Cunningham, 2006; Nation et al, 2005; Share, 1999). Hiebert et al (2010a) and Compton et al (2004) both found that word decodability did not predict reading accuracy nor did it predict reading fluency. While not a single study has demonstrated the necessity nor the efficacy of LTTM decodable texts for beginning readers, millions of children every year are required by state mandates to read LTTM decodable texts to help them learn to read, regardless of whether or not they can actually read the texts successfully.

What Juel & Roper-Schnieder (1985) actually offers about the effects of using LTTM decodable texts for 1st grade reading instruction is a sobering warning. According to the authors, "the results do suggest that selection of text used very early in first grade may, at least in part, determine the strategies and cues children learn to use, and persist in using, in subsequent instruction" (p. 150). Yet when the selection of texts is such that struggling early readers are unable to read them accurately or fluently, the strategies they learn to use and persist in using are not the ones intended. When faced with texts they are unable to read, struggling early readers frequently stop attempting to read independently and ask teachers to read for them, they stop attempting to make sense of what they're reading and decode mechanically or they stop paying attention to the text and invent a story rather than read the text that is present.

Juel & Roper-Schnieder argued, "the types of words which appear in beginning reading texts may well exert a more powerful influence in shaping children's word identification strategies than the method of instruction" (p. 151). But what is the effect of the consistent use of low-frequency words outside of children's oral vocabulary? What word identification strategies are fostered when a child attempts to decode a word they don't recognize even if they decode it successfully? What is the effect of having over 70% of the words in LTTM decodable texts appear only once in the entire series (Foorman et al, 2004)? Hiebert et al's (2010a) conclusion was that the texts end up actually fostering the exact type of visual memorization strategies the texts are designed to avoid. Since a small pool of irregular sight words are repeated frequently across all the texts, those words are memorized visually. The remaining words appear so infrequently, beginning readers rarely encounter new words enough times to begin to recognize them automatically—they simply continually practice decoding. The only words they "know" are ones they have memorized visually.

None of these findings or conclusions are new discoveries, recently brought to light. From the earliest use of linguistic readers, critics have argued that the unusual language patterns and vocabulary choices cause difficulty for beginning readers, especially struggling early readers. Yet proponents of the use of LTTM decodable texts and the state boards of education mandating their use have "doubled down" so to speak, ignoring both the historical critiques of their effectiveness as well as the growing recent body of research arguing against their effectiveness. While middle-performing students learn to read using these texts, they offer little or no support for advanced readers. But the situation for struggling early readers is much worse. There is clear and growing evidence that LTTM decodable texts disproportionately disadvantage struggling early readers—the very group the tests are supposedly designed to help.

This idea of a disproportionate effect of text type on the low-achieving students can perhaps be understood in light of an earlier study by McNamara, Kintsch, Songer & Kintsch (1996) where they found a similar but opposite effect of a disproportionate effect of educational

materials depending on the achievement level of the students using the materials. Their study identified an interaction on a post-test between the informational quality of the reading materials used in the intervention and the level of student background knowledge specific to the topic of the research in question. The use of highly informative reading materials lead to higher learning outcomes for low-knowledge readers as compared to the use of less informative reading materials—as would be expected. For the high-knowledge students, however, use of the less informative materials actually lead to greater gains on the post-test than did use of the more informative materials. In explaining their findings, the authors proposed that the minimally informative instructional materials offered the high-knowledge students a chance to fill in the missing information, requiring a degree of engagement and analysis that lead to greater learning gains. The low-knowledge students, not possessing the requisite background information, were unable to make the types of inferences and conclusions required by the less informative texts and were thus unable to benefit from the "opportunity" provided by the less informative texts.

So how do the McNamara et al (1996) findings relate to the results from this study? As described in the section analyzing text features affecting student reading performance, the decodable texts used in this study (and LTTM decodable texts as a genre for the most part) tend to be too difficult for struggling beginning readers. They include low-frequency words that students do not know; over 70% of the words appear only once across the entire series of texts and the requirement that 75% of the words be decodable based on previous instruction leads to atypical language patterns. In order to focus on the task of decoding, pictures are minimally supportive. In order to avoid students using context to guess unfamiliar words, vocabulary and phrasing are often intentionally unpredictable.

While mid- and high-level readers encountered the same texts, the resources they brought to the task were dramatically different (Hiebert, 2009; Neuman & Celano, 2006), enabling them to overcome these obstacles. Struggling readers were frequently unable to overcome these obstacles, producing error rates and fluency rates that significantly undermined learning gains. Over 70% of the time bottom tercile students read a decodable text, it was below Bett's (1946) cutoff for frustration level, even though they had read the text twice before the day previously. Occasional errors happen even with expert readers and the small number of errors produced by most middle tercile students caused their reading systems to bend, but not break. For bottom tercile readers, there were repeated signs of breakdown in their reading systems. The percentages of errors and the lack of fluency represent a tipping point where Adams (1997) even argues, ""If too many words of a text are unfamiliar, then the reading experience becomes tedious and frustrating. In these cases, children should be given help or, better yet, an easier text" (online document). Unfortunately for struggling early readers, the requirement that the same text be read by all students on the same day, regardless of ability level, means that struggling early readers are disproportionately impacted by the texts being used to teach them to read.

Are There Better Ways to Support Struggling Beginning Readers?

First, if the goal of the books is to accurately reflect the content that students must learn, and it is accepted that reading books that are too hard undermines learning rather than supports it, then struggling students will either need alternate materials or be allowed to progress through the standard materials at a different pace than high-performing or mid-performing students. If we know that reading overly difficult texts undermines reading development, requiring students to read texts that are too difficult would be tantamount to educational malpractice. Asking students

to reread the same early decodable texts over and over again—low-interest texts with uncommon language patterns and unfamiliar vocabulary—would be educational malpractice as well.

One possibility is to attempt to provide enough external support such that even struggling beginning readers could read the current crop of LTTM decodable texts accurately and fluently. Classroom instruction could be redesigned to offer better book introductions, more strategic pre-teaching of vocabulary and low-frequency words, additional practice with target phonics content and more time spent practicing sight words. But each of these activities are already part of current instructional practice, so assuming more of the same will lead to improved outcomes is potentially problematic. Additionally, the block of time currently assigned to literacy development is already the largest instructional block and takes up the greatest percentage of a tightly packed school day. If two hours of literacy instruction/day doesn't prepare students to read the decodable texts, counting on an additional half hour of instruction to accomplish the task seems like wishful thinking. And while one could imagine hiring aids for 1st grade classrooms to support struggling readers or offering technology support through computer instructional programs or providing teachers with additional professional development—all of these are costly and complicated solutions for a very simple problem facing millions of public elementary school children on a daily basis.

It is obviously possible to "rework" the current design strategies for LTTM decodable texts to improve their usefulness for struggling beginning readers. The simplest and most basic change would be to "ban" the use of low-frequency words that are likely outside the oral vocabulary of struggling readers. Given that there is no attempt to teach these words as new vocabulary and they appear only as examples of LTTM decodable words, more common, more recognizable words could be selected. The mandate for 75% of the words in a text meeting LTTM decodability standards could be reduced to allow for more natural language and the decision to have over 70% of the words appear only a single time across an entire series of texts could be replaced with a strategy of word repetition that supports the development of automatic word recognition as well as reflects patterns of normal language. Each of these changes would most likely significantly improve the reading outcomes of struggling beginning readers using these texts.

Some of these suggestions, however, fly in the face of two critical arguments made by proponents of LTTM decodable texts. The first can be roughly described as the belief that there is a threshold for the percentage of decodable words that must be present in a text for students to develop the tendency to see decoding as the primary strategy for identifying unfamiliar words. Beck's (1997) comments regarding the percentage of words in a text that needed to be decodable in order to establish the decoding habit seem to have echoed in the ears of state education officials in Texas who have recently mandated even stricter requirements for text decodability than had been previously decided. If LTTM texts were redesigned, however, such that fewer decodable words were required, then students would spend less time decoding and would more frequently encounter words where attempting to decode would be counterproductive. The second tenet is the idea that if students can identify unfamiliar words through the use of context, looking at pictures, text predictability or the use of syntax, then less attention is paid to the orthographic details of words, students don't remember the orthographic details as well, and again, the habit of decoding loses its preeminence.

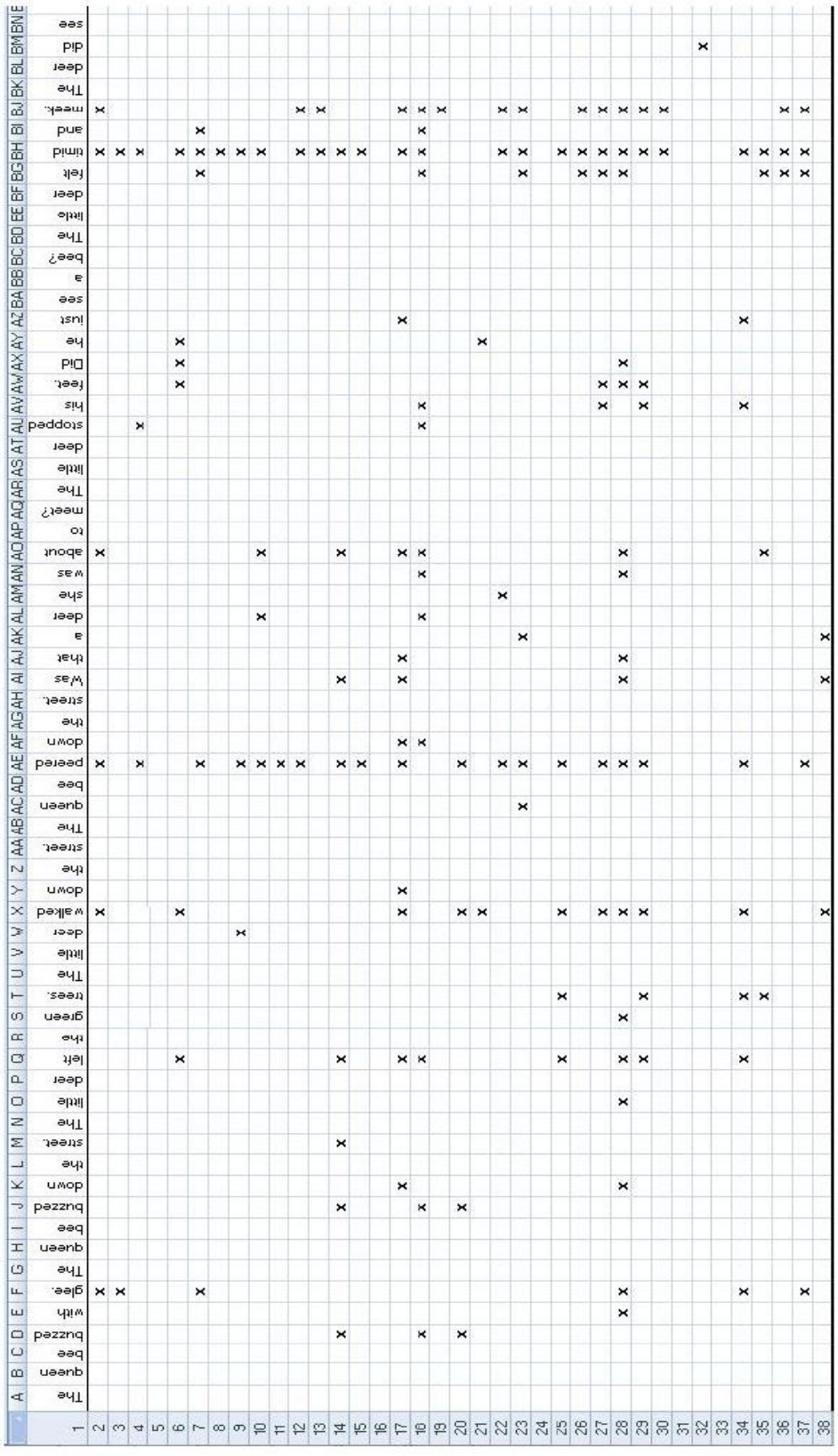
While Adams (2009) and Nation et al (2005) write in favor of these arguments supporting the use of LTTM decodable texts, Hiebert et al (2010b) and Allington (2005) have questioned the methods, findings and conclusions. What is clear, however, is that the research base

analyzing these questions is underwhelming and inconclusive. Given the degree to which these two arguments frame and shape the debate over the nature of early reading materials, it is essential that literacy researchers address these critical assumptions. Is it true that students who learn to read with books with fewer than 75% decodable words won't be able to decode and will be less effective readers? Is it true that if students read books with informative pictures, supportive context and useful syntactic cues that they will be less effective readers and less competent decoders? Are LTTM decodable texts the best available resource for students who are unable to read the texts accurately and fluently? No one has answered these questions with any degree of confidence, yet state boards of education and commercial curriculum designers continue on as if the answers were established facts. Given the number of students that struggle to learn to read by the end of 1st grade and the number of students that continue struggling to read throughout elementary school, it is dangerous to ignore the growing body of evidence challenging the assumptions undergirding the design and use of LTTM decodable texts.

Appendix 1 - 20 Most Frequently Missed Words

Word	Number of students missing the word
female	18
problem	18
pups	18
twirled	18
Arthur	18
peered	19
turned	19
dashed	20
recent	20
yanked	20
corner	21
narrator	21
curb	22
wriggling	22
blasted	23
curled	23
babbling	24
Birch	25
tramped	25
trudged	33

Appendix 2



This figure depicts the words in a specific story (The Bee and the Deer) across the top of the worksheet and the list of student id's represented by the line numbers along the left edge of the worksheet. The blank cells represent words that students read correctly, cells with x's in them represent words where students made errors. While there is clearly a degree of randomness to the error data, there are clearly present vertical patterns that represent specific passages in the book where as many as 70% of students made mistakes at the exact same place. These error patterns were analyzed in an attempt to identify text features that corresponded with consistently high error rates.

References

Adams, M. (1990). *Beginning to read*. Cambridge, MA: MIT Press.

Adams, M.J. (1997). Myths and realities about words and literacy. School psychology review, volume: 26 issue: 3 page: 425-436.

Adams, M. J. (2009). Decodable Text: Why, When, and How? In E. H. Hiebert & M. Sailors (Eds.), *Finding the Right Texts: What Works for Beginning and Struggling Readers* (pp. 23-46). New York: Guilford Press.

Allington, R. L. (1997). Overselling phonics. *Reading Today*, 15-16.

Allington, R. L. (2005). Ideology Is Still Trumping Evidence. *Phi Delta Kappan*, 462-468.

Amsterdam, L., Ammon, P., & Simons, H. D. (1990). Children's Elicited Imitations of Controlled and Rewritten Reading Texts. *Journal of Educational Psychology*, 82(3), 486-490.

Armbruster, B. B., Lehr, F., Osborn, J., Adler, C. R., & National Institute for Literacy (U.S.). (2001). *Put reading first: The research building blocks for teaching children to read: kindergarten through grade 3*. Washington, D.C.: National Institute for Literacy, National Institute of Child Health and Human Development, U.S. Dept. of Education.

Beck, I. L. (1997, October/November). Response to "overselling phonics." *Reading Today, 17.*

Beck, I. L., & McCaslin, E. S. (1978). *An analysis of dimensions that affect the development of code-breaking ability in eight beginning reading programs* (Report No. 6). Pittsburgh, PA: Learning Research and Development Center.

Beck, I.L., & Block, K.K. (1979). An analysis of the dimensions that affect the development of code-breaking ability in eight beginning reading programs. Pittsburgh: University of Pittsburgh, Learning Research and Development Center.

Betts, Emmett A. (1946). *Foundations of Reading Instruction with Emphasis on Differentiated Guidance*. New York, New York: American Book Company.

Bloomfield, L., & Barnhart, C.L. (1961). *Let's read: A linguistic approach*. Detroit, MI: Wayne State University Press.

Bryant , N. D. (*1975*). *Bryant Test of Basic Decoding Skills*. New York: Teachers College Press.

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

Brysbaert, M., & New, B. (December 01, 2009). Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods, 41*, 4, 977-990.

California State Board of Education. (2006, April 17). *Criteria for evaluating instructional materials: Reading/language arts*. Retrieved from www.cde.ca.gov/ci/rl/im/.

Compton, D. L., Appleton, A. C., & Hosp, M. K. (August 01, 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 and Practice*, *19*, 3, 176-184.

Cunningham, A. E. (2006). Accounting for children's orthographic learning while reading text: Do children self-teach? *Journal of Experimental Child Psychology*, 95, 56-77.

Duong, T., Callahan, M., Pearson, P. D., & Price, P. (2006, June). *Update on 2006 data collection results from the Northern California team*. Paper presented at the Annual Meeting of the TBAL Assessment Consortium, Los Angeles, CA.

Ehri, L. C. (1998). Grapheme-phoneme knowledge is essential for learning to read words in

English. In J. L. Metsala & L. C. Ehri (Eds.), Word recognition in beginning literacy (pp. 3-40). Mahwah, NJ: Lawrence Erlbaum Associates.

Ehri, L.C. (2005). Learning to Read Words: Theory, Findings, and Issues. *Scientific Studies of Readings*, 9(2), 167-188.

Ehri, L. C., & Roberts, K. T. (1979). Do beginners learn printed words better in contexts or in isolation? *Child Development*, 50, 675–685.

Flesh, R.F. (1955). *Why Johnny can't read*. New York: Harper & Row.

Foorman, B. R., Francis, D. J., Davidson, K. C, Harm, M. W., & Griffin, J. (2004). Variability in text features in six grade 1 basal reading programs. *Scientific Studies of Reading*, 8, 167-197.

Frey, N. (n.d.). *The national right to read foundation - resources*. Retrieved from http://www.nrrf.org/resources.htm

Frey, R. (2011). *Word Learning and Struggling Readers: A critical analysis of theories and data regarding the role of decodable texts in supporting word learning for beginning readers.* Paper presented at the annual meeting of the American Educational Research Association. New Orleans, Louisiana.

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(3), 239-256.

Good, R. H., Wallin, J., Simmons, D. C., Kame'enui, E. J., & Kaminski, R. A. (2002). System-wide Percentile Ranks for DIBELS Benchmark Assessment (Technical Report 9). Eugene, OR: University of Oregon.

Hart, B., & Risley, T. R. (1995). *Meaningful differences in the everyday experience of young American children*. Baltimore: P.H. Brookes.

Hiebert, E.H. (2005). State Reform Policies and the Task Textbooks Pose for First-Grade Readers. *The Elementary School Journal*, 105(3), 245-266.

Hiebert, E. H. (2009). The (Mis)Match between Texts and Students Who Depend on Schools to Become Literate. In E. H. Hiebert & M. Sailors (Eds.), *Finding the Right Texts: What Works for Beginning and Struggling Readers* (pp. 1-20). New York: Guilford Press.

Hiebert, E. H., & Fisher, C. W. (September 01, 2007). Critical Word Factor in Texts for Beginning Readers. *Journal of Educational Research, 101*, 1, 3-11.

Hiebert, E. & Mesmer, H. (2006). Perspectives on the Difficulty of Beginning Reading Texts. In D. Dickinson & S. Neuman (Eds.), *Handbook of Early Literacy Research: Volume 2* (pp. 395-409). New York, NY: Guilford Press.

Hiebert, E.H., Mesmer, H.A., & Cunningham, J.W. (March 7, 2010). *The Effectiveness of Decodable Texts for At-Risk Beginning Readers: A Review of Research.* Report to Student Achievement Partners for the Text Complexity.

Hiebert, E.H., Stewart, J.D., & Uzicanin, M. (July 10, 2010). *A Comparison of Word Features Affecting Word Recognition of At-Risk Beginning Readers and Their Peers.* Paper to be presented at the annual meeting of the Society for the Scientific Study of Reading. Berlin, Germany.

Hoffman, J. V., Sailors, M., & Patterson, E. U. (2002). Decodable Texts for Beginning Reading Instruction: The Year 2000 Basals. Center for the Improvement of Early Reading Achievement.

Huey, E. B. (March 27, 1908). A STUDY OF VISUAL FIXATION. *Science, 27,* 691, 501-502.

Jenkins, J. R., Peyton, J. A., Sanders, E. A., & Vadasy, P. F. (2004). Effects of Reading Decodable Texts in Supplemental First-Grade Tutoring. *Scientific Studies of Reading*, 8(1), 53-85.

Juel, C., & Roper-Schneider, D. (1985). The Influence of Basal Readers on First Grade Reading. *Reading Research Quarterly*, 20(2), 134-152.

Landi, N., Perfetti, C. A., Bolger, D. J., Dunlap, S., & Foorman, B. R. (June 01, 2006). The role of discourse context in developing word form representations: A paradoxical relation between reading and learning. *Journal of Experimental Child Psychology*, *94*, 2, 114-133.

Lipson, M. Y., & Wixson, K. K. (1986). Reading Disability Research: An Interactionist Perspective. *Review of Educational Research*, 56(1), 111-136.

Mathews, M. M. (1966). *Teaching to read: Historically considered*. Chicago, IL: University of Chicago Press.

McKeown, M. G., and I. L. Beck. "Direct and Rich Vocabulary Instruction." In *Vocabulary Instruction*, edited by J. F. Baumann and E. J. Kame'enui, 13-27. New York: Guilford Press, 2004.

McNamara, D. S., Kintsch, E., Songer, N. B., & Kintsch, W. (1996). Are Good Texts Always Better? Interactions of Text Coherence, Background Knowledge, and Levels of Understanding in Learning From Text. *Cognition and Instruction* 14(1), 1-43.

Menon, S., & Hiebert, E. H. (January 01, 2005). A comparison of first graders' reading with little books or literature-based basal anthologies. *Reading Research Quarterly*, 40, 1.)

Mesmer, H. A. E. (2001). Decodable Text: A Review of What We Know. *Reading Research and Instruction*, 40(2), 121-142.

Mesmer, H.A.E. (2010). Textual Scaffolds for Developing Fluency in Beginning Readers: Accuracy and Reading Rate in Qualitatively Leveled and Decodable Text. *Literacy Research and Instruction*, 49, 20-39.

Nation, K., Snowling, M.J., & Clarke, P. (2007). Dissecting the relationship between language skills and learning to read: Semantic and phonological contributions to new vocabulary learning in children with poor reading comprehension. *Advances in Speech-Language Pathology*, 992), 131-139.

Neuman, S. B., & Celano, D. (April 01, 2006). The Knowledge Gap: Implications of Leveling the Playing Field for Low-Income and Middle-Income Children. *Reading Research Quarterly*, 41, 2, 176-201.

National Institute of Child Health and Human Development. (2000). *Report of the National Reading Panel. Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction* (NIH Publication No. 00-4769). Washington, DC: U.S. Government Printing Office.

Pearson, P. D. (2002). American reading instruction since 1967. In N. B. Smith, *American reading instruction* (special edition; pp. 419-486). Newark, DE: International Reading Association.

Saunders, A. (1999). Eureka! School reform works. *San Francisco Chronicle*, June 29, A21.

Share, D. L. (1995). Phonological recoding and self-teaching: Sin qua non of reading acquisition. *Cognition*, 55, 151–218.

Share, D. L. (1999). Phonological recoding and orthographic learning: a direct test of the self teaching hypothesis. *Journal of Experimental Child Psychology*, 72, 95–129.

Share, D. L. (January 01, 2004). Knowing letter names and learning letter sounds: a causal connection. *Journal of Experimental Child Psychology*, 88, 3, 213-33.

Share, D.L. (2008). On the Anglocentricities of Current Reading Research and Practice: The Perils of Overreliance on an "Outlier" Orthography. *Psychological Bulletin*, 13494), 584-615.

Smith, F., & Goodman, K. S. (January 01, 1971). On the Psycholinguistic Method of Teaching Reading. *The Elementary School Journal*, 71, 4, 177-181.

Roit, M. (2008). *Sra imagine it! professional development administrator's guide*. Columbus: SRA/ McGraw-Hill.

Stanovich, K.E., Siegel, L.S. & Gottardo, A. (1997). Converging evidence for phonological and surface subtypes of reading disability, *Journal of Educational Psychology,* 89: 114–127.

Torgesen, J.K. (2002). The Prevention of Reading Difficulties. *Journal of School Psychology*, 40(1), 7-26.

University of Oregon. (2012). *Big ideas on beginning reading*. Retrieved from http://reading.uoregon.edu/big_ideas/.