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

Rick Chan Frey
University of California, Berkeley
April 8th, 2011

Abstract:
The leading theory of word learning posits that repeated successful attempts to decode words in text lead to more fluent processing with the eventual outcome that words are no longer decoded, but read as entire units, as "known words". The critical assumption underlying this model is that words are decoded accurately when first encountered in text. After reviewing the research literature on the theory of word learning and recent research analyzing the relationship between accuracy and word learning, data is presented demonstrating low accuracy scores for struggling early readers with typical decodable texts. These findings are discussed in relation to the primary model of word learning and present a challenge to the current almost exclusive usage of decodable texts as the primary reading materials for beginning readers.

Running Head: Word learning and struggling readers

For reprints or questions contact: rickchanfrey@berkeley.edu

The simple theory of word learning

It is rare in a field as contentious as early reading research to find an important topic where there's close to unanimous agreement. Yet when it comes to understanding how developing readers learn words while reading text, researchers across a wide range of perspectives over the last thirty years have settled on a fairly simple description of the basic process. What I hereafter refer to as the simple theory of word learning describes the process of students reading a word in context, successfully decoding that word, and slowly, over repeated encounters, developing a degree of fluency or automaticity in their reading such that the word becomes "known". In order to get a clearer sense of this theory and the important constraints and ramifications, let me briefly review some of the foundational research to frame the state of the field in relation to theories of word learning.

Ehri (2005)

> People used to believe that readers memorized associations between visual features such as the shapes of words and their meanings. This was one justification for the look–say, whole-word method of teaching reading. However, the visual explanation is inadequate.
>
> To summarize, readers learn to process spellings of words as phonemic maps that lay out elements of their pronunciations visually. Beginners become skilled at computing these mapping relations spontaneously when they read new words. This is the critical event for sight word learning. Grapheme–phoneme connections provide a powerful mnemonic system. They provide the glue that bonds letters in written words to their pronunciations in memory along with meanings. Once the alphabetic mapping system is known, readers can build a vocabulary of sight words easily. (Ehri, 2005, pg. 172)

Ehri (2005), building upon earlier work (Ehri, 1998), offers a recent history of the research on learning words, beginning with a brief overview of her original theory of four stages of reading development and word analysis. Ehri proposed that beginning readers moved through a series of stages marked by their growing attention to the visual details of written language which allows them to analyze words phonetically in more and more complex ways. Ehri outlines the steps in how experienced readers read words automatically and how beginning readers develop the skills and knowledge that enable them to do the same.

Adams (2009)

> To sound-out a word, a student must examine the letters left to right, in sequence. This causes the ordered, left-to-right sequence of letters to leave a trace of itself in memory. At the same time, because the student is sounding the word, the trace that results includes the letters' connection to their phonology or speech sounds. That is true for the individual letters and groups of letters, as well as the word as a whole. … 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. Its spelling, pronunciation, and meaning seem to come to mind at once. The word has become a *sight word*. (pg. 33, Finding the Right Texts)

Adams (2009) provides an overview of her work in the field of reading research over the last twenty plus years and chronicles the development of her thinking in relation to word learning and how best to design texts to support students' efforts to develop decoding skills. Building on Adams (1990), she provides an overview of past research, framed uniquely towards the task of discussing the role that text design plays in supporting beginning readers. She describes the ways in which a well-designed text

allows readers opportunities to utilize the phonics information they've been taught through classroom instruction, thus engraining the habit of attending to the letter-sound relationships that make up a specific word. Adams argues that as a child successfully decodes an individual word, not only are they engaging in the process by which that word can become automatic, they're also developing an awareness of the patterns and combinations of letters that form words and growing in their ability to use increasingly complicated combinations in order to successfully analyze new words.

Share (1999)

> According to this [self-teaching] model, phonological recoding (print-to-sound translation) performs a self-teaching function enabling the learner to acquire the detailed orthographic representations necessary for fast, efficient visual word recognition. … Each successful identification (decoding) of a new word in the course of a child's independent reading of text is assumed to provide an opportunity to acquire the word-specific orthographic information on which skilled visual word recognition is founded.

Share (1999) presents what he sees as an alternative or at least orthogonal take on word learning in beginning readers. For Share, the context for analysis is the independent reader and their ability to learn about print while they read. After discussing a number of alternative models for how a student could solve an unknown word (contextual guessing, direct instruction) Share argues for the primary role of phonological analysis and its ability to provide insight into the wide range of words students encounter in text. Share's take on the exact process, however, focuses more on the phonological recoding and the role that plays in helping students both to recognize a word know orally and to disambiguate unclear phonetic information. Share provided evidence that in the absence of phonological recoding (i.e. if students are allowed to focus only on the visual letter information) students learned significantly fewer

words across a variety of word learning measures. Share argues for a significant role in utilizing contextual information to help determine "exact word pronunciations on the basis of partial decoding." Share sees beginning readers as beginning self-teachers, even when their decoding skills are not completely sufficient for a specific text.

While the topic of word learning has been a historically well researched theme, these three perspectives represent highly overlapping yet significantly independent takes on both the history of theories of word learning and the current state of the literature. All three come to almost identical conclusions over the basic, fundamental elements of the word learning system for learning words while reading text:

1. The student must be able to use phonological information to successfully decode the word in question (or at least partially successfully).
2. The student recodes the word phonetically (either out loud or inside their head if reading silently)
3. The student recognizes the word and builds a connection between the word they know and the orthographic form of the word.
4. Through repeated encounters with the word in text, the decoding/word recognition process becomes more automatic.
5. The word eventually can be read automatically and becomes a known word.

Clarifications and Questions

While both Ehri (2005) and Adams (2009) use the name "sight words" to refer to words that students have learned as a result of decoding in text, this use is often confusing. Ehri (2005) presents a thoughtful argument in defense of this label, referencing the way in which experienced readers recognize words instantly, at a glance, as single units. Share (1999) argues that the use of the label sight words creates ambiguity over the manner in which the word was learned (i.e. that the word was memorized visually)

and doesn't clearly refer to the final state of the knowledge. In this paper I will reserve the label sight words for the set of phonetically irregular, highly common words that are memorized visually by beginning readers. Words that have been learned through repeated successful decoding will be referred to as known words.

It's difficult to discuss the literature on word learning without at least mentioning the issue of the number of repetitions required for the trace of a word to become established in memory. Share (1999) directly references this question and presents evidence that four encounters with a word were sufficient to establish the word in memory. While Adams (2009) argues that even a single exposure provides useful and lasting information, Hiebert (2005) argues that all words are not equivalent and cites evidence from Metsala (1999) and Laing & Hume (1999) that specific factors of individual words such as length, decodability, imageability and frequency all affect the rate at which a student can learn a word. Additionally, the question of number of repetitions is significantly impacted both by the depth or shallowness of the orthography of the language in question (Share, 2008) and the level of phonological skill of the reader (Ehri, 2005; Share, 1999). While it's not part of the purpose of this project to weigh in on these issues, the simple fact that most researchers argue for strong visual impressions from a handful of encounters is an issue that can have both positive as well as negative effects—an issue I'll be addressing shortly.

Alternatives to the simple theory of word learning

There are two primary alternative models for word learning present in the research literature. The first alternative, visual word learning, dominated the research agenda for much of the 20th century (Pearson, 2002). Given that a young children with little to no knowledge of phonics or decoding can recognize and

print their own name, there is clearly a mechanism in place where people can recognize and even reproduce words apart from any significant alphabetic knowledge. It has been well documented that in the specific situation of phonetically irregular high frequency words such as *the* or *was* that students form a visual picture of the words and recognize them as units (Torgesen 2002; Ehri, 2005). There is little debate that these "sight words" are learned primarily visually, but, as Ehri (2005) describes, this model is inadequate to explain the learning taking place as students develop the ability to read more and more complex texts automatically. No one argues that a certain percentage of words are learned visually, the only debate (primarily a historical one at this point) was whether or not the visual model of word learning could account for the majority of students' word learning). A typical 3rd grade student can recognize thousands of words automatically and there is no evidence that these words are memorized as visual images. On the contrary, there is significant evidence that these words, while recognized as units and able to be read as such are recognized as comprised of phonetic elements that map to their phonetic recoding (Adams, 1990; Ehri, 1998).

A second alternative, primarily proposed by Goswami (1990) proposes that students learn new words by making analogies between previously learned words and words with similar spelling patterns they encounter in text. While there is significant argument concerning the validity of this model of word learning (Nation, Allen & Hulme, 2001), Adams (2009), Ehri (2005) and Share (1999) all describe the ways in which readers progress in their ability to process increasingly larger chunks of letters rather than having to process each letter individually. A skilled reader would process the letters *i, n, g* for example, not as individual units but as the chunk "-ing". Especially in deep orthographies such as English, a significant part of the task of becoming an expert reader is learning the variety of spelling patterns and their associated phonetic recodings (Share, 2005). For the sake of the arguments presented in this

paper, it is reasonable to consider that decoding and learning new words based on recognizing previously learned spelling patterns, whether recognized as part of word analogies or as independent chunks, as equivalent examples of what the simple view of word learning already proscribes and thus doesn't represent an alternative view of word learning so much as a minor variation in the order of the steps or the focus of attention.

To help clarify this point, consider when a student with sufficient decoding skills encounters the word *hike* in a text. They don't typically sound out each individual letter, h-i-k-e. According to the simple theory of word learning, they would, assuming they have the requisite phonics knowledge, recognize the vowel-> consonant-> e at the end spelling pattern and recognize that the "e" is silent and the "i" is long. Putting the "h" sound in the front, they would generate hike. Alternately, as per Goswami's analogy theory, it could be they recognize that hike is spelled similarly to bike and like and use that information to treat the –ike rime as a chunk to be used in identifying the word hike. In either case, the word hike has been accurately de-coded (using either method), phonetically re-encoded, the child produces the word hike and the process of automatic and fluent word reading has begun.

Critical role of accuracy in supporting word learning:

A critical constraint of the simple theory of word learning is that students must decode the word accurately when they encounter it in text. Adams (2009) elaborates on her description of the word learning process in order to clarify this point:

> Importantly, however, acquiring new sight words is the direct outcome of neither careful instruction in phonics, phonemic awareness, or letters, nor even of prior decoding sophistication. All of those factors are but enablers. Rather, the prepotent determinant of sight

> word acquisition is whether, on encountering a new word in print, the student actually does try and does succeed in decoding it. (pg. 35, Finding the Right Texts, emphasis mine)

While Adams (2009) and Ehri (2005) each talk about the importance of successful decoding in supporting word learning, there exists little research looking at the effect that inaccurate reading has on the word learning process and the degree to which student decoding errors during independent reading undermine or potentially negate the positive effects typically associated with decoding practice.

Torgesen (2002) briefly touches on this topic, noting that while successful reading leads to fluent word identification, inaccurate reading of a text actually undermines word learning outcomes. Citing the frequent difficulty struggling readers have with grade level texts that are often too difficult for them, he argues that since automatic word learning is dependent on multiple successful trials, inaccurate reading slows the growth of word learning. Along this same line, data from Share (1999) demonstrates the way in which students who made decoding errors during reading tended to make errors on a post-test word identification task in line with their decoding errors. When offered alternative spellings for a target word they had read, students tended to select spellings that matched what they had incorrectly said out loud rather than the correct choice they had seen and attempted to decode. While there is strong agreement and evidence that successful decoding is critical in supporting word learning and there is some mention of the idea that errors during text reading undermine word learning, Cunningham (2006) was one of the first and clearest studies to demonstrate this effect.

Cunningham (2006)

Cunningham (2006) 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 solving

unknown words (such as contextual clues) they will be less likely to attend to the orthographic features of words and thus less likely to form a visual representation of a word's spelling and less likely to learn the word orthographically. She presented students with a set of 16 target words, selected for their likelihood to be known orally but unlikelihood 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 solve 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 piece of data 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 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) went down. 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 correctly across any of the word learning measures. In order to make as level of a playing field in testing the validity of her primary hypothesis, Cunningham only compared word learning outcomes 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.

What does this have to do with decodable texts?

Given this foundation based on an analysis of the research literature about the theory of word learning and the role of accuracy in supporting word learning outcomes, the issue now turns to the question: what does this have to do with decodable texts?

While texts based on the phonetic regularity of the English Language have been used for years (see Pearson, 2002 for a recent history), the current definition of decodability, as mandated by states such as California and Texas is that texts for beginning readers require a substantial majority of words to be decodable as defined by being made up of letters and spelling patterns that have been previously taught in the classroom (a concept referred to as lesson-to-text-matching) (Hiebert et al, 2010). This definition of decodability represents a design strategy for texts used in beginning reading instruction that is often called single-criterion (Mesmer, 2010). Previous text design considerations such as vocabulary control, sentence length, word repetition and word frequency are not controlled in decodable texts (Hiebert, 2009). The overall concern of text difficulty is replaced by a focus on lesson-to-text-matching since all texts created by such a design strategy should be manageable by students who have been taught the

appropriate phonics information and sight words, seemingly independent of earlier theories of text difficulty (Pearson & Hiebert, 2010; Hiebert 2005).

Over the last ten years, two of the most frequently used commercial curriculum packages across the country have been Houghton Mifflin Reading and SRA/McGraw Hill's Open Court Reading. 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 decodables modeled after those of Open Court, 2000 (Hiebert, online source, 2010). These packages are used widely in public schools across the United States and students are taught to read using decodable texts designed according to the principle of lesson-to-text-matching. Both Open Court and Houghton Mifflin contain large numbers of decodable texts that students are given in class, supported in their initial readings and then directed to take home and reread independently. And while both programs have been certified at all levels as researched based instructional practices, there is no existing data isolating the accuracy of student's reading of these decodable texts.

While a number of researchers looking at 1st grade reading development have assessed early reading success with a variety of earlier basal series (Hoffman, Sailors & Patterson, 2002; Mesmer 2001), the overwhelming majority of recent data on 1st grade student reading behavior in text comes from widely used measures such as the DIBELS oral reading fluency assessment. While the DIBELS data is highly informative and researchers have done significant work in analyzing what it tells us (Hiebert et al, 2010), the design criteria for the DIBELS passages are radically different than the design criteria for the decodable texts students read every day at school. Additionally, most school districts do not even use the accuracy data contained in the DIBELS assessment. During reading of the oral reading fluency

passages, students are scored for each error they make. That data, however, is only used as a modifier for calculating the fluency score and no data on student accuracy is reported. This same pattern has been true for most curriculum based assessments—that student accuracy data is only calculated so as to modify the fluency score such that it represents a reliable measure of words read accurately in a given time period. While the data exists in the individual student testing booklets, it is never scored or reported on independently.

Mesmer (2010) highlights some of the difficulties in assessing student reading performance with decodable texts. Mesmer worked with 74 1st grade students, having them read an appropriately leveled decodable text as well as an appropriately leveled qualitatively-leveled text. Students read the same two books at three points during the year to assess development in accuracy and fluency. In order to determine an appropriate decodable text, texts were selected based on phonics information for which students demonstrated mastery. 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). 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 that specific decodable text than a typical struggling reader. Even with the supportive definition of appropriately decodable text, in two of the four reported readings, students' accuracy scores were below 90% with fluency rates averaging under 30 words/minute. And in relation to

Mesmer's primary research question, the data were inconclusive whether decodable or qualitatively-leveled texts better supported student accuracy during reading.

The paucity of research looking at the specific effects of text on student reading behaviors and learning outcomes and the specific challenges students face in working with a significantly new class of texts leads to a gap in the research literature. Juel & Roper-Schneider (1985) demonstrated clearly that the type of texts used in early reading instruction can have a significant effect on the reading behaviors and learning outcomes of students trained by them. According to that logic, it is incumbent on the research community to study the effects of students' daily, instruction decodable texts on beginning readers, to see what outcomes might be a result of the interaction. It is in order to help address this question that the following data collection project was undertaken.

Methods:

In order to assess the performance of students with typical, daily-use decodable texts, data was collected from two 1st grade classrooms in an urban school district in Northern California. Twice per week I met with students from each class and had them read their most recently read decodable text. Each of the texts included in the study had been read by the students with their classroom teacher together as a whole class with each teacher providing support focusing on identifying the specific phonics information highlighted in the text. Each student had also read the book once more independently at their desk prior to reading with me. Data for any student who had been absent during a reading of that specific text was not recorded. Due to scheduling difficulties, field trips, testing and logistical difficulties, the sequence of books from which data has been collected does not follow an exact pattern. The total number of books as of this preliminary analysis, however (n = 18), should provide a

large enough sample from which to analyze patterns in student performance and guard against specific issues arising from individual books.

On a given day I would show up to class, 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 and then have them read it out loud to me. I kept a detailed record of the sessions and scored each student's reading for accuracy and fluency, also marking not just if a student made an error, but what they actually produced as they read a specific word. In situations where I was unable to identify what a specific student had said, the response was marked simply as an error. 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 and 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 and 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 for fluency. 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. This data was kept separate from the primary data used for accuracy and fluency calculations but was analyzed in a search for patterns of text features that correlated with patterns of student errors.

Participants:

The 47 first grade students at Lockwood Elementary School (pseudonym) were 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) 16% 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 top 80% of California schools.

Students came from two reasonably similar 1st grade classes, one taught by a teacher with 14 years of experience and the other with five years of experience. Both teachers implemented the Open Court reading program with similar fidelity and had students read their decodable book of the day more than 90% of the time. Both teachers followed proscribed lesson plans in introducing new books to students and had roughly similar strategies for reading the book together as a class and then having students read the book independently, with help if necessary. Both teachers also allowed for a small number of their advanced readers (n = 11 between both classes) to skip the book instruction, read the book independently at their desk and then have time for free independent reading. All students from both classes 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 150 decodable texts that teachers sequentially give out to students on a daily basis. The decodable texts are used from the beginning of instruction until roughly 2/3 of the way through the school year at which point literature anthologies become the primary instructional materials for class and independent reading. As stated previously, this preliminary data set was collected from a subset of the books read by students representing 18 titles across four months of data collection, starting near the end of November and continuing through April. For a complete list of all books used for data collection, see the appendix.

Results

In relation to the specific research question of how accurate are 1st grade students when reading daily instructional decodable texts, there were three significant findings from the preliminary data analysis. Borrowing a framework from Betts (1946) the DIBELS assessment literature describes three classifications of texts in terms of their difficulty: frustrational texts that students read at below 94% accuracy, instructional texts that students can read at 94% accuracy or better and independent texts that students can read at 97% accuracy or better (DIBELS website, modified from Hasbrouk, 1998). The calculation for reading accuracy was done by 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). While the reading accuracy score hides the absolute number of errors made by a given student, it gives a more informative picture than simple number of errors/book by taking into account how much text the student was able to process and the ratio of accurate to inaccurate reading the student was able to produce.

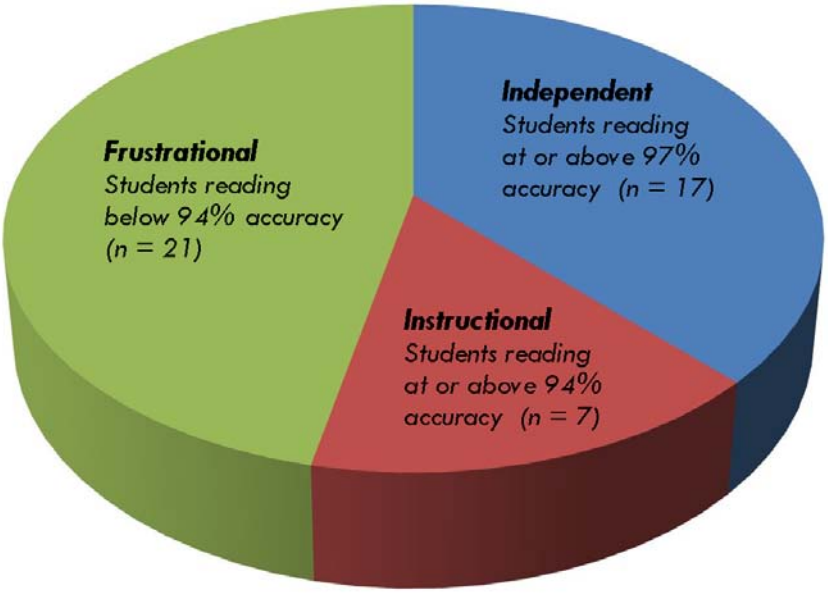

Figure 1 Number of students broken down by average accuracy scores across all 18 texts

Figure 1 shows the number of students for whom their average accuracy score across all books used in this preliminary data analysis (n = 18 books) falls into the different categories. Table 2 breaks the data down in a slightly different way, showing how many individual books students read below 94% accuracy.

Number of books read at frustrational level (< 94% accuracy) out of 18 total books read	Number of Students
0	12
1 to 3	8
4 to 6	5
7 to 9	5
10 to 12	5
13 to 15	7
16 to 18	3

Table 2 Number of students reading varying number of books at below 94% accuracy

Data from the beginning of the year DIBELS assessment was used to group students into terciles based on their composite performance. The top 3rd of students averaged 98.7% accuracy across the 18 decodable texts, while the middle third students averaged 93.8% accuracy and the bottom third averaged 85.7% accuracy. Of all the texts read by all of the students in the top 3rd, 9% of the readings were read below 94% accuracy. For the middle third of students, 40% of their readings were below 94% accuracy and for the bottom third of students, 75% if their readings were below 94% accuracy. For the top third of students, 80% of their readings were at the independent level, at or above 97% accuracy. For the middle third of students, 41% of their readings were accurate at or above 97% and for the lowest third, 10% of their readings were 97% accurate or better.

A second finding pertained to the nature of the mistakes made by students. While this aspect of the data analysis is in its early stages and a more detailed analysis will be the subject of a project currently in process, the overwhelming majority of mistakes made by students (> 85%) occurred on words that were either previously taught sight words or decodable words made up of phonics patterns that had been previously taught. For the lowest third group of readers, a large percentage of their errors occurred on words containing the most recently taught phonics patterns or recently taught sight words.

The third finding pertains to fluency scores. Table 3 shows the average fluency scores broken down by tercile. Fluency scores give an approximate measure of the degree to which both the sight words and previously taught spelling patterns have become sufficiently internalized to support fluent, automatic word recognition. Fluency scores have also been linked to accuracy (Mesmer, 2010; Fuchs, Fuchs, Hosp & Jenkins, 2001) although the relationship is somewhat unclear.

Student Tercile	Fluency Score	σ
top third	27.3 wpm	7.4
middle third	50.9 wpm	10.6
bottom third	83.4 wpm	14.2

Table 3 Average fluency scores across all 18 books organized by tercile

Discussion

The data collected provides clear evidence that struggling beginning readers have difficulty reading their daily, instructional decodable texts accurately even with limited support (i.e. my telling a student an unknown word after three seconds). The data additionally show that struggling readers make errors on words containing target instructional spelling patterns as well as on previously instructed sight words. The data also showed fluency levels that indicate students are having substantial difficulty making it through these texts. The readings for this data collection represent at least the third time the book was read by a student. The fact that most of the struggling beginning readers had fluency rates averaging under 30 words/minute is a significant concern. Both sight words and decodable words are not being read automatically, even after three repetitions with a specific book and daily practice across the entire series.

One significant methodological issue amplifies the strength of these findings especially for struggling readers--the research decision to employ the assessment technique of telling students an unknown word after three seconds. While the decision was made both in light of the data collection methodology employed with the DIBELS assessment and in order to support struggling readers, as Share (1999) points

out, this is not the situation faced by a student during independent reading. Since the books are given to students to take home, they are unlikely to receive any additional support during their attempted independent readings, a situation that is more likely true for the lowest performing students. The absence of this support would significantly impact student reading performance, both increasing the number of errors they make and decreasing the speed at which they are able to read. The average number of tells/book for the lowest third of students was 4.3, while only 1.2 for the middle third of students and zero for the top third of students. For many of the struggling readers with accuracy scores well under 80%, the assistance provided by these "tells" enabled them to continue moving through a text that otherwise would have been impossible. Telling a student an unknown word often provided sufficient context for them to solve a subsequent word using contextual cues that they were unable to solve through decoding at other places in the text. For a student reading at roughly 30 words/minute and repeatedly encountering words they don't recognize and know they can't solve, one has to wonder about the likelihood of this student persevering to read an entire text and the actual benefit (or potential harm) of these attempts at independent reading.

Different Outcomes by Tercile

The highest performing readers read through all of the books quickly and effortlessly. Their accuracy and fluency scores show them having attained a reading level beyond the difficulty of the decodable texts, and given their success at the beginning of the data collection and their scores on the beginning of the year DIBELS assessment, one could argue that they began 1st grade reading at a level beyond that of the decodable texts.

The story for the middle third of readers is slightly more complex. While their fluency scores and DIBELS scores show them to be making satisfactory progress and at low-risk of missing benchmark standards for the end of the year assessment, their accuracy scores reading the daily, instructional decodable texts are problematic. 40% of the readings from students in the middle third were below 94% accurate and their average accuracy score summed across all the books was 93.8%. These students are clearly making progress and in most cases, showing improvement over time in regards to the errors they made on earlier books. While the data analysis to fully understand, let alone explain this finding is the work of another project, the question of why low accuracy scores don't seem to have as significant of an effect on the learning outcomes of the middle third of students is an important question.

The picture is clearest and most problematic for the lowest third of students. The DIBELS scores for the bottom third show them to be in the 30^{th} – 40^{th} percentile ranks across national averages in their initial assessments of letter naming fluency, phoneme segmentation fluency, nonsense word fluency and oral reading fluency (from the mid-year assessment). Pinebranch Elementary school is in the 80^{th} percentile of schools in California, so the results from this data collection are in no way due to extremely low performing students. The data for 1^{st} grade students in the lowest tercile at Pinebranch Elementary mirrors data from students performing roughly in the bottom third of students across the nation.

The clearest take home point from this project is that the data shows that the daily, instructional decodable texts are too difficult for students in the lowest-third of reading performance. Even by their third reading, their average accuracy scores were 85.7%, and 75% of the time they encountered a book, it was at their frustrational level for reading. The lowest third of students made the overwhelming majority of their errors on the sight words and decodable spelling patterns that the books are designed

to support/teach. Given the mandate that the books be 75% decodable according to the rules of lesson-to-text-matching it's no surprise that student errors are on these words. But the evidence from these errors indicates that the books aren't meeting the essential criteria of the simple view of word learning—that the students are successful in their attempts to decode words en route to having them become automatic, easily recognized and part of their pool of known words.

An average fluency score of 27.4 words/minute for the lowest third of students serves as an additional perspective on this difficult picture. For many of the lowest-performing students, averaging over four tells per book means that they were enabled to keep going in places where otherwise they were stuck, with insufficient resources to successfully solve an unknown word. Fluency scores were hugely impacted by the methodological decision to support readers struggling with unknown words. Many of the lowest students simply gave up attempting to solve an unknown word and either directly asked me what the word was or waited for me to tell them. Some students in the bottom and middle tercile demonstrated an alternative to getting stuck or waiting to be told, they simply made their best attempt and kept going—whether or not what they read matched what was written or made any sense. While this issue of patterns in behavior of what students do when presented with words they can't solve is the topic of ongoing analysis, the most salient point from Juel & Roper-Schnieder (1985), that the type of books students use in learning to read will significantly impact their reading behaviors and learning outcomes is clearly evident in this data collection. Asking students to read texts on a daily basis that are too hard for them produces high error rates, low fluency rates and leads to development of the exact opposite types of reading behaviors and learning outcomes that the texts are designed to support.

Conclusion and ongoing questions

The critical next question for future research is to explain these findings. The two primary candidate explanations would focus in two distinctly different directions—either looking at the students as struggling beginning readers who would struggle no matter what they were reading or looking at the nature of the texts being used for beginning reading instruction and attempting to identify text features that potentially undermines students' ability to be successful (Lipson & Wixson, 1986). While studies such as Cunningham (2006) and Share (1999) have found strong evidence supporting relationships between word learning outcomes and individual students orthographic/phonetic skills (e.g. rapid alphabetic naming, Woodcock word attack, etc.), in both of those studies texts were carefully controlled for student reading ability. There is a clear and compelling case that has been made connecting student orthographic knowledge and student reading performance, but text difficulty is the trump card, so to speak, in this relationship. Give any student a book that is made up entirely of material which they have completely mastered and their learning outcomes will be significantly higher than if the book they were reading was at a difficulty level where they were frequently inaccurate and struggled to make it through. So while the ongoing research effort to understand the relationship between individual student skills and reading performance is a valuable endeavor, the specific situation of students reading books that are beyond their reading level changes the fundamental equation.

Based on the data from this study and looking across recent research looking at text features of decodable texts, there is strong reason to believe that current decodable texts, while strong on their integration with classroom instruction and abundant in their support of opportunities to decode suffer from a number of design flaws that undermine their ability to be read successfully by beginning readers, especially struggling 1st grade readers. The primary design theory supporting successful reading of daily,

instructional decodable texts is lesson-to-text-matching and scientifically-based classroom instruction. As has been pointed out by a number of researchers (Mesmer & Cumming, 2009; Pearson & Hiebert, 2010; Allington, 1998, 2005) by exchanging previous strategies for controlling text difficulty for the single-criterion of lesson-to-text-matching, text designers have adopted a strategy with no empirical evidence supporting its applicability. Additionally, the model of lesson-to-text-matching harkens back to the most basic and traditional frameworks of transmission models of teaching. The simple belief that because a teacher teaches something therefore students have learned it and are capable of using it successfully almost immediately is a frequently reoccurring myth in education. JJuel & Roper-Schnieder (1985) clearly demonstrated that beginning reading texts needed to support students in practicing what they had been taught in class. The inverse, however, that because something had been taught in class it therefore works as a single design criteria for creating a text that students can read successfully, does not logically follow.

Apart from any other concerns about vocabulary control, word frequency, word repetitions, awkward language use and structure, each of which significantly impact the difficulty of a text (Hiebert & Mesmer, 2006), the simple issue of pacing in lesson-to-text-matching demonstrates the Achilles heel with this strategy for controlling text difficulty. As Hiebert (2009) addressed, the pace of introduction of new material, both sight words and phonics information, has been accelerated significantly over the last 20 years. Much of what used to be taught in 1st grade has been pushed down to kindergarten and 1st grade standards have been expanded with the goal of better preparing students for nationally mandated tests at the end of 2nd grade. But imagine a simple thought experiment. At what point does instructed not mean learned? At what point does the pace of introducing new material get so difficult and so overwhelming that it is developmentally inappropriate and unrealistic to assume 1st grade students can

learn at that pace? Clearly there's a point where instructed can't mean learned, where lesson-to-text-matching relies on learning from lessons that are introducing new material at a pace that cannot be achieved. So the fundamental design theory of decodable texts and the primary tenet of state mandates regarding the design of early reading materials relies not just on an idealized transmission model of teaching, but on a set of assumptions about pacing that are proving to be consistently unrealistic.

The evidence from school districts across the country is resoundingly clear—many students are not learning to read by the end of 1st grade. While this is obviously an issue far larger than the scope of this project and it again raises the question of the nature of the problem, the instructional program vs. the attributes and behaviors of students, at what point does the NCLB mandate for evidence based instructional practices ever start to filter backwards? How many years of similar outcomes where the lowest 30 - 40% or so of readers not meeting national benchmarks becomes evidence that the current models of instruction aren't working? Data such as those presented in this research project will hopefully help both the research community and those charged with mandating curricular standards to at least reconsider the text design criteria of the daily, instructional decodable texts and their role in early reading instruction.

Appendix: Titles of books read by students

Jen's Pen
Zack the One Man Band
Chuck's Chest
A Blur With Fur
Kim's Trip
The Stand
The Cold Troll
At the Vet
A Gift for Me
Steve Sells Vans
The Bee and the Deer
Craig Sails
Dean's Pies
The Farmer and the Doe
Mew, Mew & Rescue the Cat
Max the Grouch
The Knight Who Didn't Know
The Choice

References

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

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

Allington, R. L., & Woodside-Jiron, H. (1998). Decodable Text in Beginning Reading: Are Mandates and Policy Based on Research? In R. L. Allington (Ed.), *Big Brother and the National Reading Curriculum: How Ideology Trumped Evidence* (pp. 195-216). Portsmouth, NH: Heinemann.

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

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

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

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.

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.

Goswami, U. (1990). Phonological Priming and Orthographic Analogies in Reading. *Journal of Experimental Child Psychology,* 49, 323-340.

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

Hiebert, E.H. (2008). 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 for beginning and struggling readers: Research-based solutions (pp. 1-21). NY: Guilford.

Hiebert, E.H., Stewart, J.D. & Uzicanin, M. (2010) A Comparison of Word Features Affecting Word Recognition of At-Risk Beginning Readers and Their Peers, in-press

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.

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

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

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., & Cumming, S. (2009). Text-Reader Matching: Meeting the Needs of Struggling Readers. In E. H. Hiebert & M. Sailors (Eds.), *Finding the Right Texts: What Works for Beginning and Struggling Readers* (pp. 149-176). New York: Guilford Press.

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., Allen, R., & Hulme, C. (2001). The Limitations of Orthographic Analogy in Early Reading Development: Performance on the Clue-Word Task Depends on Phonological Priming and Elementary Decoding Skill, Not the Use of Orthographic Analogy. *Journal of Experimental Child Psychology*, 75-94.

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.

Pearson, P.D., Hiebert, E.H. (2010). National Reports in Literacy: Building a Scientific Base for Practice and Policy. *Educational Researcher,* 10(10), 1-9.

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

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