Features of Known and Unknown Words for First Graders of Different Proficiency Levels in Winter and Spring

Elfrieda H. Hiebert [1,*] **, Yukie Toyama** [2] **and Robin Irey** [3]

[1]   TextProject, Santa Cruz, CA 95060, USA
[2]   Graduate School of Education, University of California, Berkeley, CA 94720, USA; yukie.toyama@berkeley.edu
[3]   Graduate School of Education, Stanford University, Stanford, CA 94305, USA; robin.irey@gmail.com
[*]   Correspondence: hiebert@textproject.org

check for updates

Abstract: This study describes the features of words known and unknown by first graders of different proficiency levels in six instances of an oral reading fluency assessment: three in winter and three in spring. A sample of 411 students was placed into four groups (very high, high, middle, and low) based on their median correct words per minute in spring. Each word in the assessment was coded on 11 features: numbers of phonemes, letters, syllables, blends, morphemes, percentages of multisyllabic and of morphologically complex words, concreteness, age of acquisition, decodability, and U function. Words were classified as known if more than 50% of the students within a group were able to correctly read those words. Features of known and unknown words were contrasted for all but the highest group, which made no errors, at each point in time. An analysis of the patterns of known words across groups from winter to spring shows that students followed a similar general progression in the number and type of words recognized. The most prominent feature of unknown words in winter and spring for the middle group of students was the presence of multiple syllables. The lowest-performing group of students continued to be limited by word length and frequency in their recognition of words, but on both features, their proficiency increased from winter to spring. The discussion addresses several critical issues, most notably the relationship of words in oral reading assessments to the word recognition curriculum of many beginning reading programs.

Keywords: word recognition development; assessment texts

1. Introduction

This study is an examination of the words known and unknown by first graders of different proficiency levels in winter and spring on a widely used oral reading fluency assessment. The underlying aim of this work is to support theory building and research as well as instruction and intervention by adding to the field's understanding of word recognition in the context of typical texts. While research on student and word factors that influence word recognition is substantial [1–3], studies often use word lists or carefully constructed texts that control words, sometimes even through the use of nonsense words [4]. However, typical texts can be expected to have a variety of words, some orthographically regular and others not. Even within texts written to emphasize specific grapheme–phoneme correspondences (GPCs), students encounter numerous words with irregular correspondences [5]. After all, among the 100 unique words that account for 48% of the total words in texts [6,7], 29 have vowel patterns with an irregular relationship between graphemes and phonemes [8]. Reading multisyllabic words is often viewed as a skill that follows consolidated alphabetic knowledge [9], despite the presence of

high-frequency multisyllabic words (e.g., *over*) and words with inflected endings (e.g., *going*) in early texts. Neither the match between word recognition proficiencies required by the texts of evaluation and instruction nor students' knowledge at particular points in development has been established.

In theoretical perspectives of word recognition, at least a modicum of repetition with words or word patterns is assumed to serve as a mechanism for development of alphabetic knowledge [9]. A substantial change occurred in perspectives on beginning texts, however, in the early 1990s, when large states mandated use of authentic texts rather than texts with controlled vocabulary [10]. As Foorman et al. [11] showed in their analysis of first-grade reading programs published after this change, even texts designed to promote decodable words have a variety of word types. To date, empirical investigations have not documented the words that students recognize in texts at different points in their learning progression. For models of word recognition to be comprehensive and for empirical research to extend knowledge of word features, students' recognition of words in the context of typical texts used for assessment and instruction requires attention.

In addition to contributing to theory on the features of words that influence the progression of word recognition, information on known and unknown words is fundamental to the creation of appropriate instruction and intervention. Students' oral reading fluency, as measured by the quantity of words recognized when reading text, has been influential when making decisions about tiers of instruction and intervention [12,13]. Summaries of these assessments, however, fail to provide information on the types of words students know at specific points in development. Without information on what students know, some students may be placed into interventions driven by pedagogical assumptions about progression rather than on evidence of the kinds of words that students can and cannot recognize.

We review two areas of literature for this investigation of known and difficult words for students of different proficiency levels at two points during first grade. To analyze words that are difficult or known by young readers requires an understanding of the features of words that influence word recognition. To provide this background, we review critical variables that have been shown to influence word recognition acquisition. Further, since evaluations of readers' word recognition are invariably based on the content of assessments, we review a second area of literature—the nature of words in assessment and instructional texts.

2. Word Features that Influence Word Recognition

Numerous variables have been shown to predict young children's early reading acquisition, including short-term memory, ability to segment and blend phonemes, letter-name knowledge, concepts of print, and vocabulary [14,15]. Instructional activities that enhance children's preparedness for reading have typically been the purview of preschool and kindergarten, although evidence points to increased expectations for children to begin recognizing words in kindergarten [16]. The curricula of current core reading programs indicate that lessons on recognizing words are prominent at the end of kindergarten and definitely at the beginning of first grade. For example, the second lesson of the first-grade component of a widely used program identifies *new*, *try*, *great*, *enjoy*, *excited*, and *nervous* as focus words [17]. These words are taught through lessons and activities, including reading a text. This study focuses on the features of words that are related to children's independent recognition of text, not on variables that may have direct or indirect effects on children's word recognition overall.

Hiebert et al. [18] found that two of nine variables—frequency of a word's appearance in written English and age of acquisition—predicted second through twelfth graders' recognition of word meanings. For students who were English Language Learners, word length, number of syllables, and concreteness also predicted knowledge of word meanings. None of these measures directly addressed the orthography of words, measures that are essential when considering young readers' recognition of words in oral reading tasks. For that reason, we used the four constructs that were identified in a recent analysis of the features of words in first- and third-grade texts: orthography, length, familiarity, and morphology [19]. The research on each of these variables and their effects on

word recognition is extensive [1–3]. Our aim is to provide a succinct summary of research on these four constructs, which serve as the foundation for this study.

2.1. Orthography of Words

Consonant clusters in which several graphemes represent a single phoneme affect the speed of word recognition [20], but it is the vowel patterns of English where variability in GPCs is greatest and with which beginning readers are most challenged [21]. A single phoneme for a vowel can be represented by unique graphemes as well as multiple graphemes. Further, the consistency and number of words that share GPCs for vowels can influence the speed of recognition [22].

Among the many studies focused on word recognition [1–3], information on the progression of students' ability to recognize words as a function of the complexity of the vowel GPCs has been limited. One of the few studies of students' progression in proficiency with vowel GPCs across the reading acquisition period was conducted by Guthrie and Seifert [23]. Their sample consisted of typical readers in grades one through three and a comparison sample of fourth and fifth graders who were struggling readers. At three points in time across a school year, students were assessed on a set of five tasks: consonant–vowel combinations (e.g., *da*); short-vowel words, including with consonant blends (e.g., *hat*, *brim*); long-vowel words (e.g., *tape*, *shave*); nonsense words with 50% short vowels, 25% long vowels, and 25% special vowels; and special-rule word production (e.g., *join*, *saw*). Guthrie and Seifert reported that both proficient and struggling readers followed a similar order. This progression from easy to difficult was the following: consonant–vowel, short vowel, long vowel, special rule, and nonsense.

Guthrie and Seifert [23] considered whether the order of proficiency reflected instruction or the complexity of GPCs. The classrooms from which participants came followed a core reading phonics curriculum in which short- and long-vowel words were emphasized in first grade, moving to an emphasis on more complex vowels and special-rule vowels in second grade. Some follow-up of the latter occurred in third grade. Of the words in the first-grade texts, 31% had short vowels, 8% had long vowels, and 9% contained special-rule vowels. The remaining 52% of the words either had rare vowel patterns or were multisyllabic. Guthrie and Seifert concluded that the stages of learning were based on the complexity of the rules to be learned and not an artifact of the curriculum because all GPCs had been taught to both typical and struggling readers.

A second study that validates a progression in vowel patterns of monosyllabic words is that of Pirani-McGurl [24], who conducted a validation study of the Phonics Diagnostic Inventory (PDI). The PDI was administered to 375 students from grades two through five with varying ability levels. The sequence identified by Pirani-McGurl follows a similar progression to that identified by Guthrie and Seifert [23]. However, Pirani-McGurl's analysis is more specific with regard to the order within a given vowel pattern, such as short vowels. For example, whereas words with short vowels were learned before words with long vowels, words with the short /a/ were not learned first, as had been hypothesized. Students performed better with words with short /i/, /e/, and /u/ than they did with words with short /a/.

In both the Guthrie and Seifert [23] and Pirani-McGurl [24] studies, data were not reported separately by grade or proficiency level. The two samples comprised students who ranged from grades two through five. The breadth of this age group makes it unclear when during the learning period students become consistent in their application of specific GPCs. Further, although Guthrie and Seifert included special-rule vowels in their study, neither study addressed patterns that Fry [25] has described as rare—patterns that are irregular and occur in only a small group of words. A sizable number of words that have rare vowel patterns are among the most frequent words in written language (e.g., *of*, *from*). To date, studies of the progression of word recognition have not compared students' facility with words with rare patterns relative to the categories of short vowels, long vowels, and special-rule vowels.

2.2. Word Length

It has been well established that the number of letters in the words of alphabetic languages influences word recognition, especially at the beginning stage of reading acquisition [26,27]. However, it is not clear whether this effect is due to the number of letters or, rather, due to the number of phonemes in words. Numerous words contain graphemes that are not represented by a phoneme (e.g., g in *gnat*), and in other words, two or more graphemes can be pronounced as a single phoneme (e.g., th in *that*).

In a study of Dutch words [28], stimulus words had an equivalent number of graphemes but varied in number of phonemes. They found that naming onset latencies were longer for phonologically longer words as compared to phonologically shorter words. Gagl et al. [20], hypothesizing that the presence of consonant clusters could have contributed to the word length effect reported by Marinus and de Jong [28], contrasted students' performances reading German words in which each grapheme was represented by a single phoneme in a three-letter word (e.g., *Bad*) and words containing four to five letters but of three word types: (a) multiple-letter graphemes represented by a single phoneme (e.g., "sch" in *Fisch*); (b) words containing a cluster, in which every grapheme in a monosyllabic word was represented by a phoneme (e.g., *Prinz*); and (c) two-syllable words with a single grapheme to phoneme match (e.g., *Salat*). They found that, whereas phoneme and letter length contributed to the length effect in naming latencies, words with consonant clusters elicited the largest length effect. Thus, the role of the number of letters and phonemes in beginning readers' word recognition warrants further attention.

The influence of multiple syllables on the speed with which words are processed is widely recognized in the literature [29]. Most theories of reading development, as Seymour [30] observes, suggest that attention to syllables is needed only after the consolidation of knowledge of smaller units [31]. At least some multisyllabic words, however, can be expected to appear in beginning texts. Of the 100 most frequent words that account for 48% of the total words in texts [7], 12% of the words are multisyllabic. Among the 200 words that account for an additional 10% of the words in texts, 35% are multisyllabic. The prolific nature of compounding among words with Anglo-Saxon origins [32] means that compound words with at least two syllables (e.g., classroom) can be expected in beginning texts as well. This observation is confirmed by Masterson et al. [33], who reported that multisyllabic words accounted for 38% of word types in United Kingdom (U.K.) schoolbooks at the reception level (equivalent to kindergarten in the United States (U.S.)).

2.3. Word Familiarity

Familiarity refers to students' prior experience with words and their associated concepts [34]. The rapid and automatic word recognition of skilled readers relies on high-quality lexical representations [35,36]. High-quality lexical representations require that readers have had at least some previous exposure to words. Semantic knowledge has been shown to influence recognition of both regular and irregular words in English [37].

Often, frequency in written language has been used as an indicator of exposure or familiarity [38]. Frequency can be a nebulous variable because it is confounded with features such as polysemy and length. Even so, in that automaticity with words is viewed as a function of exposure [9], the influence of frequency of appearance in written language merits attention in establishing the words beginning readers are familiar with in the context of texts. This variable may assume an especially critical role in the texts currently used for beginning reading instruction in that repetition of words does not appear to be a criterion for text selection and design [11,39].

At the same time, the age at which words enter students' oral language also requires consideration. For example, young children who have never heard the words *yawl* and *brawl* in oral language are likely to recognize these words more slowly in a text than words used in their oral language, such as *draw* and *yawn*. Age of acquisition (AoA), the age at which children first learn a word in their oral language, has been found to be a strong predictor of how quickly words are processed [40,41].

Another variable that influences semantic access and, therefore, merits attention is the concreteness/abstractness of a given word. The term concreteness is often used interchangeably with imageability in that highly concrete words elicit mental images more rapidly than abstract words. Concreteness of words has been shown to influence the speed with which words are processed [42,43]. More recently, Steacy and Compton [44] found that imageability played a role in irregular word learning and that students who received imageability training required fewer exposures to reach mastery.

2.4. Morphology

GPCs are frequently viewed as a priority in beginning reading instruction in contrast to morphology, which is the domain of the middle to upper grades [32,45]. However, the texts used for beginning reading instruction contain numerous words with inflected endings and affixes. Among the 1000 most frequent words that populate beginning reading texts [33,46], approximately 20% are words with inflected endings (e.g., *wearing*), simple suffixes (e.g., *helpful*), compounds of two base words (e.g., *classroom*), and complex suffixes (e.g., *attention*). Indeed, even at first- and second-grade levels, morphological awareness has been shown to predict both word reading and reading comprehension [47,48]. Therefore, research focused on reading development needs to consider recognition of words with multiple morphemes in texts used to assess the proficiency levels of beginning readers.

3. Features of Words in Beginning Assessment and Instructional Texts

In an overview of a special journal issue on curriculum-based measurement (CBM), Petscher et al. [49] concluded that reliability estimates of oral reading fluency (ORF) measures are uniformly high. In contrast, the validity of measures of ORF, specifically standardized ones (e.g., DIBELS, AIMSweb), has been infrequently addressed. Given that standardized ORF measures emerged from the CBM model [50], they have frequently been assumed to represent the typical beginning reading curriculum. Reviews of the beginning reading curriculum are few, but research such as that of Pirani-McGurl [24] leads to the expectation that assessments of young readers should focus on monosyllabic words with common and consistent GPCs.

Available information on the development of texts on CBM ORF assessments challenges the assumption that these measures represent the typical beginning reading curriculum. Texts on the widely used DIBELS ORF assessments, according to Powell-Smith et al. [51], are based on three criteria: word complexity (as measured by number of characters and syllables within words, as well as percentage of words with three or more syllables and of words with seven or more characters), semantic difficulty (number of rare words), and syntactic difficulty (median words per sentence).

Words with seven or more characters are undoubtedly more difficult for young students to read than words with two to four letters, primarily because seven-character words are likely to be multisyllabic. An emphasis on multisyllabic words in the measurement of word complexity of a first-grade assessment is unexpected based on existing research on word recognition. As the review of research indicated, theoretical models typically emphasize proficiency with monosyllabic words in the beginning phases of reading acquisition [29,52]. For the DIBELS text design model, neither number nor complexity of GPCs is mentioned as part of the measurement of word complexity.

Our review found no studies of the word recognition curricula that underlie current programs used for beginning reading instruction in English-speaking countries to verify that the DIBELS model captures the enacted curriculum. There are, however, studies that provide descriptions of the word-recognition task required by typical instructional texts. These studies explicitly address the distributions of high-frequency words and the GPCs often viewed as critical to students' acquisition of proficient word recognition [52]. We include a brief review of these studies, as they provide comparative lenses from which to view the text difficulty of an ORF assessment such as DIBELS.

Hiebert [53] provides insight into the distributions of words across the texts of a long-standing reading program from five decades (1962, 1993, 2000, 2008, 2013). The percentage of unique words among the 1000 most frequent words [7] fell from 60% in 1962 to 34% in 1993. That percentage remained

fairly consistent in the subsequent decades. The presence of phonetically regular words, defined as the percentage with regular vowels patterns [8], increased from 17% in 1962 to 29% in 1993. Since 1993, the percentage of words with regular vowel patterns has remained at 29%.

Another source of information on the features of words in beginning texts comes from statistical analyses of corpora, such as that conducted by Spencer [54] on a database of words from first-year texts used in beginning reading instruction in the U.K. [55]. Even in a limited database, the number of unique words was substantial: 6731. Within the words that accounted for the majority of the texts—the 1000 most frequent words—Spencer identified 217 combinations of phonemes and graphemes.

Spencer [54] next asked 105 six-year-old children to read the 150 most frequent words in the database. The variables that accounted for the largest proportion of students' word reading were the frequencies of individual words and of individual GPCs. Frequency, Spencer concluded, is the central mechanism for beginning readers' word knowledge.

4. The Current Study

In this study, we were interested in describing the features of words known and unknown by first graders of different proficiency levels in six passages of an oral reading fluency assessment: three in winter and three in spring. Specifically, this study addressed four research questions (RQs):

RQ1. Do features of words in ORF passages vary between winter and spring of Grade 1?

RQ2. How does the performance (words read correctly per minute (WCPM)) of proficiency groups differ in winter and spring?

RQ3. How do features of known and unknown words differ for different proficiency groups and at different points in time?

RQ4. Do the features of words that characterize word recognition follow a similar pattern for students acquiring proficiency at different trajectories? That is, are words known by low-group students in spring similar in kind and amount to those known by middle-group students in winter, and those of middle-group students in spring to those of high-group students in winter?

5. Method

5.1. Sample of Texts and Words

We chose to examine first graders' word recognition on ORF because of its widespread use in U.S. classrooms [56]. We focused specifically on the ORF measure of DIBELS for two reasons. First, of the four assessments of ORF that Toyama et al. [57] studied, DIBELS had the most consistent progression in difficulty within and across grades. Second, DIBELS is on many state-approved lists see [56,58].

The version used in this study was DIBELS-Next [59]. This version of DIBELS included three texts that were administered at the middle of the school year (winter) and three different texts that were administered at the end of the school year (spring) during Grade 1. The winter assessment was administered in early January and the spring assessment in May. The period of time between assessments, then, is approximately four months.

Each of the individual first-grade DIBELS texts averaged 235 words in length. A total of 1137 word tokens from six DIBELS ORF (DORF) passages (three from winter and three from spring) were analyzed to characterize DORF passages (RQ1). Of the 1137 tokens in the texts, 418 unique words were part of the current analysis. The remaining unique words appeared later in texts and were not read by any participants.

Because students at different proficiency levels can be expected to vary in the amount of text they read, features of words in different sections of the passages were analyzed (RQ1). Each text was clustered into segments of 25 words and was analyzed up to the fifth segment (or 125th word token per passage) because, according to Hasbrouck and Tindal's [60] norms, students at the 90th percentile can be expected to read 97 words in the winter and 116 in the spring.

A word was judged to be difficult/unknown if less than half the students in a proficiency group reached it but did not read it correctly. This level was chosen because the number of words known by students in the bottom half of the distribution was small, especially in winter. The 50% correct criterion ensured a sufficient number of known words to characterize features for different levels of readers. Features of known/unknown words were examined both within and across groups and across time (i.e., winter to spring). For RQs 3 and 4, the unit of analysis was word type, not word token.

5.2. Sample of Students

For this study, we analyzed word-level performance on the DORF assessments for 411 first-grade students at four schools in a school district in the Midwest from three academic years (2013–2014, 2014–2015, and 2015–2016). The racial/ethnic distribution of the school district was 62% Caucasian and 33% Hispanic, with the remaining students distributed between American Indigenous and African American. In its state's testing program, the district attained a level of 44% proficiency in reading.

The sample on which the current analyses are based consists of 411 students for whom complete data were available. An additional 64 students did not read the full set of six DORF passages and, consequently, were not included in the sample. For most of the 64 students, the incomplete data were a result of attendance at the school for only one point in time or a failure to have read all three passages in either winter or spring. The data of one student whose scores declined dramatically from winter to spring were also excluded. For those dropped students for whom there was winter data for at least one passage ($n = 39$), the average WCPM was 26.62 (SD = 28.15). This average contrasts with an average of 46.92 (SD = 29.07) for the 411 students in the sample. For the dropped students with spring data ($n = 34$), the spring average was 56.35 (SD = 36.77). This average contrasts with that for the entire sample of 71.79 (SD = 30.81).

The research interest of the current study centered on the nature of word recognition errors and not on the relative performances of students compared to one another, so this discrepancy is not concerning. The percentage of students with missing data, however, is a consideration from the vantage point of the representativeness of the current sample. In the sample of students with complete data, 64% were above the 50th percentile according to Hasbrouck and Tindal's [60] norms. The majority of the students with incomplete data performed below the 50th percentile, indicating that the distribution of the whole group including the students with missing data reflected a more typical distribution.

Of the 411 students with complete data for both points in times, four groups were formed based on students' median WCPM in the spring assessment in accordance with 25, 50, and 75 quartile points in the Hasbrouck and Tindal [60] norms. WCPM ranges for the four groups are provided in Table 1. Additionally included in Table 1 are the number of students in each group and the ranges for accuracy (percentage of words read correctly over total words read).

Table 1. Number of words read correctly per minute (WCPM) and accuracy by time and group.

Group [1] (*n*)	Hasbrouck and Tindal ORF Spring Norms [2]	WCPM: Mean [3] (SD)		Accuracy Mean [3] (SD)	
		Winter	Spring	Winter	Spring
Very High (102)	92–	83.68 (24.62)	110.49 (19.29)	96.37 (4.75)	98.6 (1.57)
High (160)	61–**91**	47.21 (5.14)	76.55 (9.34)	88.83 (7.68)	96.64 (3.13)
Middle (97)	35–**60**	26.34 (7.98)	49.47 (6.72)	76.26 10.56)	90.85 (5.51)
Low (52)	0–**34**	12.37 (5.76)	22.88 (9.51)	56.02 (13.77)	71.97 (17.41)

[1] Groups were formed on students' spring median WCPM based on Hasbrouck and Tindal's 2017 spring norms. [2] Numbers in bold are norms for 25, 50, and 75 percentiles [60] [3] Each student's median WCPM (out of three WCPMs per testing period) was averaged (i.e., mean of medians).

For each testing period, students read three passages, and their performance on each word was coded as correct (1), incorrect (0), or missing/unreached (NA). Additionally, for each assessment passage, the number of WCPM was recorded.

6. Variables for Analysis

6.1. Orthography

Decoding System Measure (DSyM): The DSyM [61] is a quantitative measure of word-level decoding difficulty that incorporates three variables, each an important predictor of word recognition accuracy and latency.

The measure is based on the premise that the frequency with which readers see a word influences their word recognition. Although derived from the same database as the frequency measure that is part of the representation of familiarity in this study, the nature of the variable and its weight in the DSyM make it distinct. Word frequency is calculated by subtracting the word frequency percentile score from the Standard Frequency Index (SFI) [7]. For example, the SFI of 0.8 for *you* is subtracted from 1 to get 0.2.

The second component is letter–sound discrepancy, which is the difference between the numbers of letters in a word and the number of phonemes. Because *you* has two phonemes and three letters, its letter–sound discrepancy score is 1 (3 − 2 = 1).

The third component of the DSyM is based on the conditional probabilities of a word's GPCs. This metric is obtained by subtracting the frequency of a grapheme–phoneme match based on the Berndt et al. [62] database from 1.0. In the example of the word *you*, the "ou"-/oo/ has a 0.04 frequency in the Berndt et al. corpus. The conditional probability score of 0.96 was obtained by subtracting 0.04 from 1.0.

These three measures are combined to produce the DSyM for a word. The word *you* receives a DSyM of 2.16 (the total of 0.2 [SFI] + 1 [letters − phonemes] + 0.96 [conditional for GPC]). By comparison, the word *it* has a DSyM of 0.456 (0.172 + 0 + 0.284), and *but* has a DSyM of 1.137 (0.239 + 0 + 0.898). In terms of relative complexity, *it* would be predicted to be the easiest of the three words, followed by *but* and then *you*.

Number of blends (nblends): The number of blends is included in the DSyM system [61], but is reported as a distinct score. Each blend is counted within a word for a score representing the total number of blends contained in each word. Only consonant blends, not digraphs, are included in this variable. The number of blends ranged from 0 to 2.

Word length: Three measures are indicators of length in this study: number of letters, phonemes, and syllables.

Number of letters (nletters): The length of the word is the number of letters as calculated by a digital word analyzer [63]. The code within the software is simply: NLet = length(word). Word length of the focus words ranged from 1 to 11.

Number of phonemes (nphonemes): The number of phonemes refers to the number of sounds contained in each word. For example, the word *ship* has four letters but only three sounds because the "sh" digraph makes a single sound. The number of phonemes ranged from 1 to 10.

Number of syllables (nsyllables): The number of syllables is the number of phonological syllables contained in each word. The Oxford English Dictionary (OED) [64] was consulted to confirm the syllabication breakdown of words. The number of syllables ranged from 1 to 4.

6.2. Word Familiarity

The familiarity measures pertain to overall frequency of words, age of acquisition, and concreteness.

Age of acquisition (AoA): This variable captures the age at which students typically understand or use a word in oral language. The data for this variable came from Kuperman et al. [65]. AoA ranged from 2.4 to 10.5. Fifteen words (3.2% of the sample) did not have a value for this variable. These words fell into three groups: proper names, contractions, and interjections.

Concreteness (concrete): In this study, a word's concreteness was based on the norms developed by Brysbaert et al. [66]. They defined a concrete word as something that can be experienced through the senses in contrast to an abstract word, which must be explained via language. Words were evaluated

on a five-point scale, where 1 was assigned to very abstract words (e.g., *charity*) and 5 was assigned to very concrete words (e.g., *chair*). Concreteness within the current sample of words ranged from 1.1 to 5, with 2.8% (or 13 words) missing a value for this variable. As was the case with AoA, these words were proper names, contractions, and interjections.

U function: The U function measure was used as the word frequency variable. Carroll et al. [67] originally identified the U function, which predicts the appearances of a word per million words of text as adjusted for a word's distribution across content areas. For the current study, the data on the U function came from Zeno et al.'s [7] Educator's Word Frequency Guide (EWFG), which was based on more than 17 million words of texts that represent school content areas and grade levels from first grade through college. The U functions of the words in this study ranged from a low of 0.7 to a high of 68,006.

6.3. Morphological Structure

Morphological structure was represented in three ways: number of morphemes, percentage of unique words that are multisyllabic, and percentage of unique words that are morphologically complex.

Number of morphemes (nmorphemes): Data on number of morphemes came from the OED [64]. Both word origin and etymology were considered when classifying words as morphologically complex and determining the number of morphemes in a word.

Multisyllabic words (%multisyllabic): Multisyllabic words were defined as those with more than one syllable. This measure was presented as a percentage of unique words in the sample of known or unknown words during a specific time period. For example, of the 46 words known by the middle-group students in winter, 19.6% were multisyllabic (e.g., *brother*). Among the 15 unknown words, 73.3% were multisyllabic (e.g., *compete*).

Morphologically complex words (%morphocomplex): Morphologically complex words were compound words (e.g., *sunlight*), words with inflected endings (e.g., *played*), derived words (e.g., *unusual*), or words with a combination of these categories (e.g., *refilled*). As with the previous measure, this measure was presented as a percentage of unique words. For example, among the 46 words that were known by the middle-group students in winter, 17.4% were morphologically complex (e.g., *going*). Of the 15 unknown words, 46.7% were morphologically complex (e.g., *younger*).

The examples of multisyllabic and morphologically complex words illustrate that not all multisyllabic words are morphologically complex (e.g., *brother*) and not all morphologically complex words are multisyllabic (e.g., *played*).

7. Results

Our interest lies in the features of words known by students in different proficiency groups and at two points in time over the first-grade academic year. As context for those analyses, we first present an overview of the features of words on the assessment.

RQ1. Do features of words in ORF passages vary between winter and spring of Grade 1?

Establishing the equivalence of passages administered in winter and spring is critical in that equivalence is needed if proportional changes in specific word features are to be attributable to changes in student proficiency. If passage difficulty is not equivalent over time, changes could be due to passage design.

Descriptive statistics for word features for the winter and spring DORF assessment passages are provided in Table 2. The unit of analysis was the word token. That is, if the word appeared five times in a section, each appearance was counted when computing a mean and a standard deviation. The information is also given for five consecutive segments of 25 words to determine whether demands of the first and second sections of text (that are usually the furthest that students in the bottom quartile get on the assessment) are similar to those of later text sections (that are read by more proficient students).

Table 2. Features of words in text segments in two periods of time: means (standard deviations).

Text Segment	Winter						Spring					
	All	**1**	**2**	**3**	**4**	**5**	**All**	**1**	**2**	**3**	**4**	**5**
						Orthography						
DSyM	2.0 (1.1)	2.0 (1.1)	2.2 (1.1)	2.1 (1.2)	2.0 (1.0)	2.0 (1.0)	2.1 (1.0)	2.1 (1.0)	2.0 (1.0)	2.2 (1.0)	2.1 (1.2)	2.0 (1.0)
nblends	0.27 (0.5)	0.25 (0.5)	0.37 (0.5)	0.34 (0.6)	0.23 (0.5)	0.23 (0.5)	0.31 (0.5)	0.29 (0.5)	0.26 (0.4)	0.38(0.6)	0.32 (0.6)	0.32 (0.6)
						Word Length						
nletters	3.9 (1.5)	3.8 (1.6)	4.2 (1.8)	4.1 (1.7)	3.8 (1.4)	3.8 (1.4)	3.94 (1.6)	4.1 (1.5)	4.0 (1.6)	3.9 (1.5)	3.8 (1.6)	3.96 (1.7)
nphonemes	3.1 (1.2)	3.1 (1.3)	3.5 (1.6)	3.3 (1.4)	3.0 (1.1)	3.0 (1.1)	3.23 (1.3)	3.3 (1.4)	3.2 (1.2)	3.2 (1.2)	3.2 (1.3)	3.4 (1.6)
nsyllables	1.2 (0.46)	1.2 (0.54)	1.2 (0.54)	1.2 (0.50)	1.1 (.37)	1.1 (0.44)	1.2 (0.51)	1.2 (0.43)	1.2 (0.60)	1.2 (0.38)	1.2 (.47)	1.3 (0.65)
						Word Familiarity						
AoA	4.6 (1.2)	4.6 (1.2)	4.5 (1.1)	4.6 (1.3)	4.8 (1.3)	4.5 (1.1)	4.6 (1.2)	4.7 (1.1)	4.6 (1.3)	4.7 (1.4)	4.7 (1.3)	4.6 (1.1)
concrete	2.8(1.2)	2.8 (1.2)	3.0 (1.3)	2.9 (1.3)	2.8 (1.3)	2.6 (1.1)	2.98	2.8 (1.2)	3.1 (1.2)	3.0 (1.3)	3.0 (1.3)	2.8 (1.3)
Ufunction	10,058 (18,976)	10,870 (20,190)	10,221 (18, 271)	8464 (17,586)	14,449 (23,883)	8996 (17,664)	10,513 (20,073)	11,458 (21,022)	6909 (14,968)	12,735 (23,064)	12,430 (23,848)	14,711 (23,153)
						Morphology						
nmorphemes	1.2 (0.4)	1.2 (0.4)	1.3 (0.5)	1.4 (0.5)	1.3 (0.5)	1.2 (0.4)	1.2 (0.4)	1.3 (0.5)	1.2 (0.4)	1.2 (0.4)	1.1 (0.3)	1.2 (0.4)
%multisyllabic [1]	17.2 (NA)	26.7 (NA)	24.0 (NA)	21.3 (NA)	12.0 (NA)	14.7 (NA)	17.8 (NA)	18.7 (NA)	21.3 (NA)	17.3 (NA)	12.0 (NA)	25.3 (NA)
%morphcomplex [1]	18.6 (NA)	18.7 (NA)	25.3 (NA)	29.3 (NA)	18.7 (NA)	13.3 (NA)	17.8 (NA)	25.3 (NA)	17.3 (NA)	18.7 (NA)	12.0 (NA)	18.7 (NA)

[1] Multisyllabic and morphcomplex are dummy variables, and values indicate percentages.

As the "all" columns in Table 2 show, the texts are similar between the two points in time. Further, sections of texts are generally comparable: Texts and text segments have words with similar averages for DSyM, AoA, concreteness, and number of letters, phonemes, syllables, and morphemes. Overall percentages of multisyllabic and morphologically complex words are similar. However, later text segments in each time period appear to include smaller proportions of multisyllabic and morphologically complex words than the earlier segments.

RQ2. How do the performances (WCPM) of proficiency groups differ in winter and spring?

Data on the WCPM and accuracy for the groups formed on the basis of Hasbrouck and Tindal's [60] norms appear in Table 1. As would be expected, both WCPM and accuracy rates varied considerably across the four groups.

Students in the very high group made few, if any, errors in either winter or spring (their mean accuracy rates were 96.4% and 98.6% in winter and spring, respectively, as shown in Table 1). The lack of errors in these students' performances meant that analyses on either known or unknown words were not relevant. Consequently, the remainder of the analyses focus on the other three groups: high, middle, and low.

For the high group, the average winter WCPM was 47.2 (SD = 5.14), and their average accuracy was 88.8% (SD = 7.68). This increased to 76.55 WCPM (SD = 9.34) in spring with 96.6% average accuracy (SD = 3.13), which was statistically significant ($p < 0.001$).

For the middle group, the average winter WCPM was 26.34 (SD = 7.98) with 76.3% average accuracy (SD = 10.56). In spring, this increased to 49.5 WCPM (SD = 6.72) with 90.6% average accuracy (SD = 5.51), which was a statistically significant change ($p < 0.001$).

Students in the low-performing group read an average of 12.4 WCPM (SD = 5.8) on the median passage in winter, and their average accuracy was 56% (SD = 13.8). In spring, the WCPM increased to an average of 22.9 (SD = 9.51), which was statistically significant ($p < 0.001$), and their average accuracy increased to 72% (SD = 17.4). The increase in accuracy levels meant that the number of words used for the word feature analysis for known and unknown words in the spring analysis was lower than for winter for the three groups (low, middle, and high).

RQ3. How do features of known and unknown words differ for different proficiency groups and at different points in time?

Words were classified as known if more than 50% of the students within a group were able to reach those words and correctly read them. Conversely, unknown words were words that more than 50% of the students in the group were unable to read correctly. Included in the latter category were words that the majority of students did not reach. These "neither known nor unknown" words were excluded from the analysis.

Features of known and unknown words in winter and spring were analyzed for each of the three groups. Means were compared between known and unknown words using the Wilcoxon tests (nonparametric method) within winter and spring because many of the word features do not follow the normal distribution. The Bonferroni correction was applied to account for inflated Type I errors through multiple comparisons. For proportions of multisyllabic and morphologically complex words, two-sample proportion tests were used.

High group: As can be seen in Table 3, known ($n = 87$) and unknown words ($n = 7$) differed on several word features for the high group (i.e., students whose WCPM in spring were percentiles 50 through 75 on Hasbrouck and Tindal's [60] norms) in winter. Compared to known words, unknown words had more letters (6.14 as compared to 4.33) and more phonemes (5.43 as compared to 3.39). Known words were more familiar as measured by age of acquisition (4.55 as compared to 6.55) and frequency (about 3600 occurrences per million as compared to about 160 occurrences). Of these differences, the number of phonemes and AoA were statistically significant ($p < 0.01$). Additionally, unknown words differed structurally from known words in that they had, on average, about one more syllable (2.14 as compared to 1.24), a difference that was statistically significant ($p < 0.001$). Further, unknown words were less concrete

than known words (3.09 as compared to 3.23) and were more challenging to decode, as determined by the DSyM (3.05 as compared to 2.32). These differences, however, were not statistically significant.

By the spring of first grade, the high group members were reading a substantial number of words correctly—135. The presence of only one unknown word in spring meant that a comparison between known and unknown words was not possible. Consequently, we report on the attributes of the known words, especially in relation to their features in winter. In other words, what had students learned over the period from January to May?

Most variables in every group had slightly higher averages, reflecting the ability of students in the high group to read the available words in the text. However, these differences were not significant. The average frequency rating was lower in spring than in winter (2750 relative to 3560) but, although suggesting that students in the high group recognized less frequent words more readily in spring than in winter, this difference was not statistically significant.

Features that had challenged students in this group in winter, such as the length of words (in letters, phonemes, and syllables) were no longer a problem. In particular, students in the high group had grown more proficient in recognizing multisyllabic words. An occasional multisyllabic word may not be recognized, but these students were reading relatively rapidly and with considerable accuracy.

Middle group: In winter, students in this group (i.e., students whose WRCMs in spring were percentiles 26 through 50 on Hasbrouck and Tindal's [60] norms) read with 75.4% accuracy. By the end of the year, they read with 96% accuracy. The number of unknown words that could be compared to known words in winter, then, was substantially greater than in spring (15 to 46 in winter; 4 to 97 in spring). The substantial number of words read correctly in spring for this group allows for a greater understanding of the kind of knowledge that this group of students—often ones who fall into the "provisional" range—have learned over the course of the year.

Features of known and unknown words for the middle group appear in Table 4. In winter, the features of known words relative to unknown words differed significantly on six features: Known words had an average of almost two more letters than unknown words, had a lower age of acquisition (4.36 vs. 6.43), were more frequent (U = 5551 vs. U = 114), and had one less syllable. Additionally, 73% of unknown words were multisyllabic and about 47% were morphologically complex, whereas only 20% and 17% of known words had such features. Further, unknown words had more phonemes, were more concrete, and had slightly greater numbers of blends and morphemes, but none of these differences were statistically different at $p = 0.05$.

In spring, known and unknown words differed statistically on the number of syllables (1.24 vs. 2) and the proportion of multisyllabic words (20% vs. 100%). Additionally, 24% of known words were morphologically complex, whereas 75% of unknown words were of this time, and this difference approached statistical significance ($p = 0.86$).

When words were unknown by middle-group students in spring, the words were multisyllabic and, in some cases, also morphologically complex: *baskets*, *needles*, *Africa*, and *island*. They were, however, able to read other multisyllabic words, such as *idea*, *money*, and *different*. Of the multisyllabic words read by the middle-group students in the winter, words were primarily inflected forms of highly frequent words (e.g., *going*, *looking*).

Low group: Unlike the other two groups, the words known by students in the low group were fewer (*n* = 16) than known words (*n* = 19) in winter. Features of these known and unknown words appear in Table 5. The majority of the word features either differed significantly between the known and unknown words, or the differences approached statistical significance. Words known by the low group were shorter (2.7 letters vs. 5 letters), more frequent (U = 13,190 vs. U = 915), and more structurally decodable as determined by DSyM (1.19 vs. 2.75), and none of them were multisyllabic and morphologically complex (whereas 53% and 42% of unknown words were such words, respectively). Additionally, words also had a close letter-to-sound match (2.4 phonemes for the 2.7 letters), had a single syllable, and had been in children's vocabularies for some time (AoA = 4 years).

Table 3. Features of known vs. unknown words: high group in winter and spring.

Word Feature	Variable	Winter [1]			Spring [2]		
		Known Mean (SD)	Unknown Mean (SD)	Z [3] (*p*)	Known Mean (SD)	Unknown Mean (SD)	Z [3] (*p*)
Orthography	DSyM	2.32 (1.14)	3.05 (0.98)	352 (1.000)	2.24 (1.02)	4.13 (NA)	NA
	nblends	0.36 (0.53)	0.5 (0.55)	291 (1.000)	0.38 (0.55)	1 (NA)	NA
Length	nletters	4.33 (1.6)	6.14 (1.07)	498 (0.228)	4.3 (1.56)	6 (NA)	NA
	nphonemes	3.39 (1.27)	5.43 (0.98)	526 (0.019)	3.53 (1.29)	5 (NA)	NA
	nsyllables	1.24 (0.48)	2.14 (0.38)	542.5 (<0.001)	1.26 (0.53)	2 (NA)	NA
Familiarity	AoA	4.55 (1.05)	6.55 (1.33)	513 (0.029)	4.7 (1.26)	7.41 (NA)	NA
	concrete	3.23 (1.14)	3.09 (1.19)	254.5 (1.000)	3.16 (1.11)	4.96 (NA)	NA
	Ufunction	3561.73 (9213.7)	161.86 (210.21)	143.5 (1.000)	2749.47 (7519.62)	129 (NA)	NA
Morphology	nmorphemes	1.27 (0.45)	1.29 (0.49)	273 (1.000)	1.26 (0.44)	1 (NA)	NA
	%multisyllabic	24.1 (NA)	100 (NA)	14.4 (<0.001)	23.0 (NA)	100 (NA)	NA
	%morphcomplex	24.1 (NA)	57.1 (NA)	2.1 (0.145)	25.2 (NA)	0 (NA)	NA

[1] Winter: Known = 87 words; Unknown = 7 words. [2] Spring: Known = 135 words; Unknown = 1 word. [3] Means for known and unknown words were compared in winter using the Wilcoxon test with Bonferroni correction. Similarly, for %multisyllabic and %morphcomplex, two sample proportion tests were conducted, and the reported values are chi-squared test statistics. Significance tests were not performed for spring because there was only one unknown word for this group.

Table 4. Features of known and unknown words: middle group in winter and spring.

Word Feature	Variable	Winter [1]			Spring [2]		
		Known Mean (SD)	Unknown Mean (SD)	Z [3] (p)	Known Mean (SD)	Unknown Mean (SD)	Z [3] (p)
Orthography	DSyM	2.06	3.02	493	2.15	3.16	311
		(1.13)	(1.07)	(0.407)	(1.04)	(0.70)	(1.000)
	nblends	0.31	0.47	378.5	0.28	1	326
		(0.51)	(0.64)	(1.000)	(0.5)	(0.00)	(0.117)
Length	nletters	3.85	5.6	556	4.2	6	315.5
		(1.51)	(1.3)	(0.016) [4]	(1.55)	(1.41)	(1.000)
	nphonemes	3.11	4.6	510	3.39	5.50	338
		(1.07)	(1.68)	(0.124)	(1.21)	(1.29)	(0.331)
	nsyllables	1.17	1.8	537	1.24	2	340
		(0.44)	(0.56)	(0.006) [4]	(0.54)	(0.00)	(0.038) [4]
Familiarity	AoA	4.36	6.43	531	4.59	7.21	341
		(0.85)	(1.15)	(<0.001) [4]	(1.14)	(2.34)	(0.308)
	concrete	2.98	3.55	368	3.07	4.34	303.5
		(1.19)	(1.04)	(1.000)	(1.08)	(1.26)	(1.000)
	Ufunction	5550.67	114.05	73.5	3315.39	71.75	35.5
		(12035)	(235.38)	(<0.001) [4]	(8697.86)	(47.86)	(0.297)
Morphology	nmorphemes	1.21	1.33	320	1.25	1.75	255.5
		(0.41)	(0.49)	(1.000)	(0.43)	(0.5)	(1.000)
	%multisyllabic	19.6	73.3	12.50	20.6	100	9.33
		(NA)	(NA)	(<0.001) [4]	(NA)	(NA)	(<0.001) [4]
	%morphcomplex	17.4	46.7	3.76	23.7	75	2.94
		(NA)	(NA)	(0.050) [4]	(NA)	(NA)	(0.086)

[1] Winter: Known = 46 words; Unknown = 15 words. [2] Spring: Known = 97 words; Unknown = 4 words. [3] Means for known and unknown words were compared within a testing period (e.g., winter) using the Wilcoxon test with Bonferroni correction. For %multisyllabic and %morphcomplex, two sample proportion tests were conducted; chi-squared test statistics are reported. [4] Significant at 0.05 or less.

Table 5. Features of known and unknown words: low group in winter and spring.

Word Feature	Variable	Winter [1]			Spring [2]		
		Known Mean (SD)	Unknown Mean (SD)	Z [3] (p)	Known Mean (SD)	Unknown Mean (SD)	Z [3] (p)
Orthography	DSyM	1.19	2.75	258	1.89	2.74	269
		(0.66)	(1.07)	(0.003) [4]	(0.98)	(0.68)	(0.62)
	nblends	0.13	0.47	178	0.21	0.62	226.5
		(0.35)	(0.70)	(1.000)	(0.41)	(0.74)	(1.000)
Length	nletters	2.69	5.05	289	3.72	5.50	281
		(0.79)	(1.39)	(<0.001) [4]	(1.3)	(1.31)	(0.186)
	nphonemes	2.4	4	232.5	3.05	5.00	292.5
		(0.63)	(1.60)	(0.060) [5]	(1)	(1.51)	(0.057) [5]
	nsyllables	0.94	1.58	236.5	1.14	1.50	234
		(0.25)	(0.61)	(0.031) [4]	(0.35)	(0.53)	(1.000)
Familiarity	AoA	4.06	5.52	212	4.24	5.46	230
		(0.66)	(1.25)	(0.076) [5]	(0.79)	(1.65)	(1.000)
	concrete	2.34	3.45	198.5	2.9	3.64	219
		(1.07)	(1.00)	(0.388)	(1.11)	(1.38)	(1.000)
	Ufunction	13,191.25	476.93	38.0	5833.98	185.50	48.0
		(18153)	(915.02)	(0.009) [4]	(12,114.06)	(197.28)	(0.069) [5]
Morphology	nmorphemes	1.09	1.39	128.5	1.15	1.62	194.5
		(0.30)	(0.50)	(1.000)	(0.36)	(0.52)	(0.306)
	%multisyllabic	0.00	52.60	9.35	14	50.00	3.05
		(NA)	(NA)	(0.002) [4]	(NA)	(NA)	(0.061) [5]
	%morphcomplex	0.00	42.10	6.51	11.60	62.50	8.08
		(NA)	(NA)	(0.011) [4]	(NA)	(NA)	(0.004) [4]

[1] Winter: Known = 16 words; Unknown = 19 words. [2] Spring: Known = 43 words; Unknown = 8 words. [3] Means for known and unknown words were compared within a testing period (e.g., winter) using the Wilcoxon test with Bonferroni correction. For %multisyllabic and %morphcomplex, two sample proportion tests were conducted; chi-squared test statistics are reported. [4] Significant at 0.05 or less. [5] Significant at 0.051 to 0.076.

In contrast, unknown words were longer, had more phonemes, were acquired later orally, had lower expected frequency in written language, had more than one syllable, and were more difficult to decode. The only pattern that did not follow the hypothesized direction was that of concrete words. Known words were more abstract than unknown words: 2.3 for the former and 3.5 for the latter. The explanation for this pattern is quite straightforward; high-frequency abstract words, such as *was* and *get*, were prominent in the known word corpus.

In spring, the number of known words for the low group outnumbered unknown words: 43 to 8. Known words increased in number and also in complexity from winter to spring, although only the proportion of morphologically complex words was found to differ significantly between the known and unknown words for the low group (12% vs. 63%). Three additional features' differences approached statistical significance: Known words were, on average, two phonemes less than unknown words and of higher expected frequency in written language (U = 5834 vs. U = 186), and only about 12% of them were morphologically complex, whereas 63% of unknown words had such characteristics. Unknown words were longer in terms of number of letters, phonemes, morphemes, and syllables. The unknown words are also likely to be in the oral language environments of students later. Once again, unknown words were more concrete than known words. However, these differences were not statistically significant.

Students in the low group did not know any multisyllabic or morphologically complex words in winter. By spring, there were still multisyllabic words (e.g., *market*) and morphologically complex words (e.g., *wanted*) that students in this group did not recognize. However, there were some multisyllabic words (e.g., *brother*) and morphologically complex (e.g., *helped*) words that they did recognize. They had made some progress in recognizing words beyond the highest-frequency words that comprised the majority of their repertoire in winter.

RQ4. Do the features of words that characterize word recognition follow a similar pattern for students acquiring proficiency at different trajectories? That is, are words known by low-group students in spring similar in kind and amount to those known by middle-group students in winter, and those of middle-group students in spring to those of high-group students in winter?

This question considers whether the progress of students resembles that of students in an adjacent group in the previous period in both the number of words read and in features of known words. Specifically, we compared the changes in the words known by students in the middle group in spring to patterns of their peers in the high group in winter and by students in the low group in spring to patterns of middle group in winter. Features of known words in winter and spring for the low, middle, and high groups are summarized in Table 6.

Comparison of the patterns in middle and high groups: By spring, students in the middle group had surpassed the number of words known, on average, by high-group students in winter—10 additional words. The features that characterized known words for the middle group in spring were almost identical to those of the high group in winter on a number of measures: numbers of phonemes, syllables, letters, and morphemes, as well as AoA and morphologically complex words. There were differences in DSyM, blends, and percentage of multisyllabic words, but these were not substantial. For example, the average DSyM for the middle group in spring was 2.15 (e.g., *air*), whereas the average DSyM for the high group in winter was 2.32 (e.g., *sure*). The two illustrative words fall into a similar group of complex GPCs pertaining to the influence of "r" on a vowel. Only four words were unknown by the middle group in spring, and all of these were multisyllabic, similar to the words that were not known by the high group in winter. These patterns suggest that the students in the middle group are following a similar progression in knowledge of words to that of the students in the high group.

Table 6. Features of known words from winter to spring for low, medium, and high groups.

Word Feature	Variable	Low Group			Middle Group			High Group		
		Winter M (SD) (*n* = 16)	Spring M (SD) (*n* = 43)	Z [1] (*p*)	Winter M (SD) (*n* = 46)	Spring M (SD) (*n* = 97)	Z [1] (*p*)	Winter M (SD) (*n* = 87)	Spring M (SD) (*n* = 135)	Z [1] (*p*)
Orthography	DSyM	1.19 (0.66)	1.89 (0.98)	185 (0.447)	2.06 (1.13)	2.15 (1.04)	2031.5 (1.000)	2.32 (1.14)	2.24 (1.02)	5813.5 (1.000)
	nblends	0.13 (0.35)	0.21 (0.41)	298 (1.000)	0.31 (0.51)	0.28 (0.50)	2192 (1.000)	0.36 (0.53)	0.38 (0.55)	5465 (1.000)
Length	nletters	2.69 (0.79)	3.72 (1.3)	173.5 (0.078) [2]	3.85 (1.51)	4.2 (1.55)	1914.5 (1.000)	4.33 (1.6)	4.3 (1.56)	5891.5 (1.000)
	nphonemes	2.4 (0.63)	3.05 (1)	207 (0.867)	3.11 (1.07)	3.39 (1.21)	1887 (1.000)	3.39 (1.27)	3.53 (1.29)	5213.5 (1.000)
	nsyllables	0.94 (0.25)	1.14 (0.35)	277.5 (1.000)	1.17 (0.44)	1.24 (0.54)	2169.5 (1.000)	1.24 (0.48)	1.26 (0.53)	5885.5 (1.000)
Familiarity	AoA	4.06 (0.66)	4.24 (0.79)	273 (1.000)	4.36 (0.85)	4.59 (1.14)	1888.5 (1.000)	4.55 (1.05)	4.7 (1.26)	5114.5 (1.000)
	concrete	2.34 (1.07)	2.9 (1.11)	218 (1.000)	2.98 (1.19)	3.07 (1.08)	1951.5 (1.000)	3.23 (1.14)	3.16 (1.11)	5535.5 (1.000)
	Ufunction	13,191.25 (18,153.07)	5833.98 (12,114.06)	444.5 (1.000)	5550.67 (12,034.93)	3315.39 (8697.86)	2420 (1.000)	3561.73 (9213.68)	2749.47 (7519.62)	5715 (1.000)
Morphology	nmorphemes	1.08 (0.3)	1.15 (0.36)	170.5 (1.000)	1.21 (0.41)	1.25 (0.43)	1556 (1.000)	1.27 (0.45)	1.26 (0.44)	4735.5 (1.000)
	%multisyllabic	0.0 (NA)	14.0 (NA)	1.2 (0.275)	19.6 (NA)	20.6 (NA)	6.8 (1.000)	24.1 (NA)	23.0 (NA)	0.001 (0.989)
	%morphcomplex	0.0 (NA)	11.6 (NA)	0.08 (0.368)	17.4 (NA)	23.7 (NA)	0.4 (0.552)	24.1 (NA)	25.2 (NA)	(< 0.001) (0.986)

[1] Means for known and unknown words were compared across the two testing periods using the Wilcoxon test with Bonferroni correction. For %multisyllabic and %morphcomplex, two sample proportion tests were conducted; chi-squared test statistics are reported. [2] Significant at 0.078.

Comparison of patterns in the low and middle groups: The low group had made substantial progress from winter to spring. They had seven fewer errors than the middle group had had in winter. However, at the same time, neither had they attained the WCPM in spring that the middle group had attained in winter, nor was their accuracy level as high. On all of the 11 measures, the low group in spring had lower averages than the middle group had had in spring. The word recognition of the low group had progressed substantially from winter to spring, but in only one case did the difference approach significance—number of letters. They could read words with one more letter than in the winter, but the kinds of words they could recognize were still heavily influenced by the frequencies of the words, as indicated by the DSyM and U function. The students in the low group had made movement in their recognition of words over the semester, but they had not quite attained the levels of their middle-group peers by the end of first grade.

8. Discussion

How students are evaluated as readers is based on their performances on ORF as well as word recognition tasks. Influential commentaries on deficiencies in students' reading processes or in the reading instruction they receive typically do not address the texts on which students are assessed or instructed e.g., [68]. This study provides insights into the word-level proficiencies of first graders in the passages of a widely used ORF assessment. These insights offer relevant next steps to both researchers and practitioners.

8.1. The Nature of the Task

True to CBM procedures [50], the spring expectations were the basis for evaluating winter performances as well as spring performances. At both points in time, the texts were almost identical in the features of words. What were the word features that characterized the texts on which first graders were evaluated in the middle and end of the year?

First, successful first-grade readers needed to be proficient with a broad range of orthographic patterns. The measure of orthography, DSyM, used in this study captures the decoding demands of a variety of words, including those with irregular vowel patterns and multisyllabic words [61]. The average words in the winter and spring texts had a DSyM of 2.0 and 2.1, respectively. Words with a DSyM of 2 are *eggs*, *old*, *wood*, and *both*. Only one of these words, *eggs*, has a vowel pattern with the typical foci of early phonics instruction (short vowel, long vowel). The other words have one of the following patterns: (a) regular but complex (*wood*); (b) unusual patterns, as described by Fry [25], where GPCs for vowels are not regular but appear in a relatively sizable group (e.g., *old*); or (c) rare, which Fry described as appearing in a limited number of words (e.g., *both*). The words in the texts had a DSyM that ranged from 0.46 (*in*) to 8.01 (*throughout*), which gives an indication of the variety of recognition of orthographic patterns expected of first graders.

To illustrate the decoding task from the perspective of first graders, especially those in the low group who typically do not get beyond the first two sentences of a text, we applied a simple rubric to the first two sentences of the three winter texts. The rubric and percentages for each category follow: (a) 36%—short or long vowels (the typical emphasis of the first-grade phonics curriculum); (b) 40%—monosyllabic words with rare vowel patterns (e.g., *friends*); and (c) 24%—multisyllabic words. A phonics curriculum that focuses on short and long vowels will not be sufficient for students to be successful with these assessment texts.

Another indication of the task for readers on the assessment texts comes from measures of length. The average word was 3.9 letters. This might suggest that demands for overall word recognition are reasonable. However, words with four letters represent a range of tasks. If words have a short-vowel pattern, they will have a blend or digraph (e.g., *fish*). Other words with four letters will require students to navigate GPCs that do not have a one-to-one correspondence (e.g., *game*, *wood*), words with two morphemes (e.g., *bees*), or words with two syllables (e.g., *over*).

In the design of DIBELS [51], the concern lies with word length in the form of words with seven or more characters and words with three or more syllables, rather than with morphological complexity. Only one word per passage had three syllables, but the percentage of two-syllabic words was, as already noted, 24%. Further, the percentage of seven-letter words may have seemed low to test developers—12%—and although the length of a word influenced the word recognition of students, especially those in the low group, the percentage of morphologically complex words was high—18%. Words that continued to challenge low-group students in spring included words that did not reach the seven-letter criterion, but had inflected endings (e.g., *wanted*). In summary, success with these first-grade assessment texts requires that students have facility with a range of orthographic patterns, including ones with irregular patterns and in multisyllabic and morphologically complex words.

8.2. The Nature of Student Performances

Even on what is an unarguably challenging task, many first graders were relatively successful as early as mid-first grade. That was not the case for all, but already in the winter of first grade, students who ended up in the very high group in spring could be described as proficient beginning readers. They read words in a passage with a range of word types with accuracy and automaticity. By the end of the year, they were, for all intents and purposes, ready for most texts. Texts in the grades that follow will be similar in their distribution of monosyllabic and multisyllabic words and relatively frequent and rare words, although rare words can be expected to increase in length, morphological complexity, and age of acquisition [19].

By the end of first grade, students in the high group attained the level of proficiency of their peers in the very high group in winter. They were able to accurately read almost all words (99%) at a fairly rapid rate. By most definitions of beginning reading acquisition, this group of students had become capable readers over the course of the school year.

Students in the middle group made substantial progress over a semester, attaining a moderate rate of oral reading and an accuracy level of 96%. Clinical observations have identified this accuracy level as adequate for comprehension [69,70], although analyses of adults learning English as a second language place the threshold as high as 98% [71,72]. Middle-group students appear to be on the cusp of the levels required for successful reading, although metaphorically, they can be described as not being out of the woods just yet. In particular, they struggle with the multisyllabic words that appear to be relatively prominent in first-grade texts [39], words such as *needles*, *baskets*, and *onto*.

Students in the low group made progress from winter to spring. The length of words that these students recognized had increased by a letter and were somewhat less frequent, and some words were multisyllabic. However, there were still many words that students were unable to read. The multisyllabic words that challenged the middle group were among these (*needles*, *basket*). However, students in the low group were also unable to read words with regular GPC patterns (*yet*, *shapes*) and words with inflected endings (*filled*, *wanted*). Further, their rate of reading was exceedingly slow. Young children may speak more slowly than adults, but a rate of 50 words per minute is slow, even for young children [73]. These students will depend greatly on the quality of the second-grade instruction and curriculum if they are to become proficient readers.

8.3. Limitations

In relation to the vastness of the English lexicon and the number of children who are taught to read in English each year, our samples of both words and students are small. However, previous analyses support the conclusion that the words on the DIBELS first-grade assessment are similar to those on other ORF assessments [57].

Further, we have reasons to be confident that the present results are not specific to a single school district. Not only did data come from three different groups of first graders, each representing a different academic year in five different schools, but this district's performance on summative performances is

below the average on the summative assessments of a state that performed in the middle range of the National Assessment of Educational Progress on the 2019 assessment [74].

An even more compelling reason for suggesting that these data are not isolated to one context comes from Hasbrouck and Tindal's [60] revision of ORF norms. From their last norms [75] to their current norms in 2017, first graders at the end of the school year at the 75th, 50th, and 25th percentiles recognized an average of seven additional words. A likely explanation for these changes in the oral reading rates of students is an earlier initiation of reading instruction. Comparisons of the Early Childhood Longitudinal Study from 1999 and 2011 show that kindergarten expectations for the students in the 2011 cohort were accelerated as compared to those of their counterparts in 1999 [16]. The literacy curriculum of kindergarten is no longer focused on alphabet and sound recognition activities, as was described several decades ago [76]. An analysis of the kindergarten texts of a core reading program in 2013 showed that the features of these texts were comparable to those of first-grade texts in 1965 [53].

Replications of this study with larger samples of students and additional words merit the attention of scholars investigating the progression of word recognition in beginning reading. The performances of this sample of first graders suggest that analyses should begin with entering first graders, if not with kindergartners. Investigations that address the depth of treatment required to learn specific GPCs and how these treatments differ for students with different entry skills are needed.

Further research on the individual progressions followed over the beginning reading period would benefit greatly from validation of measures of vowel GPC knowledge. The DSyM does not place students in terms of their knowledge along a linear phonics curriculum, as is the case with categorical measures, such as those of Menon and Hiebert [77] and Pirani-McGurl [24]. The DSyM [61] is highly useful in that it captures the variety and complexity of the word recognition task. However, measures that consider averages of word features in texts fail to put a spotlight on what may challenge students, such as words with the variety of r-influenced vowel GPCs. Measures such as the 10 categories used by Menon and Hiebert are efforts to describe decodability demands more directly. The latter proved to distinguish first- and second-grade students' reading achievement in the Fitzgerald et al. [10] study. However, the Menon and Hiebert measure would benefit from validation of the difficulty that blends and digraphs add to short and long vowels and of patterns of vowel + r. A study such as that conducted by Pirani-McGurl but with kindergarten and first-grade students, not students who already have a modicum of reading proficiency, is needed.

Finally, a shortcoming of our study was the lack of data on comprehension. Measures of comprehension in ORF assessments, including informal reading inventories, with beginning readers can be variable [78]. The high correlation between ORF performances in first grade and students' comprehension on silent reading assessments in later grades gives credence to the hypothesis that, at least in first grade, students' oral reading rate and accuracy are a viable proxy for comprehension [79]. Based on the rate of reading and accuracy levels of students in the bottom quartile in this study, even a modicum of comprehension on the tested passages is unlikely. Consider, for example, the words that were recognized in the first sentence of one of the winter passages by the low group: *it*, *was*, *the*, *of*, and *the*. The words that were not recognized carry the meaning of this sentence: *day*, *jump*, *rope*, and *contest*. Additional research is necessary to determine the decoding threshold required for comprehension with beginning readers, similar to research with middle- and high-school students conducted by Wang et al. [80].

8.4. Implications and Issues

Just as the patterns in the current study suggest directions for researchers, the study's findings have implications for those who work on the design and implementation of reading programs. A striking aspect of the data is the wide range of proficiency within a first-grade group. A portion of the sample—the very high group—have sufficient word recognition to be successful with texts that fall well into the second- to third-grade step of the Common Core State Standard's staircase of text complexity. The high group—those with accuracy levels of 99% and relatively high rates of word

recognition—are ready for more challenging texts, although their performances on extended texts and in silent reading contexts have not been documented. The students in the middle group progressed as readers during first grade, but they are not yet reading at sufficiently automatic levels to be considered proficient beginning readers. The students in the low-performing group have a substantial distance to go in the journey to automatic word recognition.

This variability across students completing their first-grade year raises the question of the degree of adaptivity in the beginning reading curriculum. The Hasbrouck and Tindal [60] norms suggest changes in the oral reading proficiencies of first-grade students. These changes have occurred even while instructional texts have increased in complexity [10]. Whether the texts of ORF have increased in their complexity over the period represented in the Hasbrouck and Tindal norms has not been documented. However, the earlier initiation of reading instruction and changes in texts raise the question of how beginning reading curriculum and instructional practices have changed.

The introduction of the Response to Intervention (RtI) legislation might support a perception that adaptations to curriculum and instruction based on assessments of students' proficiencies are widespread. The gist of a recent spate of journalistic investigations into beginning reading instruction, however, suggests that a "one-size-fits-all" perspective dominates [81,82]. That is, an entire age group appears to be described in a similar manner. However, if the first-grade curriculum is, indeed, represented by the DIBELS texts—which are described as "curriculum-based measurement" (CBM) [50]—the variability within a first-grade group relative to the curriculum is substantial.

Our search in the research and pedagogical literature for reports of beginning reading curriculum, past and present, has produced few results. Do the texts on the DIBELS assessment truly represent the end point of the first-grade curriculum? One compelling question is whether the substantial growth during the first phases of beginning reading is represented by typical CBM assessments. Might the students in the low group in the current study, for example, have been successful if the assessment texts consisted of high percentages of words with short vowels?

A perusal of the word study curriculum for the program that Schwartz [82] identified as most widely used in American beginning classrooms—Leveled Literacy Intervention (LLI) [83]—shows a predominance of lessons on short and long vowels, but little attention to more complex vowel patterns or multisyllabic words. In the texts of the LLI program that students read for practice, however, multisyllabic words are prominent from the beginning levels [39], and a measure of word count but not decodability predicts the level at which texts are placed in leveled text programs [84]. In another widely used word study program, the bulk of attention in kindergarten and first-grade programs addresses short and long vowels as well as high-frequency words [85]. First-grade reading instruction that continues to cycle through short vowels and blends, at a point when students are expected to read these words as kindergartners, can be expected to fall short of the mark for all students—not only those who have not mastered these patterns, but also those with more proficiency but who need guidance with multisyllabic and morphologically complex words.

Without assessments that capture the variability among first graders, it is doubtful that curricula will be differentiated. The range of student proficiency calls for differentiated curricula, not a single one. Many in the first-grade group, including those in the middle group, were able to recognize monosyllabic words with a range of patterns at the end of first grade. It was multisyllabic words that challenged them. The consequences of learning to read with texts that have an abundance of multisyllabic words but no instruction other than the recommendation to "use the picture" to assist in recognizing these words [85,86] requires examination.

The lack of differentiated curriculum, the poor match between curriculum and texts, and the lack of instruction in multisyllabic words likely contribute to the poor word recognition capabilities of a portion of an American cohort. There is evidence that, when assessments include multisyllabic words, a and age of acquisition middle- and even high-school students are not adept at specific orthographic–phonic knowledge [80,87]. Wang et al. [80] reported that as many as 38% of Grade 5 students and 19% of

Grade 10 students were below a decoding threshold that predicted comprehension. These students did not improve in their reading comprehension in the following three years; their peers did.

Reading instruction currently begins earlier for American students than it did for previous generations. Overall, students are completing first grade with ORF rates higher than previous first-grade students. By third grade, students at the bottom 20th percentile read DIBELS passages with 95% accuracy, and students at the 50th percentile with 99% [88]. Why, then, are students not doing better on the NAEP? One explanation may lie in the failure of curriculum to adapt to changes in students and texts, which has already been discussed. First graders may be able to recognize more words more rapidly than in previous eras. However, without adaptations to the curriculum and instruction that recognizes the proficiencies and needs across a cohort of students, the accomplishments of students and teachers during the early years are unlikely to produce the desired outcomes.

References

1. Cain, K.; Compton, D.L.; Parrila, P.K. (Eds.) *Theories of Reading Development*; John Benjamins Publishing Company: Amsterdam, The Netherlands, 2017; Volume 15.
2. Metsala, J.L.; Ehri, L.C. (Eds.) *Word Recognition in Beginning Literacy*; Routledge: Abingdon, UK, 2013.
3. Yap, M.J.; Balota, D.A. Visual word recognition. In *The Oxford Handbook of Reading*; Pollatsek, A., Treiman, R., Eds.; Oxford University Press: Oxford, UK, 2015; pp. 26–43.
4. Joseph, H.S.; Nation, K. Examining incidental word learning during reading in children: The role of context. *J. Exp. Child Psychol.* **2018**, *166*, 190–211. [CrossRef] [PubMed]
5. Hiebert, E.H. State Reform Policies and the Task Textbooks Pose for First-Grade Readers. *Elem. Sch. J.* **2005**, *105*, 245–266. [CrossRef]
6. Vousden, J.I. Units of English spelling-to-sound mapping: A rational approach to reading instruction. *Appl. Cogn. Psychol.* **2008**, *22*, 247–272. [CrossRef]
7. Zeno, S.; Ivens, S.H.; Millard, R.T.; Duvvuri, R. *The Educator's Word Frequency Guide*; Touchstone Applied Science Associates: Brewster, MA, USA, 1995.
8. Juel, C.; Roper/Schneider, D. The Influence of Basal Readers on First Grade Reading. *Read. Res. Q.* **1985**, *20*, 134. [CrossRef]
9. Ehri, L.C. Learning to read words: Theory, findings, and issues. *Sci. Stud. Read.* **2005**, *9*, 167–188. [CrossRef]
10. Fitzgerald, J.; Elmore, J.; Relyea, J.E.; Hiebert, E.H.; Stenner, A.J. Has first-grade core reading program text complexity changed across six decades? *Read. Res. Q.* **2016**, *51*, 7–28.
11. Foorman, B.R.; Francis, D.J.; Davidson, K.C.; Harm, M.W.; Griffin, J. Variability in text features in six grade 1 basal reading programs. *Sci. Stud. Read.* **2004**, *8*, 167–197. [CrossRef]
12. Ardoin, S.P.; Christ, T.J.; Morena, L.S.; Cormier, D.C.; Klingbeil, D.A. A systematic review and summarization of the recommendations and research surrounding curriculum-based measurement of oral reading fluency (CBM-R) decision rules. *J. Sch. Psychol.* **2013**, *51*, 1–18. [CrossRef]
13. Wanzek, J.; Vaughn, S. Tier 3 Interventions for Students with Significant Reading Problems. *Theory Pr.* **2010**, *49*, 305–314. [CrossRef]
14. Cabell, S.Q.; Justice, L.M.; Konold, T.R.; McGinty, A.S. Profiles of emergent literacy skills among preschool children who are at risk for academic difficulties. *Early Child. Res. Q.* **2011**, *26*, 1–14. [CrossRef]
15. Davidse, N.J.; De Jong, M.T.; Bus, A.G.; Huijbregts, S.C.; Swaab, H. Cognitive and environmental predictors of early literacy skills. *Read. Writ.* **2011**, *24*, 395–412. [CrossRef] [PubMed]
16. Bassok, D.; Latham, S.; Rorem, A. Is kindergarten the new first grade? *AERA Open* **2016**, *2*, 2332858415616358. [CrossRef]
17. Ada, A.F.; Beers, K.; Campoy, F.I.; Carroll, J.A.; Clemens, N.; Cunningham, A.; Hougen, M.C.; Izquierdo, E.; Jago, C.; Palmer, E. *Into Reading (Book 1)*; Houghton Mifflin Harcourt: Boston, MA, USA, 2020.

18. Hiebert, E.H.; Scott, J.A.; Castaneda, R.; Spichtig, A. An analysis of the features of words that influence vocabulary difficulty. *Educ. Sci.* **2019**, *9*, 8. [CrossRef]

19. Kearns, D.M.; Hiebert, E.H. The words of primary-level texts: Differences between first and third grade in widely-used curriculum. **2020**, unpublished manuscript.

20. Gagl, B.; Hawelka, S.; Richlan, F.; Schuster, S.; Hutzler, F. Parafoveal preprocessing in reading revisited: Evidence from a novel preview manipulation. *J. Exp. Psychol. Learn. Mem. Cogn.* **2014**, *40*, 588–595. [CrossRef]

21. Venezky, R.L. *The American Way of Spelling: The Structure and Origins of American English Orthography*; Guilford Press: New York, NY, USA, 1999.

22. Leslie, L.; Calhoon, A. Factors affecting children's reading of rimes: Reading ability, word frequency, and rime-neighborhood size. *J. Educ. Psychol.* **1995**, *87*, 576–586. [CrossRef]

23. Guthrie, J.T.; Seifert, M. Letter–sound complexity in learning to identify words. *J. Educ. Psychol.* **1977**, *69*, 686–696. [CrossRef]

24. Pirani-McGurl, C.A. The Use of Item Response Theory in Developing a Phonics Diagnostic Inventory. Open Access Dissertations. 2009. Available online: https://scholarworks.umass.edu/ (accessed on 11 December 2020).

25. Fry, E. Phonics: A Large Phoneme—Grapheme Frequency Count Revised. *J. Lit. Res.* **2004**, *36*, 85–98. [CrossRef]

26. Barton, J.J.; Hanif, H.M.; Eklinder Björnström, L.; Hills, C. The word-length effect in reading: A review. *Cogn. Neuropsychol.* **2014**, *31*, 378–412. [CrossRef]

27. Zoccolotti, P.; De Luca, M.; Di Filippo, G.; Judica, A.; Martelli, M. Reading development in an orthographically regular language: Effects of length, frequency, lexicality and global processing ability. *Read. Writ.* **2008**, *22*, 1053–1079. [CrossRef]

28. Marinus, E.; De Jong, P.F. Variability in the word-reading performance of dyslexic readers: Effects of letter length, phoneme length and digraph presence. *Cortex* **2010**, *46*, 1259–1271. [CrossRef] [PubMed]

29. Yap, M.J.; Balota, D.A. Visual word recognition of multisyllabic words. *J. Mem. Lang.* **2009**, *60*, 502–529. [CrossRef]

30. Seymour, P.H. Foundations of orthographic development. In *Learning to Spell: Research, Theory, and Practice across Languages*; Perfetti, C.A., Rieben, L., Fayol, M., Eds.; Routledge: Abingdon, UK, 1997; pp. 319–337.

31. Rastle, K. The place of morphology in learning to read in English. *Cortex* **2019**, *116*, 45–54. [CrossRef] [PubMed]

32. Henry, M.K. *Unlocking Literacy: Effective Decoding & Spelling Instruction*; Paul H Brookes Publishing Company: Baltimore, MD, USA, 2003.

33. Masterson, J.; Stuart, M.; Dixon, M.; Lovejoy, S. Children's printed word database: Continuities and changes over time in children's early reading vocabulary. *Br. J. Psychol.* **2010**, *101*, 221–242. [CrossRef] [PubMed]

34. Nagy, W.E.; Hiebert, E.H. Toward a theory of word selection. In *Handbook of Reading Research*; Kamil, M.L., Pearson, P.D., Moje, E.B., Afflerbach, P.P., Eds.; Longman: White Plains, NY, USA, 2011; Volume 4, pp. 388–404.

35. Perfetti, C.A.; Stafura, J.Z. Word Knowledge in a Theory of Reading Comprehension. *Sci. Stud. Read.* **2014**, *18*, 22–37. [CrossRef]

36. Taylor, J.N.; Perfetti, C.A. Eye movements reveal readers' lexical quality and reading experience. *Read. Writ.* **2016**, *29*, 1069–1103. [CrossRef]

37. Ricketts, J.; Davies, R.; Masterson, J.; Stuart, M.; Duff, F.J. Evidence for semantic involvement in regular and exception word reading in emergent readers of English. *J. Exp. Child Psychol.* **2016**, *150*, 330–334. [CrossRef]

38. Clifford, G.J. Words for schools: The applications in education of the vocabulary researches of Edward, L. Thorndike. In *Impact of Research on Education*; Suppes, P., Ed.; National Academy of Education: Washington, DC, USA, 1978; pp. 107–198.

39. Murray, M.; Munger, K.A.; Hiebert, E.H. An analysis of two reading intervention programs: How do the words, texts, and programs compare? *Elem. Sch. J.* **2014**, *114*, 479–500. [CrossRef]

40. Bonin, P.; Barry, C.; Méot, A.; Chalard, M. The influence of age of acquisition in word reading and other tasks: A never ending story? *J. Mem. Lang.* **2004**, *50*, 456–476. [CrossRef]

41. Juhasz, B.J. Age-of-Acquisition Effects in Word and Picture Identification. *Psychol. Bull.* **2005**, *131*, 684–712. [CrossRef]

42. Altarriba, J.; Bauer, L.M.; Benvenuto, C. Concreteness, context availability, and imageability ratings and word associations for abstract, concrete, and emotion words. *Behav. Res. Methods Instrum. Comput.* **1999**, *31*, 578–602. [CrossRef] [PubMed]

43. Sabsevitz, D.; Medler, D.; Seidenberg, M.; Binder, J. Modulation of the semantic system by word imageability. *NeuroImage* **2005**, *27*, 188–200. [CrossRef] [PubMed]

44. Steacy, L.M.; Compton, N.L. Examining the role of imageability and regularity in word reading accuracy and learning efficiency among first and second graders at risk for reading disabilities. *J. Exp. Child Psychol.* **2018**, *178*, 226–250. [CrossRef] [PubMed]

45. Adams, M.J. *Beginning to Read: Thinking and Learning about Print*; MIT Press: Cambridge, MA, USA, 1990.

46. Graves, M.F.; Elmore, J.; Fitzgerald, J. The Vocabulary of Core Reading Programs. *Elem. Sch. J.* **2019**, *119*, 386–416. [CrossRef]

47. Carlisle, J.F.; Fleming, J. Lexical processing of morphologically complex words in the elementary years. *Sci. Stud. Read.* **2003**, *7*, 239–253. [CrossRef]

48. Carlisle, J.F.; Nomanbhoy, D.M. Phonological and morphological awareness in first graders. *Appl. Psycholinguist.* **1993**, *14*, 177–195. [CrossRef]

49. Petscher, Y.; Cummings, K.D.; Biancarosa, G.; Fien, H. Advanced (Measurement) Applications of Curriculum-Based Measurement in Reading. *Assess. Eff. Interv.* **2013**, *38*, 71–75. [CrossRef]

50. Fuchs, L.S.; Fuchs, D.; Hosp, M.K.; Jenkins, J.R. Oral Reading Fluency as an Indicator of Reading Competence: A Theoretical, Empirical, and Historical Analysis. *Sci. Stud. Read.* **2001**, *5*, 239–256. [CrossRef]

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

52. Ehri, L.C. Orthographic mapping and literacy development revisited. In *Theories of Reading Development*; Cain, K., Compton, D.L., Parrila, R.K., Eds.; John Benjamins Publishing Company: Amsterdam, The Netherlands, 2017; pp. 127–146.

53. Hiebert, E.H. Changing readers, changing texts: Beginning reading texts form 1960 to 2010. *J. Educ. Sci.* **2015**, *195*, 1–14. [CrossRef]

54. Spencer, K. Predicting children's word-reading accuracy for common English words: The effect of word transparency and complexity. *Br. J. Psychol.* **2010**, *101*, 519–543. [CrossRef]

55. Stuart, M.; Dixon, M.; Masterson, J.; Gray, B. Children's early reading vocabulary: Description and word frequency lists. *Br. J. Educ. Psychol.* **2003**, *73*, 585–598. [CrossRef] [PubMed]

56. Croft, A. *State Adoption and Implementation of K-2 Assessments*; ACT: Iowa City, IA, USA, 2016; Available online: https://www.act.org/content/dam/act/unsecured/documents/5738_Issue_Brief_State_Adoption_of_K-2_Assess_WEB_secure.pdf (accessed on 16 December 2020).

57. Toyama, Y.; Hiebert, E.H.; Pearson, P.D. An Analysis of the Text Complexity of Leveled Passages in Four Popular Classroom Reading Assessments. *Educ. Assess.* **2017**, *22*, 139–170. [CrossRef]

58. Colorado Department of Education. READ: Approved K-3 Diagnostic and Summative Assessments. Available online: https://www.cde.state.co.us/coloradoliteracy/readdiagnosticandsummativeassessments (accessed on 28 July 2020).

59. Good, R.H., III; Kaminski, R.A. *Dynamic Indicators of Basic Early Literacy Skills (Next ed.)*; Dynamic Measurement Group: Eugene, OR, USA, 2011.

60. Hasbrouck, J.; Tindal, G.A. *An Update to Compiled ORF Norms (Technical Report No. 1702)*; Behavioral Research and Teaching, University of Oregon: Eugene, OR, USA, 2017.

61. Saha, N.M.; Cutting, L.E.; Del Tufo, S.; Bailey, S. Initial validation of a measure of decoding difficulty as a unique predictor of miscues and passage reading fluency. *Read. Writ.* **2020**, 1–31. [CrossRef]

62. Berndt, R.S.; Reggia, J.A.; Mitchum, C.C. Empirically derived probabilities for grapheme-to-phoneme correspondences in English. *Behav. Res. Methods Instrum. Comput.* **1987**, *19*, 1–9. [CrossRef]

63. Hiebert, E.H. *WordZone Analyzer*; TextProject: Santa Cruz, CA, USA, 2012.

64. Simpson, J.A.; Weiner, E.S.C. *Oxford English Dictionary*, 2nd ed.; Clarendon Press: Oxford, UK, 1989.

65. Kuperman, V.; Stadthagen-Gonzalez, H.; Brysbaert, M. Age-of-acquisition ratings for 30,000 English words. *Behav. Res. Methods* **2012**, *44*, 978–990. [CrossRef] [PubMed]

66. Brysbaert, M.; Warriner, A.B.; Kuperman, V. Concreteness ratings for 40 thousand generally known English word lemmas. *Behav. Res. Methods* **2014**, *46*, 904–911. [CrossRef] [PubMed]

67. Carroll, J.B.; Davies, P.; Richman, B. *The American Heritage Word Frequency Book*; Houghton-Mifflin: Boston, MA, USA, 1972.

68. Seidenberg, M. *Language at the Speed of Sight: How We Read, Why so Many Can't, and What Can Be Done about It*; Basic Books: New York, NY, USA, 2017.

69. Betts, E.A. *Foundations of Reading Instruction, with Emphasis on Differentiated Guidance*; American Book Company: Woodstock, GA, USA, 1946.

70. Stahl, S.A.; Heubach, K.M. Fluency-Oriented Reading Instruction. *J. Lit. Res.* **2005**, *37*, 25–60. [CrossRef]

71. Hsueh-Chao, M.H.; Nation, P. Unknown vocabulary density and reading comprehension. *Read. Foreign Lang.* **2000**, *13*, 403–430.

72. Schmitt, N.; Cobb, T.; Horst, M.; Schmitt, D. How much vocabulary is needed to use English? Replication of van Zeeland & Schmitt (2012), Nation (2006) and Cobb (2007). *Lang. Teach.* **2015**, *50*, 212–226.

73. Sturm, J.A.; Seery, C.H. Speech and Articulatory Rates of School-Age Children in Conversation and Narrative Contexts. *Lang. Speech Hear. Serv. Sch.* **2007**, *38*, 47–59. [CrossRef]

74. National Center for Education Statistics. *The Nation's Report Card: NAEP Reading Assessment*; National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education: Washington, DC, USA, 2019.

75. Hasbrouck, J.; Tindal, G.A. Oral Reading Fluency Norms: A Valuable Assessment Tool for Reading Teachers. *Read. Teach.* **2006**, *59*, 636–644. [CrossRef]

76. Hiebert, E.H.; Papierz, J.M. The emergent literacy construct and kindergarten and readiness books of basal reading series. *Early Child. Res. Q.* **1990**, *5*, 317–334. [CrossRef]

77. Menon, S.; Hiebert, E.H. A comparison of first graders' reading with little books or literature-based basal anthologies. *Read. Res. Q.* **2005**, *40*, 12–38. [CrossRef]

78. Kim, Y.-S.; Petscher, Y.; Schatschneider, C.; Foorman, B. Does growth rate in oral reading fluency matter in predicting reading comprehension achievement? *J. Educ. Psychol.* **2010**, *102*, 652–667. [CrossRef]

79. Wanzek, J.; Roberts, G.; Linan-Thompson, S.; Vaughn, S.; Woodruff, A.L.; Murray, C.S. Differences in the Relationship of Oral Reading Fluency and High-Stakes Measures of Reading Comprehension. *Assess. Eff. Interv.* **2010**, *35*, 67–77. [CrossRef]

80. Wang, Z.; Sabatini, J.; O'Reilly, T.; Weeks, J. Decoding and reading comprehension: A test of the decoding threshold hypothesis. *J. Educ. Psychol.* **2019**, *111*, 387–401. [CrossRef]

81. Hanford, E. Why are we still teaching reading the wrong way? *New York Times*, 26 October 2018. Available online: https://www.nytimes.com/2018/10/26/opinion/sunday/phonics-teaching-reading-wrong-way.html (accessed on 16 December 2020).

82. Schwartz, S. The Most Popular Reading Programs Aren't Backed by Science. Education Week. 2019. Available online: https://www.edweek.org/teaching-learning/the-most-popular-reading-programs-arent-backed-by-science/2019/12 (accessed on 16 December 2020).

83. Fountas, I.C.; Pinnell, G.S. *Leveled Literacy Intervention*; Heinemann: Portsmouth, NH, USA, 2008.

84. Cunningham, J.W.; Spadorcia, S.A.; Erickson, K.A.; Koppenhaver, D.A.; Strum, J.M.; Yoder, D.E. Investigating the instructional supportiveness of leveled texts. *Read. Res. Q.* **2005**, *40*, 410–427. [CrossRef]

85. Calkins, L.; Louis, N. *A Guide to the Phonics Units of Study, Grades K-2*; Heinemann: Portsmouth, NH, USA, 2019.

86. Fountas, I.C.; Pinnell, G.S. *Comprehensive Phonics, Spelling, and Word Study Guide*; Heinemann: Portsmouth, NH, USA, 2017.

87. Durkin, D. *The Decoding Ability of Elementary School Students (Reading Education Report No. 49)*; Center for the Study of Reading: Champaign, IL, USA, 1984.

88. Dewey, E.N.; Kaminski, R.A.; Good, R.H., III. *DIBELS Next National Norms 2012–2013 (Technical Report No. 17)*; Dynamic Measurement Group: Eugene, OR, USA, 2014.