Reading And Writing, **in press (April 23, 2026)**

The Role of Text Vocabulary in Word Recognition, Reading Rate, and Comprehension of First-Grade Students

Lori Bruner
School of Education
University at Albany
lbruner@albany.edu

ORCID: 0000-0001-8845-0206

Elfrieda H. Hiebert
TextProject
hiebert@textproject.org

ORCID: 0000-0002-2864-7549

Laura S. Tortorelli
College of Education
Michigan State University
ltort@msu.edu

ORCID: 0000-0001-9266-5939

Author Notes:

1. The authors report no conflicts of interest.

2. Correspondence regarding this article should be addressed to Dr. Lori Bruner, Assistant Professor of Literacy, University at Albany: **lbruner@albany.edu**

Abstract

This study examined how vocabulary complexity influences first-grade students' word recognition, reading rate, and comprehension. Seventy-eight first-grade students read six 100-word narrative passages equivalent in Guided Reading Level (GRL J) but systematically varied in Mean Log Word Frequency (MLWF), a Lexile Framework index of vocabulary difficulty. Multilevel modeling indicated that MLWF significantly predicted accuracy, words correct per minute, and maze comprehension, accounting for 12%, 21%, and 1% of text-level variance respectively. Exploratory quantile regression indicated that vocabulary difficulty had stronger effects on students with lower and mid-range fluency than on their more proficient peers, suggesting a developmental "vulnerability window" during which lexical demands are especially consequential. A binomial model of word-level errors showed that words with more phonemes, lower spelling-to-sound consistency, and later ages of acquisition were more likely to be misread. Together, these findings demonstrate that texts labeled at the same Guided Reading Level can pose markedly different lexical demands, with disproportionate consequences for less fluent readers. We discuss implications for text selection, assessment-to-text matching, and the design of materials that more precisely calibrate vocabulary demands for beginning readers.

The Role of Text Vocabulary in Word Recognition, Reading Rate, and Comprehension of First-Grade Students

Over the past six decades, elementary core reading programs in the United States have undergone shifts in their linguistic and lexical features (Authors, 2013; Authors, 2016; Hoffman et al., 2002). An analysis of seven editions of a first-grade core reading program published between 1962 and 2013 documented a steady increase in text complexity, driven largely by changes in vocabulary characteristics such as age of acquisition and word frequency (Authors, 2016). These shifts have introduced words acquired later and occurring less often in children's language, potentially increasing the demands placed on young readers. Yet, despite documented changes in vocabulary profiles, little is known about how vocabulary complexity influences foundational reading processes, including word recognition, fluency, and comprehension.

To address this gap, we examined how variation in vocabulary complexity relates to first-grade students' reading performance. Specifically, we held Guided Reading Level (GRL) constant while systematically varying Mean Log Word Frequency (MLWF), a vocabulary metric from the Lexile Framework. Guided Reading Levels (GRLs) are a widely used qualitative leveling system in U.S. schools, whereas MLWF provides a quantitative indicator of word frequency and lexical demand. By analyzing texts classified at the same GRL but differing in MLWF, we isolated vocabulary's independent contribution to students' reading outcomes.

Variation in lexical demands has important implications for how texts are selected and used in classrooms. Many early-grade reading curricula rely heavily on leveled texts organized by systems such as GRLs. However, texts classified at the same instructional level can differ substantially in their lexical demands (Authors, 2022a). These differences may disproportionately affect students who are still consolidating the decoding and language skills

necessary to navigate unfamiliar words in text (Perfetti, 2007, 2010; Stanovich, 1986). By clarifying how vocabulary complexity relates to accuracy, rate, and comprehension, and identifying the types of words most likely to result in errors, this study offers empirical guidance for more responsive text selection and instructional scaffolding for young readers. Specifically, we investigate (a) how MLWF relates to reading accuracy, rate, and comprehension; (b) whether the association between MLWF and reading performance differs across proficiency levels; and (c) which word-level features are associated with higher error rates.

Conceptual Foundations and Empirical Grounding

Reading comprehension reflects the coordinated interaction of decoding and language comprehension (Gough & Tunmer, 1986). However, understanding the role of vocabulary within this interaction has proven theoretically complex. Although the Simple View of Reading situates vocabulary within oral language comprehension, vocabulary knowledge also supports decoding efficiency (Perfetti, 2007, 2010), or the speed and automaticity with which readers recognize words. Vocabulary's contribution, therefore, cannot be neatly confined to a single component of this model. The Lexical Quality Hypothesis (Perfetti & Hart, 2002) provides a more integrated account, proposing that high-quality lexical representations combine phonological, orthographic, and semantic information into well-specified word knowledge. For beginning readers, this integration means that words well established in oral language are recognized more quickly and accurately because they activate more fully specified representations. Greater efficiency in word recognition, in turn, frees working memory for constructing meaning. From this perspective, vocabulary is not simply a component of comprehension; it is also central to fluent decoding. Managing vocabulary demands in first-grade texts, therefore, may be critical for supporting both

efficient processing and meaning-making. The present study applies this framework to three interrelated outcomes: word recognition, reading rate, and comprehension.

Word Recognition

For beginning readers, vocabulary supports word recognition by facilitating mapping between orthographic forms and existing phonological and semantic representations. When readers encounter unfamiliar orthographic forms of words in their oral vocabulary, they apply letter-sound knowledge to generate a phonological candidate. If that candidate sufficiently matches a word in their oral vocabulary, the word is identified and the associated letter-sound correspondences are strengthened. With repeated exposure, these connections become more efficient, and the word may become part of the reader's sight word repertoire. From a lexical quality perspective (Perfetti & Hart, 2002), variation in word recognition efficiency reflects how fully specified words are within the reader's lexicon. Words that occur more frequently, appear across diverse contexts, and possess greater semantic richness accumulate stronger representations (Brysbaert & Cortese, 2011), whereas less familiar or weakly specified words require more effortful processing.

Empirical research supports this account. Vocabulary knowledge predicts both the accuracy and speed of word reading, including decodable and irregular words (Kim et al., 2013; Ouellette, 2006; Ricketts et al., 2007). Kim et al. (2013), for instance, found that semantic knowledge uniquely contributed to word reading accuracy after accounting for phonological awareness, alphabet knowledge, and letter-writing automaticity. Similarly, Ricketts et al. (2007) demonstrated that children with stronger vocabulary knowledge showed more accurate reading of words that do not follow regular letter-sound patterns, suggesting that well-specified semantic representations support recognition beyond what phonological decoding alone affords. These

findings indicate that robust lexical representations built through exposure to frequent and contextually diverse words facilitate both comprehension and efficient word recognition.

Reading Rate

Vocabulary supports reading rate by enabling the automatic word recognition necessary for fluent processing. Encountering unfamiliar words requires readers to pause and decode, disrupting flow (LaBerge & Samuels, 1974). Although fluent reading does not require knowing every word in a text, the proportion of unfamiliar words must remain low enough to avoid frequent interruptions. When this proportion increases, cognitive resources shift from comprehension to word identification. Words that are less well established in the reader's lexicon therefore demand greater effort, slowing reading rate and constraining meaning-making.

In a review of text difficulty effects on elementary students' reading, Amendum et al. (2017) found that most studies reported slower reading rates as text difficulty increased, particularly for younger and less skilled readers; studies showing no relationship involved more skilled or older elementary students. This pattern reflects the greater processing demands associated with weak lexical representations: readers with less developed word recognition systems are more susceptible to fluency disruptions as text difficulty increases. Mesmer (2009) similarly demonstrated that vocabulary familiarity shapes reading rate. First-grade students read leveled texts (i.e., commercial books organized by difficulty) and decodable texts (i.e., materials structured around phonetic patterns) under practiced and unpracticed conditions at three time points. Leveled texts contained more high-frequency words likely to be part of students' oral vocabularies, whereas decodable texts included orthographically regular but less familiar words (e.g., *jug*, *vat*). Across conditions, students read leveled texts at faster rates. Together, these

findings indicate that vocabulary familiarity supports the automatic processing necessary for fluent reading in beginning readers.

Comprehension

Finally, vocabulary shapes comprehension by enabling efficient meaning construction and supporting the integration of ideas within a text. Beyond its role in facilitating word recognition, vocabulary provides the semantic precision necessary to generate inferences, connect prepositions, and construct a coherent situation model. When readers encounter unfamiliar or weakly specified words, limitations in meaning, not merely slower decoding, can disrupt integration and weaken coherence. Even when words are decoded accurately, imprecise lexical representations may obscure causal connections and distort relationships among events.

Cunningham and Stanovich (1997) demonstrated that first-grade vocabulary breadth predicts eleventh-grade comprehension, reflecting a self-reinforcing cycle: vocabulary knowledge shapes which texts children can access, and those texts in turn determine opportunities for further growth. Breadth supports automatic processing; depth supports interpretive precision. Texts aligned with readers' lexical knowledge therefore promote both fluent processing and coherent understanding. For these reasons, vocabulary demands are not incidental in research on beginning reading. Aligning texts with readers' lexical knowledge supports efficient word recognition and the semantic precision required for comprehension, shaping whether first graders can construct meaning from print.

Measuring Vocabulary in Text Complexity Systems

Two systems commonly guide text selection in U.S. schools: the Lexile Framework and Guided Reading Levels (GRLs). The Lexile Framework assigns quantitative measures based on mean sentence length (MSL) and Mean Log Word Frequency (MLWF), a log-transformed

frequency metric that captures the overall vocabulary load of a text. MLWF is grounded in a fundamental principle of language acquisition: words that occur more frequently in a language are typically learned earlier and are more accessible to readers (Nation, 2008). By calculating the logarithmic mean of word frequencies across a text, MLWF provides a stable estimate of overall vocabulary load. Lennon and Burdick (2004) reported that after 20 years of research comparing MLWF against approximately 50 alternative semantic and linguistic variables (i.e., measures that variously captured word features such as concreteness, morphological complexity, and semantic relationships), MLWF consistently outperformed these indices in predicting text difficulty.

In contrast, the Guided Reading Level (GRL) system (Fountas & Pinnell, 2006) relies on qualitative judgment across ten criteria. One criterion is vocabulary, which is evaluated through indicators such as word frequency, word patterns (e.g., letter–sound correspondences), and support provided by illustrations. However, the relative weight of each criterion has not been empirically reported. Analyses of leveled texts indicate that MLWF does not systematically differentiate among GRL levels (Authors, 2022a), suggesting that vocabulary demands may vary within texts assigned the same designation. By holding GRL constant while varying MLWF, the present study isolates vocabulary complexity as an independent dimension of text difficulty. If students perform differently on texts assigned to the same level but differing in MLWF, such findings would indicate that vocabulary demands are not fully captured by existing leveling systems and warrant more systemic consideration in text selection.

Word Features that Influence Recognition of Word Meaning

Although quantitative and qualitative text-level metrics such as MLWF estimate overall vocabulary load, they provide limited insight into the specific lexical features that may pose challenges for beginning readers. To better understand variation in word-level demands, we build

on the categorization proposed by Authors (2022b) and examine three types of features likely to influence word recognition: structural, semantic, and contextual.

Structural Features

Structural features are intrinsic properties of words that remain stable across contexts. For example, a word's spelling remains the same whether it appears in a first-grade text or a scientific journal. We examine two structural dimensions: word length, indexed by syllables, phonemes, and morphemes, and orthographic consistency. Word length reliably influences reading accuracy and rate in the elementary grades, both in isolation and in connected text (Chetail et al., 2015; Gagl et al., 2015; Juel & Roper-Schneider, 1985). Orthographic consistency also shapes recognition. Although readers benefit from large orthographic neighborhoods (Andrews, 1997), English's quasi-regular orthography introduces variability, particularly in vowel correspondences (Treiman et al., 1995), which can increase decoding effort and slow access to meaning (Perfetti, 1992). To capture this variability, we include measures of feed-forward (spelling-to-sound) and feed-backward (sound-to-spelling) consistency (Chee et al., 2020), which quantify the reliability of orthographic–phonological mappings.

Semantic Features

Semantic features capture the conceptual properties of words that shape access to meaning. We examine two semantic dimensions: concreteness and polysemy. Concreteness distinguishes words referring to tangible entities (e.g., *apple*) from abstract concepts (e.g., *freedom*) (Paivio, 2014). Concrete words are processed more quickly and accurately than abstract words (Ding et al., 2023), and young readers show lower accuracy for abstract vocabulary in lexical decision and naming tasks (McFalls et al., 1996). Abstract words are also harder to infer from context and retain over time, with mastery developing gradually across the elementary

grades (Caramelli et al., 2004). Polysemy refers to words with multiple related meanings (e.g., *bank*, *light*). Although acquiring additional senses may initially challenge younger readers (Authors, 2015), research consistently demonstrates faster recognition of polysemous words than unambiguous words (Gottlob et al., 1999), likely reflecting richer lexical representations (González-Fernández & Schmitt, 2020). Together, these semantic properties may influence how efficiently beginning readers access and integrate word meaning.

Contextual Features

Contextual features reflect how words are distributed across language use and vary across texts. We examine four contextual variables that shape word learning and recognition: frequency, dispersion, trajectory, and age of acquisition (AoA). Word frequency is among the most robust predictors of reading performance, particularly for beginning readers (Share et al., 1984). High-frequency words are recognized more quickly and accurately and require less cognitive effort in connected text (Altani et al., 2020; Juhasz et al., 2019). Although frequency does not distinguish among multiple meanings of a word, it serves as a proxy for exposure and lexical consolidation (Kirby et al., 2008). Dispersion indexes how evenly a word appears across domains (Carroll et al., 1971); high-dispersion words provide repeated exposure across contexts, strengthening orthographic and semantic representations (Lawrence et al., 2022), whereas low-dispersion words are concentrated in specialized contexts and may be less accessible to young readers.

Trajectory captures how a word's frequency changes across grade levels (Juhasz et al., 2019). Some words increase in prevalence in later grades, others remain stable, and some decline, indicating whether a word is foundational in early texts or becomes important later. Age of acquisition refers to when a word typically enters children's oral vocabularies (Kuperman et al., 2012). Words acquired earlier are recognized more efficiently in reading, even after

controlling for frequency (Juhasz et al., 2019; Zevin & Seidenberg, 2002), whereas words outside students' oral vocabularies may slow access to meaning (Perfetti & Hart, 2002). These contextual properties capture distributional patterns that may facilitate or constrain word recognition in beginning readers.

The Present Study

Building on prior theoretical and empirical work, this study examines how variation in vocabulary complexity, as measured by MLWF, relates to first-grade students' reading accuracy, rate, and comprehension. Guided by the Lexical Quality Hypothesis, we analyzed texts assigned to the same GRL (Level J) while allowing MLWF to vary to isolate vocabulary's independent contribution to these outcomes. We addressed three research questions: (1) How does MLWF relate to reading accuracy, rate, and comprehension when text level is held constant? (2) How does the association between MLWF and reading performance differ across levels of proficiency? (3) Which structural, semantic, and contextual word-level features predict higher error rates during oral reading?

Methods

Participants

Study participants included seventy-eight first-grade students from four elementary schools in the midwestern United States. The sample was 45% male; 85% of students identified as white, 5% as Latinx, 5% as Black, 3% as Indigenous, and 2% as two or more races. Five percent of students received English as a Second Language services, and 64% qualified for free or reduced-priced lunch. Sixty percent met spring first-grade DIBELS Oral Reading Fluency (ORF) benchmark scores (University of Oregon Center on Teaching & Learning, 2020). DIBELS benchmark scores represent empirically established performance targets for early

literacy skills; students meeting benchmarks are considered on track for grade-level reading proficiency, whereas those scoring below benchmark are identified as at-risk and recommended for instructional support. We collected all reading data in the spring of the academic year.

Procedures

We developed 18 passages organized into three parallel sets of six. We assigned each student to one set, and students completed the six passages within that set as oral reading tasks and the same six passages as maze tasks. Each of the three passage sets was completed by approximately one-third of the sample. We counterbalanced passage sets across students, and further counterbalanced passages to control for potential order and practice effects. This design ensured that each passage and its constituent words were read by students with a range of reading abilities. The final dataset included 462 oral reading passages and 437 maze passages.

For oral reading fluency, students read each passage aloud for one minute while researchers recorded the number of words read correctly and documented all errors, including mispronunciations, substitutions, and omissions. Two researchers independently scored 15% of the oral readings to assess inter-rater reliability. Agreement for words correct per minute (WCPM) was high (ICC = .95), indicating strong consistency in scoring. For the maze assessment, students completed passages in a group setting within their classrooms. During this timed three-minute task, students selected the correct word from three options to complete each sentence. We scored responses using a fixed answer key, awarding one point for each correct answer; total scores reflect the number of correct responses.

Measures

External ORF Scores

Classroom teachers provided spring DIBELS ORF scores (University of Oregon Center

on Teaching & Learning, 2020), collected as part of the school's regular literacy assessments. These scores served as an external measure of students' fluency levels and were included as a covariate to account for individual differences in reading proficiency. Controlling for students' baseline ORF allowed us to assess the unique contributions of MLWF to students' accuracy, rate, and comprehension beyond what baseline fluency explains.

Internal ORF and Maze Scores

We used the same texts for the ORF and maze assessments. To select and prepare the texts, we (1) ensured equivalence across GRL, genre, and Lexile level; (2) established graduated vocabulary levels; (3) designated sections of text; and (4) constructed the maze passages. We explain each one in greater detail below.

Ensuring Equivalence Across GRL, Genre, and Lexiles. To isolate vocabulary as the sole independent variable, we systematically controlled other textual characteristics, including GRL, genre, and Lexile level. To do so, we selected narrative texts from Level J of a commercial leveled text program not used in the district. By focusing exclusively on narrative texts, we minimized genre-based variability, as prior research indicates that genres differ in linguistic and conceptual complexity (e.g., Authors, 2019).

As shown in Table 1, we maintained Lexile scores within a constrained range across six passages (413–463), placing them at the end-of-first-grade band on the Common Core State Standards (CCSS; National Governors Association Center for Best Practices [NGA Center] & Council of Chief State School Officers [CCSSO], 2010) text complexity staircase (Nelson et al., 2012). To achieve comparable Lexile levels, we adjusted sentence length when necessary. Authors (2018) demonstrated that sentence length can disproportionately influence Lexile scores relative to MLWF. For example, the original version of *The Disappearing Moon* had a Lexile of

360. When we combined two short sentences ("He could not sleep. He could not eat.") into a single sentence ("He could not sleep and could not eat."), the Lexile increased to 380, an increase driven solely by syntactic manipulation. Given that texts composed primarily of simple sentences can impede comprehension (Pearson, 1974), and because passages with higher vocabulary bands naturally contained more complex sentence structures, we standardized sentence complexity across passages. This control allowed us to attribute differences in comprehension to vocabulary difficulty rather than syntactic variation.

Graduated Vocabulary Levels. To characterize vocabulary difficulty within Level J texts, we conducted Lexile analyses on 100 texts from the Leveled Literacy Intervention (LLI), which the district used for their benchmark assessments. Ninety-seven percent of these texts fell within a MLWF range of 3.3 to 3.8. We therefore defined six graduated vocabulary bands within this range (3.3, 3.4, 3.5, 3.6, 3.7, and 3.8) and selected study texts to represent each point. We differentiated texts at six MLWF values because the Lexile Framework operates on a logarithmic scale, such that even small numerical increases reflect substantial shifts in vocabulary complexity. For example, an increase from 3.3 to 3.4 corresponds to approximately a 26% decrease in word frequency (Lennon & Burdick, 2004). The selected range (3.3–3.8) therefore captures meaningful and progressively more demanding vocabulary likely to influence students' reading performance.

To avoid familiarity effects, we drew study texts from Reading A to Z, a program not used in the district. We transcribed and analyzed all 43 Level J texts available at the time using the Lexile Analyzer, which provides MLWF, mean sentence length (MSL), and Lexile scores. Three texts fell outside the target MLWF range, leaving 40 texts with MLWFs from 3.3 to 3.86. Within each vocabulary band, we selected texts from the midpoint of the distribution to represent

that band. For example, when five texts fell between 3.72 and 3.79, we selected the three central values to obtain an average MLWF of 3.75. When multiple texts shared similar MLWF values, we selected among them using a random numbers table. We intentionally selected and characterized passages by their MLWF values to represent a fixed range of lexical demands within Level J narrative texts. They were not intended to constitute a random sample of texts, and inferences are therefore limited to the passages included in the study.

Designating Sections of Text. To create passages, we selected 100–107 consecutive words from each text. We based this length on ORF norms indicating that only a small proportion of first-grade students read more than 100 words per minute (Hasbrouck & Tindal, 2017). Because we calculated WCPM using actual reading time, this design also accommodated students who read faster than 100 words per minute. Prior research supports the use of 100-word passages for silent reading tasks at this grade level (Authors, 2015).

We selected passages that represented a coherent event or description within the larger text. For example, the original version of *Rent a Llama* (433 words) describes a child's transition from a llama ranch in Ecuador to living in the United States; the 100-word excerpt used in this study depicts only the child's interaction with a llama during a school visit to an American ranch. We verified that each excerpt fell within the target MLWF and MSL ranges and made minimal edits when necessary to maintain alignment (e.g., removing animal nicknames in *Rent a Llama* to adjust vocabulary frequency while preserving meaning). Words with varying lexical and orthographic characteristics (e.g., frequency, age of acquisition, consistency) appeared throughout each 100-word excerpt.

We calculated accuracy as the percentage of words read correctly by dividing the number of correct words (words attempted minus errors) by the total number of words attempted;

accuracy was independent of time and reflects the proportion of words read correctly out of those attempted. We calculated rate as words correct per minute (WCPM). For students who read for the full one-minute interval, WCPM reflected the total number of words read correctly during that minute. When a student completed a passage in less than one minute, we divided the total number of words read correctly by the time required (in seconds) and multiplied by 60 to estimate words correct per minute.

Designing the Maze. We constructed maze passages using the same texts as the oral reading task. We left the first sentence intact and then replaced every seventh word with a three-option multiple-choice item. For each item, we matched distractors to the target word in length but selected words with higher frequency values (i.e., more common words). All distractors had predicted frequencies of at least 500 occurrences per million (U function ≥ 500; Zeno et al., 1995). For example, in the sentence "The rain came down hard as the wind howled," we paired the target word *hard* (U = 381) with the distractors *food* (U = 570) and *part* (U = 694). We reviewed all distractors to ensure that none constituted a plausible response. We calculated comprehension (i.e., maze score) as the total number of correct responses per passage.

Word-Level Features

We obtained structural, semantic, and contextual word-level variables from the South Carolina Psycholinguistic Metabase (SCOPE), a large-scale lexical database that aggregates psycholinguistic and corpus-based indices of word characteristics. These variables allowed us to examine how intrinsic and distributional properties of words may contribute to variation in students' oral reading errors. Structural measures included word length (letters, phonemes, morphemes, and syllables) and grapheme–phoneme consistency (feed-forward and feed-backward). Following Chee et al. (2020), we used token consistency values for these variables.

Feed-forward (spelling-to-sound) consistency reflects the proportion of a word's log frequency accounted for by "friends" (i.e., words with matching spelling–sound patterns) relative to the combined log frequencies of friends and "enemies" (i.e., words with conflicting patterns); feed-backward (sound-to-spelling) consistency reflects the same ratio in the reverse direction. Values range from 0 to 1, with higher values indicating more consistent grapheme–phoneme mappings.

Semantic variables included polysemy and concreteness (Brysbaert et al., 2014), and contextual variables included dispersion, trajectory, and age of acquisition (Kuperman et al., 2012). We obtained frequency values from the Zeno et al. (1995) database, a corpus-based index derived from K–12 reading materials that provides word frequency estimates widely used in research on word recognition and text difficulty.

Results

Descriptive Statistics

On average, students read passages at 62.9 WCPM with 90.4% accuracy and correctly answered 5.8 maze prompts (Table 1). According to Hasbrouck & Tindal's (2017) updated norms, first-grade students read an average of 60 WCPM on grade-level texts in the spring, suggesting this group performed near the national average. The 90% accuracy rate indicates participants were reading at the lower end of the instructional level range (Betts, 1946; Morris, 2013). Students answered approximately 48% of maze items correctly (out of 11–13 items per passage), placing performance near the lower bound of instructional level expectations (Betts, 1946; Morris et al., 2011). However, scores remained well above chance performance for a three-option task (33%), suggesting that students were engaging meaningfully with the passages.

When examined by passage, half of the texts yielded accuracy rates within the instructional range, while the other half fell within the frustration range, with accuracy rates

ranging from 84.6% to 95.6%. WCPM scores ranged from 38.6 to 78.5, with approximately one-third of the passages falling below 60 WCPM. Maze scores ranged from 4.38 to 8.71, with most passages averaging fewer than six correct responses.

RQ1: Effects of MLWF on Accuracy, Rate, and Comprehension

We used multilevel modeling to examine whether MLWF predicted students' reading accuracy, rate (WCPM), and comprehension while accounting for the dependence among multiple readings by the same student. We conducted analyses in HLM-7 (Raudenbush et al., 2011), with passage readings (Level 1) nested within students (Level 2). Because passages were intentionally selected to represent a fixed range of MLWF values rather than a random sample of texts, MLWF was modeled as a fixed passage-level predictor. Students' baseline reading ability (DIBELS ORF) was included as a Level-2 covariate (Raudenbush & Bryk, 2002).

Unconditional two-level models with random intercepts for students indicated substantial clustering by student. Intraclass Correlation Coefficients (ICCs) were .79 for accuracy, .89 for WCPM, and .72 for comprehension (see Table 2), indicating that 79%, 89%, and 72% of the variance in each outcome was attributable to differences between students in these models. The remaining variance reflected within-student differences across passage readings, including differences associated with the fixed lexical characteristics of the passages. Because all ICCs exceeded conventional thresholds (e.g., .05), multilevel modeling with a random intercept for students was warranted to avoid biased standard errors and inflated Type I error rates.

After controlling for baseline reading ability, a greater proportion of the remaining variance was associated with passage-level characteristics (*accuracy* = 44%; *WCPM* = 65%; *comprehension* = 75%), indicating meaningful variation across the passages beyond students' general reading skill. Within this two-level framework, MLWF significantly predicted all three

outcomes ($p < .05$) and accounted for 12% of the variance in accuracy, 21% in WCPM, and 1% in comprehension associated with differences across the passages included in the study. Model fit indices (AIC and BIC) indicated improved fit for models including MLWF when predicting accuracy and WCPM, compared to models including only baseline ability (see Table 3).

RQ2: MLWF Effects by Reading Proficiency Level

We used quantile regression to examine whether the association between MLWF and reading performance varied across the distribution of each outcome, controlling for baseline DIBELS ORF. Given the modest and relatively homogenous sample, these analyses were exploratory in nature and intended to examine potential variation in effects rather than to establish definitive subgroup differences. Unlike mean-based models, quantile regression estimates conditional effects at different points of the outcome distribution using the full sample, allowing us to explore heterogeneity in vocabulary sensitivity. Table 4 reports unstandardized slope estimates and confidence intervals at the .2, .5, and .8 quantiles for each outcome.

Accuracy

MLWF significantly predicted reading accuracy at all quantiles ($p < .001$), with similar pseudo R^2 values across levels (.018–.025). Although the proportion of explained variance was comparable, the magnitude of the MLWF coefficient differed across the distribution. The estimated effect was larger at the .2 and .5 quantiles than at the .8 quantile, suggesting that students with lower and mid-range accuracy may be more sensitive to increases in rare vocabulary than students performing at the highest levels of accuracy (Figure 1). These patterns indicate potential heterogeneity in vocabulary effects across proficiency levels; however, estimates at the tails of the distribution should be interpreted cautiously given reduced precision.

Words Correct Per Minute (WCPM)

MLWF significantly predicted WCPM for students below the 80th percentile ($p<.001$), but the association was not statistically significant at the highest quantile (see Figure 2). The estimated effect size was largest at the median quantile, suggesting that students reading at average fluency levels (i.e., approximately 70–100 WCPM) may experience greater reductions in rate as vocabulary demand increases than students at the highest fluency levels. As with accuracy, however, tail estimates should be interpreted as exploratory rather than definitive.

Comprehension

MLWF significantly predicted maze performance across all quantiles ($p<.001$), although effect sizes remained small (pseudo $R^2 = .003–.006$). Students above the 20th percentile showed slightly larger parameter estimates than those at the lowest quantile (Figure 3). Given the relatively restricted range of maze scores, these effects should be interpreted cautiously, as the structure of the maze task may have limited sensitivity to detect larger vocabulary-related effects.

RQ3: Word-Level Features Associated with Reading Errors

To examine which word features were associated with reading errors, we fit a binomial generalized linear model. Each word served as the unit of analysis. Because students were assigned to one of three balanced passage sets, exposure to individual words varied across students by design rather than solely as a function of reading rate. For every word, we recorded (a) the total number of student attempts and (b) the number of incorrect or omitted readings. These counts were used to model the proportion of errors for each word. Predictors included structural features (number of syllables, phonemes, morphemes, feed-forward consistency, feed-backward consistency), semantic properties (concreteness, polysemy), and contextual characteristics (trajectory, age of acquisition, dispersion, and frequency). We estimated all

predictors simultaneously to examine the unique association of each word-level feature with error rates while accounting for shared variance among correlated lexical characteristics. Because many psycholinguistic variables are interrelated (e.g., frequency, age of acquisition, and consistency), simultaneous modeling allows for identification of predictors that explain variance beyond overlapping lexical dimensions. Diagnostic checks indicated no evidence of overdispersion, supporting the use of a standard binomial model. This approach allowed us to estimate the odds that a word would be misread while accounting for differences in exposure.

As shown in Table 5, three predictors accounted for unique variance in error rates when considered simultaneously ($p < .05$). First, number of phonemes was positively associated with error rates; words containing more phonemes were more likely to be misread. The estimated odds ratio (OR = 1.21) indicates that each additional phoneme was associated with higher odds of an error by approximately 21%, suggesting that greater phonological complexity increases decoding demands on beginning readers.

Second, feed-forward consistency was negatively associated with error rates; words with more predictable spelling-to-sound correspondences were significantly less likely to be misread. The estimated odds ratio (OR = 0.12) indicates that higher consistency was associated with substantially lower odds of error (approximately an 88% reduction), highlighting the importance of consistent orthography–phonology mappings for accurate decoding.

Finally, age of acquisition (AoA) was positively associated with error rates. Words learned later in development were more likely to be misread than words learned earlier in development. The estimated odds ratio (OR = 3.24) suggests that later-acquired words had more than three times the odds of error compared to earlier-acquired words, indicating that developmental timing of word learning contributes to word recognition accuracy. It is important

to note that nonsignificant predictors in this model should not be interpreted as unrelated to reading accuracy; rather, they did not account for variance beyond that shared with other correlated lexical characteristics. Although the balanced passage-set design reduced systematic confounding between reader skill and word exposure, the word-level aggregation does not fully disentangle reader-, passage-, and word-level sources of variance. In particular, the number of attempts per word reflects both passage assignment and individual differences in reading rate. Accordingly, these findings should be interpreted as estimating associations within this design rather than as fully isolated word effects independent of reader characteristics.

Discussion

The question of which words influence which children in beginning reading represents a fundamental challenge in literacy education. When texts contain too many difficult words relative to a reader's current level of automaticity, cognitive overload may reduce resources available for comprehension. Conversely, when texts provide too little lexical challenge, opportunities for vocabulary growth may be constrained. Understanding the features that define challenging words enables educators to make more informed decisions about text selection, vocabulary instruction, and differentiated support. These questions are urgent given persistent achievement gaps and recognition that early reading experiences set lifelong literacy trajectories. As such, this study examined how MLWF influences reading accuracy, rate, and comprehension; how students with different profiles respond to texts with varying vocabulary demand; and the specific word characteristics that may be associated with reading errors. While we acknowledge that corpus frequency imperfectly represents individual familiarity with specific words, the Lexile Framework's validation against standardized vocabulary assessments provides evidence that MLWF correlates with actual vocabulary knowledge in elementary readers.

Vocabulary Difficulty Significantly Affects Reading Performance

Our analysis revealed that MLWF significantly predicted students' reading accuracy, rate, and comprehension performance. The 12% of variance in accuracy associated with differences across the passages represents a substantial effect that translates into meaningful differences in classroom performance. To contextualize this finding, the range of MLWF values in our study (3.3 to 3.8) corresponds to notable shifts in vocabulary difficulty, with higher values indicating less frequent words in the text. The practical consequence of this shift was associated with a 7 to 11 percentage point difference in reading accuracy as MLWF decreased from 3.8 to 3.3—a difference that moves students from independent reading levels (above 95% accuracy) to frustration levels (below 90% accuracy). This accuracy decline occurred despite texts carrying identical "Level J" designations.

The impact on reading rate is particularly noteworthy, as MLWF accounted for 21% of the variance associated with differences across the passages. The practical magnitude of this effect was substantial; students reading texts at MLWF = 3.3 (i.e., harder vocabulary) averaged approximately 15-20 fewer WCPM compared to texts at MLWF = 3.8 (i.e., easier vocabulary). This represents a difference equivalent to moving from grade-level fluency expectations (around 60 WCPM for spring first-grade) to below-benchmark performance.

The relatively modest impact on comprehension (1% of text-level variance) should be interpreted cautiously. Although mean maze performance was above chance levels for a three-option task (33%), scores clustered near the lower bound of the instructional range. This restricted range may have limited sensitivity to detect larger comprehension effects, suggesting that vocabulary influences on understanding may be detectable but modest under the current measurement conditions.

Impact Varies by Readers' Proficiency Levels

Perhaps our most educationally significant finding concerns the differential impact of vocabulary difficulty across students with varying proficiency levels. Exploratory quantile regression analyses suggest that students with lower reading proficiency may show greater sensitivity to vocabulary difficulty than their higher-performing peers. For reading accuracy, the estimated MLWF effect at the lower quantiles was substantially larger than at the upper quantile, suggesting potential heterogeneity in vocabulary sensitivity across proficiency levels.

The pattern of estimates across quantiles suggests potential differentiation along an automaticity spectrum rather than discrete subgroups. Students at the highest levels of fluency appeared less sensitive to variation in MLWF, whereas students at mid-range fluency levels showed the largest estimated effects of vocabulary difficulty. Students at the lowest levels of automaticity also demonstrated sensitivity to MLWF, although these estimates were less stable. One possible interpretation of this pattern is the presence of a "vulnerability window" in reading development, during which vocabulary complexity may become especially consequential for decoding efficiency. This window may occur when students have acquired sufficient decoding skill to attempt unfamiliar words but have not yet developed the automaticity needed to process them efficiently.

Effects of MLWF on comprehension were consistent across skill levels but were small in magnitude. As discussed in the Limitations, this apparent uniformity may partly reflect characteristics of our comprehension measure, as the maze format's cumulative cognitive demands could have reduced sensitivity to detect stronger vocabulary–comprehension relationships across proficiency levels. Accordingly, conclusions about comprehension should be interpreted in light of the study's design and measures.

Phonemic, Orthographic, and Developmental Influences on Word Reading Errors

To understand what makes words more likely to be misread, we examined linguistic characteristics associated with greater word-level error rates. Our analysis identified three key predictors: phonemic segment load (number of phonemes), spelling-to-sound consistency (feed-forward consistency), and age of acquisition.

Words with more phonemes were significantly more error-prone. Following Nickels and Howard (2004), we operationalize this feature as phonemic segment load, or the number of phonemic segments in a word. Our findings confirm that phonemic segment load imposes cognitive demand on beginning readers, particularly those consolidating decoding skills. This aligns with research showing that phoneme numbers affect segmentation and blending demands (e.g., Webber et al., 2024) and complements the finding that students with weaker decoding accuracy are especially sensitive to word-level difficulty. However, the 21% increase in error likelihood per additional phoneme should not be interpreted to mean that students developing decoding automaticity should read only three-letter words with one-to-one correspondences. Three-phoneme words in our study texts spanned consonant and vowel patterns, including complex vowels (e.g., *soon*), multisyllable words (e.g., *able*), and words where two letters represent a single phoneme (e.g., the “ea” and “ch” in *teach*). Overall, however, more phonemes increased error likelihood for students in average and least accurate groups.

Orthographic patterns also influenced student errors. Despite having only two phonemes (/ð/ /oʊ/), *though* was among the most error-prone words, reflecting its low feed-forward consistency (.29). This word has few friends and numerous enemies among high-frequency words (e.g., *enough, through*). By contrast, words with higher feed-forward consistency were significantly less likely to be misread. Predictable spelling patterns reduced errors by 88%, one

of the strongest effects observed, highlighting how irregular spellings can interfere with decoding even when students possess adequate decoding skills.

Finally, later-acquired words proved significantly more error-prone, with incorrect readings occurring at triple the rate of earlier-acquired vocabulary. This suggests that developmental timing of word learning creates lasting impacts, as earlier-learned words become more deeply integrated into linguistic systems. The robustness of age of acquisition as a predictor is noteworthy given that other word-level variables showing effects in previous research (e.g., number of syllables, concreteness, frequency) failed to account for variance in reading outcomes. These null findings could be specific to narrative texts in our study, where variation in word features may have been insufficient to reveal effects. As evidence, only 3% of words contained more than two morphemes; among this small group of 22 words, 15 were compound forms (e.g., footsteps, newspapers) or words with agentive suffixes (e.g., shoppers, workers). These morphological forms likely pose fewer challenges for young readers than words with complex affixes. Distributional properties like dispersion and frequency may be less influential when students are still developing basic decoding skills and rely heavily on phonological processing.

Implications for Educational Leaders

This investigation of how vocabulary affects students' reading accuracy, rate, and comprehension has implications for educational administrators, state departments of education, and policy decision-makers charged with selecting texts, setting mandates, and allocating resources for instructional materials. This research addresses a critical gap in text selection practices that has gained urgency as pandemic-related learning loss and heightened attention to the science of reading have brought critiques of text leveling systems into mainstream discourse.

Building on our earlier work demonstrating that GRL systems do not adequately account for word-level difficulty (Authors, 2022a), the present findings provide additional empirical evidence of limitations in leveled text systems. Despite identical "Level J" designations, some texts supported independent reading performance while others resulted in frustration-level accuracy (Morris, 2013), revealing substantial inconsistencies in vocabulary demands that current leveling systems fail to capture. More concerning, quantile regression analyses suggest that this inconsistency may disproportionately affect struggling readers who most need appropriately matched texts. Our word-level analysis helps explain why assessment-to-text matches may fail: the linguistic features that contributed to reading errors in our sample are not captured by current leveling criteria.

Our findings have specific implications for state and district leaders responsible for text adoption decisions. First, selection systems should be evaluated not merely by their categorical labels but by the consistency of word-level linguistic features within each designated level. Second, when texts are classified at the same reading level, educators should examine whether they contain comparable phonemic complexity, orthographic consistency, and developmental word familiarity— dimensions not currently reflected in most leveling systems. Third, advances in corpus analysis and lexical databases now make it feasible to select or develop materials with tighter control over these features, an approach that may be especially beneficial for struggling readers who require greater consistency in instructional supports. Rather than endorsing a particular text type, we recommend that educational leaders prioritize precision in text selection grounded in measurable linguistic features.

Implications for Classroom Teachers

These findings also suggest practical applications for classroom instruction. Rather than

relying exclusively on categorical reading levels, teachers may benefit from examining specific word-level features when selecting texts, particularly avoiding clusters of difficult-to-decode words that could overwhelm developing readers. Students within the proposed “vulnerability window” may warrant particular attention, with accuracy levels serving as a useful guide for instructional differentiation. Students reading below approximately 94% accuracy may benefit from texts that include carefully selected challenges paired with strong instructional support, whereas students reading above 98% accuracy are likely to tolerate greater vocabulary demands without compromising comprehension. Monitoring error patterns can further inform instruction, as words with irregular spellings or greater phonemic complexity may consistently disrupt decoding for beginning readers. Attention to these patterns can support more targeted instruction and more precise text matching.

Limitations

While this study provides important evidence about how specific linguistic features affect beginning readers' performance, several methodological constraints shape the scope and generalizability. First, sample homogeneity limits generalizability of findings to more diverse student populations. With 85% of participants identifying as white, the sample underrepresents students of color, particularly Latinx and Black students. This racial and ethnic homogeneity raises questions about whether the linguistic features associated with reading difficulty operate similarly for students from culturally and linguistically diverse backgrounds, including multilingual learners and students whose home languages differ from English. Additionally, while 64% of students qualified for free and reduced-price lunch, suggesting socioeconomic diversity, the racial homogeneity within this economically diverse sample limits our ability to disentangle effects of socioeconomic status from cultural and linguistic background. Research

examining how vocabulary difficulty affects reading performance in more demographically representative samples would strengthen confidence in the generalizability and practical applicability of these findings across diverse school populations.

Second, our methodological decision to focus on first-grade students reading Level J narrative texts at a single time point allowed us to isolate the relationship between vocabulary and reading outcomes but necessarily limits generalizability. Because passages were intentionally selected to represent a fixed range of MLWF values within one text level and genre, findings should be interpreted as applying to this set of passages rather than to all beginning-reader texts. Moreover, data collected at a single end-of-year time point represents a developmental snapshot rather than a stable pattern, as students' sensitivity to vocabulary demands likely evolves with instruction, practice, and cognitive maturation. Future research spanning multiple grade levels, time points, and text types, including informational texts with distinct vocabulary demands, would strengthen understanding of how vocabulary difficulty operates across developmental stages and instructional contexts.

A third limitation concerns our comprehension assessment. Although decoding accuracy was high and maze performance exceeded chance levels, scores clustered near the lower bound of the instructional range, restricting variability and limiting sensitivity to detect stronger vocabulary-related effects. Thus, the relatively modest variance explained by MLWF should be interpreted cautiously. The maze format required students to process an extended passage while responding to deletions every seventh word, combining passage-level comprehension with repeated decision-making demands. For emerging readers, this cumulative cognitive load may have constrained performance variability. In contrast, Kim et al. (2021) used a single-sentence semantic association task with strong reliability (α = .85 for taught words; .77 for untaught

words), suggesting that more localized formats may isolate vocabulary-related understanding with greater sensitivity. Future research should compare passage-level and sentence-level assessments to determine how task structure influences detection of vocabulary effects.

Finally, this study examines first-grade texts selected in alignment with the CCSS and evaluated using U.S.-based metrics including DIBELS benchmark scores and GRLs to indicate text complexity. Although some of these tools, such as GRLs, have been adopted internationally, this combination of standards, assessment frameworks, and complexity metrics reflects distinctly American approaches to text selection and reading evaluation. At the same time, the central question guiding this analysis—whether vocabulary demands placed on beginning readers are appropriately aligned with their developmental capacities—extends beyond national boundaries. Educators across contexts must make decisions about text difficulty and vocabulary load for early readers. While our findings are grounded in the U.S. context shaped by the CCSS, the relationships between vocabulary difficulty and first graders' reading accuracy, rate, and comprehension may inform similar investigations conducted under different standards, assessment systems, or text complexity frameworks. Future research examining vocabulary demands in beginning-reader texts across varied educational contexts would deepen understanding of how text selection practices influence early literacy development worldwide.

Conclusion

Matching students with appropriately challenging texts has long been recognized as fundamental to effective reading instruction, yet current approaches have remained relatively imprecise. For decades, educators have relied on Betts' (1946) framework classifying texts by accuracy percentages (i.e., independent, instructional, and frustration levels). However, concerns about the reliability of such classifications persist. For example, Burns et al. (2015) found that

only 28% of students read within expected accuracy ranges when given supposedly "matched" texts, with 58% of struggling readers performing at frustration level. Our findings provide additional evidence that GRL designations alone may not adequately capture meaningful differences in word-level demands.

The present study points toward a more precise approach to text selection grounded in linguistic analysis and supported by contemporary word databases and computational tools (Gao et al., 2020). Rather than relying solely on categorical labels, educators and researchers can examine multiple dimensions of word difficulty including average word frequency, phonemic segment load, orthographic consistency, and developmental familiarity to better understand how specific features shape reading performance for students with different skill profiles. Where earlier research asked whether reading inventories are reliable, we can now ask which word characteristics predict difficulty for particular readers. This shift from broad classification systems to fine-grained linguistic calibration offers the possibility of matching students to texts with greater precision, aligning vocabulary demands with learners' developing decoding skills and supporting more equitable access to appropriately challenging reading experiences.

References

Altani, A., Protopapas, A., Katopodi, K., & Georgiou, G.K. (2020). From individual word recognition to word list and text reading fluency. *Journal of Educational Psychology, 112*(1), 22–39.

Amendum, S.J., Conradi, K., & Hiebert, E. (2018). Does text complexity matter in the elementary grades? A research synthesis of text difficulty and elementary students' reading fluency and comprehension. *Educational Psychology Review*, *30*, 121-151.

Andrews, S. (1997). The effect of orthographic similarity on lexical retrieval: Resolving neighborhood conflicts. *Psychonomic Bulletin & Review*, *4*(4), 439-461.

Authors (2013). [Title omitted for blind peer-review]. [Journal omitted for blind peer-review].

Authors (2015). [Title omitted for blind peer-review]. [Journal omitted for blind peer-review].

Authors (2016). [Title omitted for blind peer-review]. [Journal omitted for blind peer-review].

Authors (2018). [Title omitted for blind peer-review]. [Journal omitted for blind peer-review].

Authors (2019). [Title omitted for blind peer-review]. [Journal omitted for blind peer-review].

Authors (2022a). [Title omitted for blind peer-review]. [Journal omitted for blind peer-review].

Authors (2022b). [Title omitted for blind peer-review]. [Journal omitted for blind peer-review].

Betts, E.A. (1946). *Foundations of reading instruction, with emphasis on differentiated guidance.* American Book Co.

Brysbaert, M., & Cortese, M.J. (2011). Do the effects of subjective frequency and age of acquisition survive better word frequency norms? *The Quarterly Journal of Experimental Psychology*, *64*(3), 545-559.

Brysbaert, M., Warriner, A.B., & Kuperman, V. (2014). Concreteness ratings for 40 thousand generally known English word lemmas. *Behavior Research Methods*, *46*, 904-911.

Burns, M.K., Pulles, S.M., Maki, K.E., Kanive, R., Hodgson, J., Helman, L.A., ... & Preast, J.L. (2015). Accuracy of student performance while reading leveled books rated at their instructional level by a reading inventory. *Journal of School Psychology*, *53*(6), 437-445.

Caramelli, N., Setti, A., & Maurizzi, D.D. (2004). Concrete and abstract concepts in school age children. *Psychology of Language and Communication, 8*(2), 19-34.

Carroll, J.B., Davies, P., & Richman, B. (1971). *The American Heritage word frequency book.* Houghton Mifflin Company.

Chee, Q.W., Chow, K.J., Yap, M.J., & Goh, W.D. (2020). Consistency norms for 37,677 English words. *Behavior Research Methods*, 52(6), 2535-2555.

Chetail, F. (2015). Reconsidering the role of orthographic redundancy in visual word recognition. *Frontiers in Psychology*, *6,* 645.

Cunningham, A.E., & Stanovich, K.E. (1997). Early reading acquisition and its relation to reading experience and ability 10 years later. *Developmental Psychology*, 33(6), 934-945.

Ding, J., Liang, P., Guo, X., & Yang, Y. (2023). The influence of conceptual concreteness on the reading acquisition and integration of novel words into semantic memory via thematic relations. *Frontiers in Psychology, 14*, 1132039.

Fountas, I.C., & Pinnell, G.S. (2006). *Leveled Books, K-8: Matching texts to readers for effective teaching*. Heinemann.

Gagl, B., Hawelka, S., & Wimmer, H. (2015). On sources of the word length effect in young readers. *Scientific Studies of Reading*, *19*(4), 289–306.

Gao, C., Shinkareva, S.V., & Desai, R.H. (2022). SCOPE: The South Carolina psycholinguistic metabase. *Behavior Research Methods*, *54*(1), 171-187.

González-Fernández, B., & Schmitt, N. (2020). Word knowledge: Exploring the relationships and order of acquisition of vocabulary knowledge components. *Applied Linguistics*, *41*(4), 481–505.

Gottlob, L.R., Goldinger, S.D., Stone, G.O., & Van Orden, G.C. (1999). Reading homographs: Orthographic, phonologic, and semantic dynamics. *Journal of Experimental Psychology: Human Perception and Performance*, *25*(2), 561–574.

Gough, P.B., & Tunmer, W.E. (1986). Decoding, reading, and reading disability. *Remedial and Special Education*, *7*(1), 6-10.

Hasbrouck, J., & Tindal, G. (2017). An Update to Compiled ORF Norms. Technical Report# 1702. *Behavioral Research and Teaching*.

Hoffman, J.V., Sailors, M., & Patterson, E.U. (2002). Decodable texts for beginning reading instruction: The year 2000 basals. *Journal of Literacy Research, 34*(3), 269-298.

Juel, C., & Roper-Schneider, D. (1985). The influence of basal readers on first grade reading. *Reading Research Quarterly*, *20*(2), 134-152.

Juhasz, B.J., Yap, M.J., Raoul, A., & Kaye, M. (2019). A further examination of word frequency and age-of-acquisition effects in English lexical decision task performance: The role of frequency trajectory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *45*(1), 82-96.

Kim, J.S., Burkhauser, M.A., Mesite, L.M., Asher, C.A., Relyea, J.E., Fitzgerald, J., & Elmore, J. (2021). Improving reading comprehension, science domain knowledge, and reading engagement through a first-grade content literacy intervention. *Journal of Educational Psychology*, *113*(1), 3.

Kim, Y.S., Apel, K., & Al Otaiba, S. (2013). The relation of linguistic awareness and vocabulary to word reading and spelling for first-grade students participating in response to intervention. *Language, Speech, and Hearing Services in School, 44*(4), 337-347.

Kirby, J.R., Desrochers, A., Roth, L., & Lai, S.S. (2008). Longitudinal predictors of word reading development. *Canadian Psychology*, *49*(2), 103.

Kuperman, V., Stadthagen-Gonzalez, H., & Brysbaert, M. (2012). Age-of-acquisition ratings for 30,000 English words. *Behavior Research Methods*, *44*(4), 978-990.

LaBerge, D., & Samuels, S. J. (1974). Toward a theory of automatic information processing in reading. *Cognitive Psychology*, *6*(2), 293-323.

Lawrence, J.F., Kulesz, P.A., & Francis, D.J. (2022). Reading comprehension and academic vocabulary: Exploring relations of item features and reading proficiency. *Reading Research Quarterly*, *57*(1), 197-218.

Lennon, C., & Burdick, H. (2004). *The Lexile framework as an approach for reading measurement and success* [White paper]. MetaMetrics. https://cdn.lexile.com/m/resources/materials/Lennon__Burdick_2004.pdf

McFalls, E.L., Schwanenflugel, P.J., & Stahl, S.A. (1996). Influence of word meaning on the acquisition of a reading vocabulary in second-grade children. *Reading and Writing*, *8*, 23250.

Mesmer, H.A. (2009). Textual scaffolds for developing fluency in beginning readers: Accuracy and reading rate in qualitatively leveled and decodable text. *Literacy Research and Instruction*, *49*(1), 20-39.

Morris, A.K. (2011). Creating shared instructional products: An alternative approach to improving teaching. *Educational Researcher*, *40*(1), 5-14.

Morris, D. (2013). *Diagnosis and correction of reading problems*. Guilford.

Nation, K. (2008). Learning to read words. *Quarterly Journal of Experimental Psychology*, *61*(8), 1121-1133.

National Governors Association Center for Best Practices and Council of Chief State School Officers. (2010). *Common core state standards*. NGAC and CCSSO. http://www.corestandards.org/read-the-standards/

Nelson, J., Perfetti, C., Liben, D., & Liben, M. (2012). *Measures of text difficulty: Testing their predictive value for grade levels and student performance*. New York: Student Achievement Partners.

Nickels, L., & Howard, D. (2004). Dissociating effects of number of phonemes, number of syllables, and syllabic complexity on word production in aphasia: It's the number of phonemes that counts. *Journal of Cognitive Neuropsychology*, 21(5), 535-556.

Ouellette, G.P. (2006). What's meaning got to do with it: The role of vocabulary in word reading and reading comprehension. *Journal of Educational Psychology*, *98*(3), 554-566.

Paivio, A. (2014). *Mind and its evolution: A dual coding theoretical approach*. Psychology Press.

Pearson, P.D. (1974). The effects of grammatical complexity on children's comprehension, recall, and conception of certain semantic relations. *Reading Research Quarterly*, *10*(1), 155-192.

Perfetti, C.A. (1992). The representation problem in reading acquisition. In P.B. Gough, L.C. Ehri, & R. Treiman (Eds.), *Reading acquisition* (pp. 145-174). Lawrence Erlbaum.

Perfetti, C. (2007). Reading ability: Lexical quality to comprehension. *Scientific Studies of Reading*, *11*(4), 357-383.

Perfetti, C.A., & Hart, L. (2002). The lexical quality hypothesis. In L. Verhoeven, C. Elbro, & P. Reitsma (Eds.), *Precursors of functional literacy* (pp. 189–213). John Benjamins.

Perfetti, C., McKeown, M.G., & Kucan, L. (2010). Decoding, vocabulary, and comprehension. In M.G. McKeown & L. Kucan (Eds.), *Bringing reading research to life* (pp. 291–303). Guilford Press.

Raudenbush, S.W., & Bryk, A.S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.) (Advanced Quantitative Techniques in the social sciences series, no. 1). Sage Publishers.

Raudenbush, S.W., Bryk, A.S., Cheong, Y.F., Congdon, R., & Du Toit, M. (2011). *HLM 7: Hierarchical linear and nonlinear modeling*. Scientific software International.

Ricketts, J., Nation, K., & Bishop, D.V. (2007). Vocabulary is important for some, but not all reading skills. *Scientific Studies of Reading*, *11*(3), 235-257.

Share, D.L., Jorm, A.F., Maclean, R., & Matthews, R. (1984). Sources of individual differences in reading acquisition. *Journal of Educational Psychology*, *76*(6), 1309-1324.

Stanovich, K.E. (1986). Cognitive processes and the reading problems of learning disabled children: Evaluating the assumption of specificity. *Psychological and Educational Perspectives on Learning Disabilities*, 87-131.

Treiman, R., Mullennix, J., Bijeljac-Babic, R., & Richmond-Welty, E.D. (1995). The special role of rimes in the description, use, and acquisition of English orthography. *Journal of Experimental Psychology: General*, *124*(2), 107-136.

University of Oregon Center on Teaching and Learning. (2020). DIBELS® 8th Edition Benchmark Goals. University of Oregon. https://dibels.uoregon.edu

Webber, C., Patel, H., Cunningham, A., Fox, A., Vousden, J., Castles, A., & Shapiro, L. (2024). An experimental comparison of additional training in phoneme awareness, letter-sound knowledge and decoding for struggling beginner readers. *British Journal of Educational Psychology*, *94*(1), 282-305.

Zeno, S.M., Ivens, S.H., Millard, R.T., & Duvvuri, R. (1995). *The educator's word frequency guide.* Touchstone Press.

Zevin, J.D., & Seidenberg, M.S. (2002). Age of acquisition effects in word reading and other tasks. *Journal of Memory and Language*, *47*(1), 1-29.