INTERNATIONAL LITERACY ASSOCIATION
READING RESEARCH QUARTERLY

The Word Complexity of Primary-Level Texts: Differences Between First and Third Grade in Widely Used Curricula

Devin M. Kearns
University of Connecticut, Storrs, USA

Elfrieda H. Hiebert
TextProject, Santa Cruz, California, USA

ABSTRACT

The Common Core State Standards emphasize the need for U.S. students to read complex texts. As a result, the level of word complexity for primary-level texts is important, particularly the dimensions of and changes in complexity between first grade and the important third-grade high-stakes testing year. In this study, we addressed word complexity in these grades by examining its dimensions and differences in the texts in three widely used U.S. reading programs. Fourteen measures of word complexity were computed, and exploratory factor analysis established that four dimensions—orthography, length, familiarity, and morphology—characterized word complexity. As expected, the third-grade texts have more complex words than the first-grade texts have in the four dimensions, with the greatest differences in length and familiarity. More surprisingly, the words in the first-grade texts increase in complexity over the year, but overall, the words in the third-grade texts do not. Polysyllabic words are common in texts in both grades, comprising 48% of unique words in first-grade texts and 65% in third-grade texts. Polymorphemic words comprise 13% of unique first-grade words and 19% of third-grade words (for derived words, 3% and 6%, respectively, of all words). Results show that word complexity changes markedly between grades as expected, not only in length and familiarity but also in syllabic and morphemic structure. Implications for instruction and future word complexity analyses are discussed.

Word recognition skills are not the only influence on comprehension, but they are essential to it (Hoover & Gough, 1990). As Perfetti and Hart (2002) stated, "reading is partly about words. Or, to begin the argument more forcefully, it is mainly about words" (p. 189). This assertion defines the lexical quality hypothesis, which suggests that reading comprehension is dependent on high-quality representations of the orthographic, phonological, and semantic characteristics of words (Perfetti, 2007). Simply put, long-term success in comprehension depends on strong knowledge of individual words, indicating the importance of examining word recognition demands. Features of text, such as its structure and coherence, also account for student performance as school texts increase in length and complexity (Graesser, McNamara, & Kulikowich, 2011). However, as Fitzgerald et al.'s (2015) analysis showed, even discourse measures that explained first and second graders' comprehension of text were measures of repetition of vocabulary within and across sentences.

Deliberate practice of words and word patterns has benefits, but such practice is insufficient for developing adequate comprehension skills (Cunningham & Stanovich, 1997). As students encounter words in texts, they consolidate their knowledge of letter–sound patterns and whole

INTERNATIONAL LITERACY ASSOCIATION

Reading Research Quarterly, 0(0)
pp. 1–31 | doi:10.1002/rrq.429

words (Share, 1995). Instructional texts that are dense with words for which students have insufficient decoding skills do little to support reading development (Juel & Minden-Cupp, 2000; Lesgold, Resnick, & Hammond, 1985).

In this study, we examined the complexity of word recognition proficiencies needed to read prominent instructional texts over the reading acquisition period. We also determined whether shifts in word complexity occurred within and across grades during this period. The issue of word complexity in instructional texts during the reading acquisition period is especially critical for students learning to read in English because of its deep orthography. Seymour, Aro, and Erskine (2003), in analyzing reading development in 16 European countries with 13 unique languages, showed that students finishing their first year of schooling had foundational reading proficiencies in most languages, but not in countries with instructional languages of French, Portuguese, Danish, or in particular, English. The rate of development in English was less than half the rate in shallow orthographies. Seymour et al. concluded that these effects were due to differences in syllabic complexity and orthographic depth, not age of school entry.

In a deep orthography, where orthographic components vary in size and pattern, phases would be expected in which readers are progressively introduced to increasingly more complex levels of orthography. At the foundational level, the basic elements built on letter–sound knowledge are acquired (Ehri, 2017). Increasingly complex orthographic units (e.g., rimes, syllables) and morphographic structures are progressively internalized in subsequent phrases (Seymour, 1997) until readers have an orthographic framework that represents the full complexity of the system (Plaut, McClelland, Seidenberg, & Patterson, 1996). In English-speaking countries such as the United States and the United Kingdom, reading development is a focus of the first three years of schooling. The expectation that students attain proficiency in word recognition over this period is reflected in legislation of 16 U.S. states that require retention of third graders who do not attain a specified reading level (Auck & Atchison, 2016). We examined the nature of word recognition during the primary-grade period in the present study.

We begin with a brief review of the research on how particular features of words influence reading acquisition. We then review research on the nature of the word recognition task in instructional texts across the primary-grade span. Finally, we describe the manner in which the current study extends theory and practice on reading acquisition.

Word Features That Influence Word Recognition Over the Primary-Grade Period

Our interest lies in describing the presence and progression of a set of theoretically distinct dimensions of word recognition within texts of the reading acquisition period. We selected 14 manifest variables that align with research and theory on word complexity, and used exploratory factor analysis (EFA) and confirmatory factor analysis (CFA) to determine whether a smaller subset of dimensions could characterize the variability in the manifest variables. Our goal was to advance understanding of word complexity in theory and practice.

In the empirical and theoretical literature, many variables have been identified as contributing to word complexity (e.g., Cain, Compton, & Parrila, 2017; Rueschemeyer & Gaskell, 2018; Yap & Balota, 2015). We believe that these variables are represented by three theoretically distinct constructs of words: length, familiarity, and structure. In this section, we provide a rationale for the importance of these three factors in understanding word complexity. We describe the final structure construct in depth because of its hypothesized centrality in the development of English reading fluency (Seymour et al., 2003).

Word Length

Word length is known to affect the speed and accuracy of word reading in the elementary years. This result has been found in studies that involved reading words in isolation (e.g., Gagl, Hawelka, & Wimmer, 2015; Marinus & de Jong, 2010) and words in texts (e.g., Juel & Roper/Schneider, 1985). The effect of word length is particularly critical for beginning readers, with the effect diminishing across the elementary grades (e.g., Gagl et al., 2015; Zoccolotti, De Luca, Di Filippo, Judica, & Martelli, 2009).

Word length is not merely about the number of letters; it involves the number of syllables as well. Data indicate that readers perform a visual analysis of words that involves blocking them into syllablelike units anchored by vowel letters (Chetail, Balota, Treiman, & Content, 2015). Readers might also perceive length in terms of the number of morphemes in a word, blocking the word into morphological components.

Word Familiarity

Familiarity is related to the storage of known words in the mental lexicon—words used in oral communication (Nagy & Hiebert, 2011). In reading, immediate preliminary orthographic analysis of written words in texts is followed by retrieval of information about phonological form and meaning; both types of information contribute to speed and accuracy in word reading. However, because researchers have found it difficult to estimate students' familiarity with words, researchers have turned to measures of word frequency (especially those calculated on the basis of grade-specific text corpora), the age at which words are typically acquired in the phonological lexicon (i.e., age of acquisition), or frequency of word parts (morphemes) in written texts.

Moreover, the frequency of the root words of morphologically complex words merits attention. Two aspects of frequency have been shown to influence recognition of morphologically related words: family size and frequency of family members (De Jong, Schreuder, & Baayen, 2000). *Family size* refers to the number of different words that have the same word stem (e.g., *work*, *worker*, *workable*, *working*). The frequency of all family members denotes the number of times readers can be expected to encounter a set of words with the same stem, such as those related to the word *work*.

These measures account for the likelihood that students have mental representations for bound and free morphemes that they have extracted from written words in texts. For example, Carlisle and Katz (2006) showed that measures of family size and frequency relate to the speed and accuracy of naming derived words. These metrics index the degree to which a given word may activate a reader's representations for morphologically related words. Readers are likely to read words with large morphological families more quickly and accurately than words without large families (Kearns, 2015) because of familiarity with orthographic and semantic constituents of the word from related words (Baayen, Milin, Đurđević, Hendrix, & Marelli, 2011). The frequency of family members has separate importance because elementary-age readers benefit from knowing words that have orthographic and semantic similarities, even if few other words contain the familiar pattern.

Word Structure

Considering structure to be the arrangement of and relations between parts or elements of something complex, word structure encompasses features related to the letters themselves, the relation of letters to sounds, morphological elements represented by groups of letters, and syllabic patterns. A particular focus in this study was on morphemic and syllabic patterns because their role in beginning reading is not well understood (Kearns, 2020; Roberts, Christo, & Shefelbine, 2011) and because of these variables' importance in the deep orthography of English (Seymour et al., 2003).

Letter-Related Structures

Data indicate that readers benefit when words contain letter patterns common to many other words (Andrews, 1997); that is, patterns have large orthographic neighborhoods. Readers are clearly sensitive to the co-occurrence of letters (e.g., Seidenberg, 1987), and this seems to be particularly true for polysyllabic words, where readers appear to differentiate syllabic units using consonants and vowels as boundaries (Chetail et al., 2015). Others have shown that readers' understanding of stress patterns depends on the letter patterns they perceive at the ends of words (Arciuli, Monaghan, & Ševa, 2010). Taken together, the data on letter units indicate that the size of a word's orthographic neighborhood is an important aspect of the word's structure that might differ across grades.

Letters and Sounds

The relation between letters and sounds is also part of the construct we call structure, but it reflects a different kind of complexity due to letter–sound (in)consistencies, particularly for vowels (Treiman, Mullennix, Bijeljac-Babic, & Richmond-Welty, 1995). English has many exemplary regularities (Perfetti, 2003), and numerous studies have shown the value of helping students understand its letter–sound system (e.g., National Institute of Child Health and Human Development, 2000). However, readers find it more difficult to read words with inconsistent grapheme–phoneme correspondences (GPCs) than words with consistent ones, both for monosyllabic (Glushko, 1979) and polysyllabic words (Perry, Ziegler, & Zorzi, 2010).

Despite the evident importance of the consistency of letter–sound connections, no sources provide consistency data for a large number of words (especially polysyllabic ones), so letter–sound consistency has not been examined carefully in text analyses. The lack of data on consistency is mainly associated with the many challenges associated with consistency calculations for polysyllabic words. Only one prior study has attempted to calculate consistency for polysyllabic words: the study by Yap and Balota (2009), who found relations between multiple instantiations of letter–sound consistency and polysyllabic word naming and lexical decision performance using the English Lexicon Project data.

Letters and Morphemes

Bound morphemes are, like GPCs, sublexical units but differ in that they also convey meaning. Both their orthographic and semantic characteristics can benefit readers. As orthographic units, morphemes usually contain two or more letters and often compose whole syllables, significantly reducing the number of units needed to recognize unknown words (Levesque, Kieffer, & Deacon, 2017). The semantic benefit is that morphemes link spelling directly to meaning. This may benefit readers when words' letters do not have transparent links to phonology but connect with other known concepts with similar spellings, such as *sign* for *signature* (Gonnerman, Seidenberg, & Andersen, 2007).

In addition, the number of morphemes within a word contributes to word complexity (Dawson, Rastle, & Ricketts, 2018). Compound words contain multiple free morphemes, but many words also contain derivational or inflected morphemes. Readers clearly encounter inflections and compounds in early language and reading experiences (Anglin, 1993), but the number of words encountered that contain derivational morphemes grows over time (Carlisle & Kearns, 2017). Therefore, considering word derivation is central to understanding morphology as part of a word's structure.

Syllabic Patterns

There has been limited research on the manner in which syllabic patterns and the presence of polysyllabic words in instructional texts influence word recognition for beginning readers (Roberts et al., 2011). Derivational and inflected morphemes result in multiple syllables, but there are also multisyllabic words in which the meaning of the derivation has long been lost (e.g., *surprise* from *super + prehendo*) or in which the phonological–orthographic relations are reflected in more than one syllable (e.g., *cavity*).

For students beyond the beginning reading period, polysyllabic polymorphemic words have been shown to require greater attention on the part of readers than monosyllabic monomorphemic words (Kearns, 2015). Theories of reading development generally suggest that readers only learn to attend to words at the syllabic and morphemic levels after consolidating knowledge of smaller units. For example, Seymour (1997) proposed that young students first learn the alphabetic principle, followed by an orthographic development phase involving syllables and then a phase focused on words' morphological units. Ehri (2005) characterized the end of the consolidated alphabetic phase as involving the development of sensitivity to words' syllabic and morphographic structures.

None of these models provides a grade level at which readers should build these understandings, and that is precisely why our analysis of first- and third-grade texts' word recognition demands is helpful: We can learn whether polysyllabic, polymorphemic words occur often in first- and third-grade texts, are salient only in third grade, or are trivial in both grades. Each possibility has obvious implications for the type of word recognition instruction that will facilitate reading success in current texts.

Summary of Constructs

Measures within the three constructs and from different constructs may well contribute to more than a single factor. Taking length as an example, readers probably attend to multiple aspects of this dimension at one time; the ability to read a word is not affected solely by the number of letters it has. How many syllables or morphemes a word has is important as well, and readers address these distinct aspects of the words in parallel. Harm and Seidenberg (2004) and Plaut et al. (1996) are among researchers who have shown that readers have distributed representations of words such that they do not simply map *A* in *cat* to /æ/. Rather, the reader regards *cat* in terms of *C*, *A*, *T*, *CA*, *AT*, and *CAT* and decides which sounds to produce by using all information regarding how these letters and letter groups are pronounced. Thus, defining length by the number of letters, although a common approach (Kearns, 2015), may ignore meaningful variations related to other units.

Of the three constructs, structure is likely the most complex. For example, the orthographic structure of words has relevance on its own and in letters' relations with sounds, morphology, and parts of syllables. These complexities make it difficult to propose that these diverse elements of structure will map onto a single construct. The variables chosen to represent this construct will likely exhibit relations among themselves in ways both predicted (letter-related, letter–sound-related, and letter–morpheme-related factors) and unexpected.

The Presence of Word-Level Constructs in Texts of the Reading Acquisition Period

Texts that support students in developing proficient word recognition over the reading acquisition period would be expected to show differences in the three constructs of length, familiarity, and structure. Our review of research indicated that studies of word features in reading texts have been few and have not addressed how features of length, familiarity, and structure change over the texts of the reading acquisition period. The handful of examinations of individual words in texts have typically categorized words into discrete categories, such as rare words (Hayes, Wolfer, & Wolfe, 1996), high-frequency words (Chall, 1967), or decodable complexity (Juel & Roper/Schneider, 1985). (An extended review of the existing research on word features in the texts of the reading acquisition period can be found in Appendix A.)

In that multiple properties of a word can influence its recognition, the existing studies did not describe the proficiencies required by readers to identify the words in typical texts. An analytic scheme based on single features, such as frequency, or on decodable complexity will not capture the underlying skills required to read the typical words in beginning texts. For example, none of the previous studies has described the word recognition demands that are evident in a widely used first-grade assessment where approximately 17.5% of the unique words are multisyllabic (e.g., *different*, *busy*) and 18.2% morphologically complex (e.g., *showed*, *beginning*; Hiebert, Toyama, & Irey, 2020).

Additionally, features that can influence word recognition, such as concreteness/abstractness and age of acquisition, have been restricted because of the lack of comprehensive databases. Databases on concreteness/abstractness (Brysbaert, Warriner, & Kuperman, 2014) and age of acquisition (Kuperman, Stadthagen-Gonzalez, & Brysbaert, 2012) of words are substantially larger than they were even a decade ago. Further, statistical procedures have become available over the past decade that make it possible to examine features of words in large corpora of texts with ease and depth that was previously impossible. An understanding of word proficiencies required to read instructional texts that span the reading acquisition period is essential for the design and implementation of instruction, especially for challenged readers. We designed the current study to apply analytic schemes and statistical procedures to understand the multiple features of words that are present in the texts that cover the reading acquisition period.

Research Questions

Our goal in the current project was to describe the word recognition task within and between grade levels during the reading acquisition period. This analysis of the word recognition task was not limited to the reading opportunities posed by texts in U.S. reading programs; the patterns in texts in the United Kingdom over the same period share similar characteristics. (More detail about the similarities between the words in reading acquisition programs in the United States, including those that extend beyond those in the current analysis, and those in current reading programs in the United Kingdom can be found in Appendix B.)

Our interest was in creating factors that represent word complexity dimensions from theory and extant empirical data. These dimensions can synthesize information across manifest variables in ways that better represent how readers process words. In particular, factor analysis can reduce the size of the analytical space necessary to understand complexity. Parsing the relative contributions of more than a dozen manifest variables to word complexity is challenging. A smaller set of theoretically distinct constructs can be useful for designing research and instruction on the nature of and changes in word recognition demands across the reading acquisition period.

As our first research question, we asked whether an EFA using measures related to length, familiarity, and structure constructs would produce a factor structure that supported the theoretical model and whether a subsequent CFA would indicate good fit for the identified structure. Our hypothesis was that the length and familiarity variables would map onto the expected constructs, but we did not anticipate a single-structure factor. Rather, we expected that the manifest variables used to characterize structure would produce a small set of structure-related constructs.

Our second research question addressed whether first- and third-grade texts differed in the complexity of words. Answering this question is essential to developing understanding of word recognition demands of third-grade texts and how students' first-grade reading experiences align with those demands. We expected that the first-grade texts would have lower magnitude factor scores for their words (i.e., lower complexity) than third-grade texts would overall.

Our third research question, an extension of the second, addressed whether complexity of words increases across grades within programs. We expected that the words in first-grade texts would increase in complexity from the beginning to the end of the year, given the large changes in reading development that occur over that period. More modest increases in word complexity might be expected in third grade, but the emphasis on third-grade proficiency within policies (e.g., No Child Left Behind Act, 2002) led us to expect at least a gradual increase in word complexity over this grade.

Our final research question concerned the differences in number of syllables and types of morphemes in words in first- and third-grade texts. We hypothesized that there would be more polysyllabic and polymorphemic words in third-grade texts, in both types and tokens, and that the distribution of morpheme types would change between grades. Specifically, we expected third-grade texts to have derivations, whereas polysyllabic words in first-grade texts would be primarily inflectional morphemes.

In summary, we designed this study to answer four research questions:

1. Does an EFA using multiple measures of three constructs—length, familiarity, and structure—support a hypothesized factor structure based on data regarding words in two grade levels (first and third) from three widely used reading programs, and does a CFA indicate that the identified structure will fit data outside the exploratory context?
2. Do the grade levels differ in complexity of words in texts relative to factors identified in research question 1?
3. Does the level of factors change across texts in programs and grade levels?
4. How do syllabic and morphological features of texts' words differ between first and third grade?

Method

Word Data Sources

According to research conducted for publishers (Education Market Research, 2014), 73% of elementary educators in the United States use a core reading program. The U.S. Census (Bauman & Davis, 2013) reported approximately 3.8 million students in an American first-grade cohort in the mid-2010s. Approximately 30% of the nation's students were in the states of California, Texas, and Florida; all three states have identified English language arts core reading programs that can be purchased with state funds. We consulted each state's textbook adoption lists to verify that the three reading programs analyzed in this study were included in those states' lists (California State Board of Education, 2015; Florida Department of Education, 2013; Texas Education Agency, 2015). The three programs in this study were the only ones listed on the approved lists of all three states: Houghton Mifflin Harcourt's Journeys (HMH; Baumann et al., 2014), McGraw-Hill's Wonders (MH; August et al., 2014), and Scott Foresman's Reading Street (SF; Afflerbach et al., 2013). The levels of the programs studied were first and third grades. We refer to each level of each program as a program level (e.g., HMH first grade is a program level).

Of the many components in core reading programs, we focused on the student text, the anthology, around which a week's instruction is centered. Lessons were created around a 30-week school year, with two selections per week in first grade and one selection per week in third grade. All texts were scanned and run through an optical character recognition program. Two research assistants checked text files for accuracy and applied criteria for inclusion and exclusion of characters. (More detail on the criteria for inclusion and exclusion of characters in text appears in Appendix C.) Words remaining after the vetting process formed the database. For each program, number of unique words (types) and total words (tokens) were established. The six program-level databases were merged, resulting in a final database with 8,550 types representing 129,523 tokens, as summarized in Table 1.

Word Feature Databases

The majority of word features were compiled from three large databases: CELEX (Baayen, Piepenbrock, & van Rijn, 1995), *The Educator's Word Frequency Guide* (EWFG; Zeno, Ivens, Millard, & Duvvuri, 1995), and Unisyn (Fitt, 2001). The three databases were combined into a master database, using word form as the unique identifier.

CELEX

CELEX (Baayen et al., 1995) is a database of 160,595 English words, 89,387 of which are unique. We extracted the American English spellings (CELEX provides both British and American).

EWFG

The EWFG (Zeno et al., 1995) contains 143,871 words, each with standard frequency index data. The standard frequency index is a transformation of the *U* statistic, a word's type frequency per million tokens, adjusted for the dispersion across content areas (Breland, 1996). The database contains word counts (untransformed and uncalculated per million) for each word form for texts from first grade to thirteenth, or college level.

Unisyn

Unisyn (Fitt, 2001) contains information about the spelling, pronunciation, and morphological structure of 119,356 English words. Phonology of Unisyn words was coded using the database's accent-free system (Wells, 1982) to create the unilex lexicon. The unilex database is accompanied by a Perl script to change pronunciations in the General American accent. The result is a database containing the phonology of each word written in X-SAMPA (Extended Speech Assessment Methods Phonetic Alphabet) and its spelling. Perl scripts were used to extract General American pronunciations to create the Unisyn database.

Variables for Analysis

We describe the 14 variables in three sections based on the three hypothesized constructs: length, familiarity, and structure.

Length

The four measures of length were letters, morphemes, phonemes, and syllables.

Letters

The number of letters was calculated by Stata 15 (Statacorp, 2017), using this code: generate NLet = length(word).

Morphemes

Number of morphemes was derived from CELEX and Unisyn. Morpheme counts occasionally differed between these data sources, and a linguist with expertise in morphology resolved the discrepancies for these cases. Morphological counts included free morphemes or stems, affixes, and inflections. For instance, *hunters* consists of the

TABLE 1
Words and Texts Within Each Program Analyzed

	First grade			Third grade		
Program	Unique words (types)	Total words (tokens)	Texts	Unique words (types)	Total words (tokens)	Texts
Houghton-Mifflin Harcourt's Journeys (Baumann et al., 2014)	1,738	11,597	60	4,067	30,823	25[a]
McGraw-Hill's Wonders (August et al., 2014)	1,861	12,274	60	4,135	30,148	30
Scott Foresman's Reading Street (Afflerbach et al., 2013)	1,818	10,961	60	4,534	33,720	30
All	3,293	34,832	180	7,779	94,691	86

[a]The last unit of five weeks in the program is devoted to three trade books that are not part of the student anthology; these texts were not included in the present analysis.

stem *hunt* plus the suffix *-er* and the inflection *-s* to total three morphemes. Bound or unproductive morphemes were not counted separately; thus, *aquatic* (*aqua* + *tic*) and *celebrate* (*celebr* + *ate*) were each considered one morpheme. Some words had no morphological data, including words that are colloquial, proper, foreign, onomatopoeic, or abbreviated. Morphological data for other words not included in CELEX or Unisyn were analyzed and coded individually by a linguist. The number of morphemes variable was cross-loaded on the orthographic structure factor because morphological units have a unique orthographic structure; they are sublexical units with phonological and/or orthographic consistencies that are uncharacteristic of syllables (syllabic structures that are not morphemes have no consistent orthographic structure) and therefore had the potential to have cross-loadings.

Phonemes

Both Unisyn and CELEX provided information for calculating the number of phonemes. These sources were cross-referenced, and all discrepancies were resolved by the first author and a linguist, selecting General American pronunciation when discrepancies occurred.

Syllables

Number of syllables was the number of phonological syllables calculated from the General American Unisyn pronunciations. Unisyn contains codes for all vowel pronunciations, including diphthongs, reduced vowels, and syllabified consonants (e.g., OI for /ɔɪ/, @ for /ə/, = n for /n̩/). The number of syllables was the sum of all vowels within a word, based on vowel pronunciations given by Unisyn in X-SAMPA.

Familiarity

The familiarity measures pertain to age of acquisition, word frequency overall, grade-level frequency, and frequency of root words.

Age of Acquisition

Age-of-acquisition information came from Kuperman et al.'s (2012) database, which consists of ratings for 30,121 words. Kuperman et al. extrapolated ratings to words with the same lemma but different forms (e.g., *demand*, *demanded*, *demanding*).

EWFG Log Frequency

Grade-specific EWFG frequency totals were used to create a variable based on the likelihood of a word occurring in texts that would be read by elementary-age students (grades 1–5). The sum of grade-specific frequencies was used to obtain frequency for the entire period. The variable was log-transformed to normalize the distribution.

Rank Frequency

Rank frequency was based on frequencies of words in three databases: the EWFG's standard frequency index, which includes words not in the grade-specific data set; the Hyperspace Analogue to Language from the English Lexicon Project; and the Google Ngram database. A word that occurred in at least two of the three databases was included. The rank for each word came from a given database. For example, the word *the* had the first rank in all three databases. Rank frequency was the mean of the standardized rank frequencies for all three databases, similar to that in Yap and Balota's (2009) study.

Words With First Root

To calculate total frequency of a root family, all words in a word family in the EWFG were counted and computed with data from Unisyn. For example, the root family size for the root *bird* was seven, reflecting the presence of these related words: *birdie*, *birds*, *blackbird*, *blackbirds*, *hummingbird*, and *hummingbirds*.

Structure

Structure measures addressed both orthographic and phonological features of words.

Bigram Frequency

A bigram consists of any consecutive pair of letters: two consonants, two vowels, or a consonant and a vowel. A program was constructed to calculate the bigram frequency for words, using the summed frequency of words in the EWFG for grades 1–5. For each bigram in a word, the program counted the number of times it occurred in the EWFG database for first through fifth grades. For example, the word *final* has the bigrams *_F*, *IN*, *NA*, *AL*, and *L_* (an underscore indicates that the letter begins or ends a word; initial and final letters are thus bigrams). The program counted the number of times these bigrams occurred in the EWFG for grades 1–5 on a type basis. The mean bigram frequency based on word types (i.e., mean number of times a given bigram occurs in the corpus) was used.

Orthographic Levenshtein Distance Frequency

This measure (Yarkoni, Balota, & Yap, 2008) is based on the edit distance, the number of letters that need to be changed to make a given word into its nearest neighbors (Levenshtein, 1966), based on the English Lexicon Project (Balota et al., 2007). Changes can be insertions, substitutions, or deletions. To get the OLD20 frequency for a word, the edit distance is calculated to identify the 20 nearest neighbors.

Orthographic *N*

Orthographic *N* (Coltheart, Davelaar, Jonasson, & Besner, 1977) is calculated as the number of words that can be created with a change of one letter within a given word. These data were obtained from the English Lexicon Project database (Balota et al., 2007).

Orthographic *N* for These Texts

Orthographic *N* data were calculated for the words in the texts in this analysis, again using Coltheart et al.'s (1977) procedure.

Orthographic Transparency

In a deep orthography, orthography represents a compromise between the links between spoken and written forms. Sometimes the orthographic forms change while phonological forms do not (e.g., *hurry* to *hurried*). In other words, phonological forms change (e.g., *sign* to *signature*). Orthographically transparent words are similar to *signature*. As with its phonological counterpart, orthographic transparency has been shown to correlate with students' ability to read polymorphemic words (Kearns, 2015). Words were coded dichotomously as orthographically transparent (1) or not (0).

Phonological Transparency

Words were coded dichotomously as phonologically transparent (1) or not (0). Some words are neither phonologically nor orthographically transparent (e.g., *wisdom*), so these words were coded as 0 for both variables.

Data Analytic Procedure

Research Question 1: Fit of the Hypothesized Model

We used EFA and CFA with the lavaan (Rosseel, 2012) and semTools (Jorgensen et al., 2018) packages in R (R Core Team, 2019). First, the data were prepared as either first-grade or third-grade words. The first-grade word category included all words in both first-grade and third-grade texts and all words in only first-grade texts. The third-grade word category consisted of only words in third-grade texts.

Next, we separated the words into two subsamples including only unique words so the fit of the EFA models was evaluated with an orthogonal subset of the words. The EFA models were tested with 25% of the data, and the CFA models were tested with the remaining 75%. Data were stratified by grade so the proportions of words were similar for first-grade and third-grade words.

We conducted the EFA to determine whether the structure of the data aligned with the theorized structure. EFA was appropriate because no similar studies exist in the extant literature to guide any CFAs at this stage. We used the empirical Kaiser criterion (Braeken & van Assen, 2017) in the initial analysis to establish the number of candidate models to test. The fit of each candidate model was evaluated with the comparative fit index (CFI) and Tucker–Lewis index (TLI), in which good fit values are greater than 0.95 (Kline, 2005). In the root mean square error of approximation (RMSEA), excellent fit values are less than 0.1, good fit values are less than 0.5, and mediocre fit values are less than 0.08 (MacCallum, Browne, & Sugawara, 1996); in the standardized root mean square residual (SRMR), good fit values are less than 0.8 (Hu & Bentler, 1999). We used these indices to identify the most parsimonious model with the best fit to data. Then, the CFA determined whether the factor structure identified in the EFA fit the data adequately when the model included only paths between manifest variables and factors with high loadings in the EFA. We used the same fit statistics to determine model fit.

For both EFA and CFA, an oblique geomin rotation was used to obtain factor solutions, allowing for correlated factors. We expected loadings for more than one factor, and prior data suggest that this rotation works well for complex data. Browne (2001) showed this using Thurstone's (1947) simulated data in which an orthogonal solution will fail to represent the underlying structure. In addition, use of casewise deletion from lavaan maximized the amount of data used in the analysis.

Research Question 2: Intergrade Differences in Levels of Factors

We aimed to determine whether words in first-grade and third-grade texts differed in the magnitude of each factor. This question was based on the assumption that first- and third-grade words had the same factor structure. To check this, we tested increasingly constrained models for evaluation. First, we tested configural invariance to determine whether the correlations among factor loadings were the same across grades. Then, we tested metric invariance to determine whether loadings of each manifest variable on factors were the same for each grade. The tests of metric variance were based on the comparison between fit indices for the model for each grade. The difference should be very small: $CLI \leq -0.005$, $RMSEA > 0.010$, and $SRMR > 0.025$ (Chen, 2007). Thus, the answer to research question 2 was either a comparison of the magnitudes of each factor between grades (there is invariance) or a comparison of the factor structures between grades (there is not invariance).

Research Question 3: Factor-Level Trajectories Within Programs

We extracted factor scores for each word from the CFA using all data. Then, we created aggregate factor scores for each text using the mean of each set of factor scores for words in the given text. Finally, we used a series of regression analyses for each program, with each factor as a dependent variable and order of the text in each program (i.e., whether it was taught earlier or later) as the independent variable. The magnitudes of these factor score lesson order slopes established whether programs included systematic change in word complexity for a given construct.

Research Question 4: Frequency of Polysyllabic and Polymorphemic Words

Addressing the syllabic and morphemic features of words, we explored patterns in manifest variables describing these features by grade. We examined the number of syllables and morphemes in words in each grade's texts. Additionally, we examined morpheme type (inflections, derivations, and compounds) and the words' orthographical and phonological transparency. To determine whether meaningful differences lie in distributions of number of syllables or morphological type, we used chi-square goodness-of-fit tests. Effect size is given by Cramér's V, where a small effect is considered to be approximately 0.05, a moderate effect 0.10, and a large effect 0.15. We calculated four effects: (1) number of syllables for two groups: monosyllabic and polysyllabic; (2) number of syllables using all syllable lengths as factors; (3) number of morphemes for two groups: monomorphemic or polymorphemic; and (4) morpheme types using eight categories: monomorphemic, compound, inflected, derived, compound-inflected, compound-derived, inflected-derived, and compound-inflected-derived.

Results

Research Question 1: Factor Structure

EFA

Table 2 presents descriptive statistics for the 14 variables used in the factor analyses, and Table 3 provides bivariate correlations among variables. For the EFA, the empirical Kaiser criterion suggested four factors, and we tested candidate models with two, three, and four factors. The four-factor model had the best fit (CLI = 0.99; TLI = 0.99; SRMR = 0.05; RMSEA = 0.04). Loadings of the four-factor model showed a similar pattern to what was anticipated.

Using loadings above 0.40 as a heuristic for evaluating model fit, the first factor mapped onto an orthographic structure dimension with loadings for orthographic N and for orthographic N for these texts of 0.97 and 0.89, respectively. The orthographic Levenshtein distance frequency loading was lower but showed some relation at 0.35. The second factor mapped onto the length dimension with the expected variables of letters, morphemes, phonemes, and syllables mapping at 0.89, 0.87, 0.83, and 0.54, respectively. In addition, orthographic transparency, phonological transparency, and orthographic Levenshtein distance frequency loaded at 0.64, 0.59, and −0.45, respectively. The bigram frequency loading was somewhat lower but still related at 0.36. The third factor mapped onto the familiarity dimension with strong loadings for EWFG frequency, rank frequency, and age of acquisition, at 0.87, 0.59, and −0.59, respectively. The fourth factor that mapped onto a second structure dimension was most strongly related to number of morphemes and number of words with the same first root (0.53 and 0.48, respectively). This was termed a morphological structure dimension. Table 4 displays the set of factor loadings.

CFA

The CFA was tested with 75% of the data not used for the EFA. The number of morphemes variable was specified as cross-loading on the length and morphological structure dimensions. We tested the factor structure of the EFA for all data. The fit using diagonally weighted least squares was adequate (CFI = 0.96; TLI = 0.95; SRMR = 0.09; RMSEA = 0.08, 95% confidence interval [0.081, 0.085]) and suggested that the four-factor model adequately represented the data. The factor pattern was the same as for the CFA.

Research Question 2: Intergrade Differences in Levels of Factors

The configural invariance test indicated that correlations among the factors were similar across grades (words in first- and third-grade texts versus words in third-grade texts alone). Figure 1 shows the distributions of factor scores for the first- and third-grade words and reveals the similarities in the factor structure at the configural level across the grades. However, the test of metric invariance showed that first-grade and third-grade words had different loadings on factors with differences for CFI of −0.008, TLI of −0.00553, RMSEA of 0.00420, and SRMR of 0.00503. None of these differences met Chen's (2007) criteria, so the means could not be compared between the first-grade and third-grade words on this basis. Table 5 displays descriptive statistics for the factors and the correlations between them for each grade. The table indicates that the largest difference concerns the number of words with the same first root as a given word, which does not load at all onto a morphology factor for first-grade words but loads onto this factor for third-grade words.

Research Question 3: Factor-Level Trajectories Within Programs

Regression analyses for programs showed a mixed pattern of effects, as displayed by factor covariances in Table 6. Among first-grade programs, HMH showed decreasing familiarity across the year ($\beta = 0.293$, $R^2 = .08$, $p < .05$) and increasing morphological complexity across the year ($\beta = -0.319$, $R^2 = .10$, $p < .05$). MH showed only increasing morphological complexity ($\beta = -0.395$, $R^2 = .15$, $p < .01$), and SF showed decreasing familiarity ($\beta = 0.265$, $R^2 = .07$, $p < .05$) and morphological structure ($\beta = -0.418$, $R^2 = .17$, $p < .01$). Among third-grade programs, no statistically significant changes were seen in levels of any of the third-grade word factor scores. (The figures in Appendix D depict changes in the factor scores across texts in each program in each grade.)

TABLE 2
Descriptive Statistics for Variables Used in Factor Analyses

	All words (*N* = 8,749)					First-grade words (*n* = 4,915)					Third-grade-only words (*n* = 3,831)				
Variable	*N*	*M*	*SD*	*Min*	*Max*	*N*	*M*	*SD*	*Min*	*Max*	*N*	*M*	*SD*	*Min*	*Max*
Letters	8,746	6.42	2.07	1	16	4,915	6.07	2.02	1	16	3,831	6.87	2.05	1	14
Morphemes[a]	8,280	1.68	0.65	1	5	4,722	1.60	0.63	1	5	3,558	1.77	0.65	1	4
Syllables	8,310	1.90	0.84	1	6	4,735	1.76	0.81	1	6	3,575	2.07	0.86	1	6
Phonemes	8,310	5.28	1.81	1	15	4,735	4.98	1.76	1	15	3,575	5.69	1.81	1	14
Age of acquisition[b]	7,288	6.78	2.18	2	18	4,229	6.33	2.11	2	15	3,059	7.40	2.13	2	18
EWFG log frequency[c]	8,405	3.42	2.17	0	13	4,769	4.03	2.28	0	13	3,636	2.62	1.73	0	7
Rank frequency[d]	8,445	1.46	0.30	−0.42	1.74	4,784	1.50	0.28	−0.42	1.74	3,661	1.41	0.31	−0.35	1.73
Words with first root	8,281	3.18	2.66	1	22	4,723	3.35	2.69	1	22	3,558	2.96	2.60	1	22
Mean bigram frequency[e]	8,724	321.54	133.46	1	1,199	4,899	316.81	137.55	1	1,199	3,825	327.59	127.79	2	862
Frequency of words with OLD within 20[f]	7,030	7.23	0.87	4	12	4,110	7.39	0.89	5	12	2,920	7.01	0.79	4	10
Orthographic *N* for these texts[g]	8,746	1.61	2.80	0	20	4,915	2.02	3.11	0	20	3,831	1.08	2.24	0	18
Orthographic *N*[g]	7,030	4.46	5.57	0	34	4,110	5.34	6.06	0	34	2,920	3.22	4.51	0	29
Orthographic transparency	8,263	.93	.26	0	1	4,713	.94	.24	0	1	3,550	.91	.28	0	1
Phonological transparency	8,263	.97	.17	0	1	4,713	.97	.16	0	1	3,550	.97	.18	0	1

Note. EWFG = *Educator's Word Frequency Guide* (Zeno et al., 1995); OLD = orthographic Levenshtein distance.
[a]Number of morphemes is derived from analysis of CELEX and Unisyn. [b]Age of acquisition is based on an expansion of the Kuperman age-of-acquisition norms using CELEX lemma data. [c]Log frequency is the total of EWFG words in the grades 1–5 corpus. [d]Rank frequency is based on Google Ngram, Hyperspace Analogue to Language, and EWFG standard frequency index. [e]Mean bigram frequency is derived from these data. [f]Hyperspace Analogue to Language frequency for the word's 20 nearest orthographic neighbors using the Levenshtein algorithm. [g]The first orthographic *N* is derived from these data, and the second is from the English Lexicon Project (Balota et al., 2007); both use the Coltheart formula.

TABLE 3
Bivariate Correlations for Variables Used in Factor Analyses

Variable	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1. Letters	—													
2. Morphemes[a]	.57	—												
3. Syllables	.76	.36	—											
4. Phonemes	.89	.50	.80	—										
5. Age of acquisition[b]	.32	.09	.34	.34	—									
6. EWFG log frequency[c]	−.36	−.31	−.32	−.39	−.57	—								
7. Rank frequency[d]	−.23	−.29	−.15	−.24	−.34	.69	—							
8. Words with first root	.09	.31	−.05	.03	−.25	.13	.03	—						
9. Mean bigram frequency[e]	.24	.20	.14	.18	.01	−.02	.03	.07	—					
10. Frequency of words with OLD within 20[f]	−.78	−.53	−.57	−.67	−.27	.45	.38	−.06	−.19	—				
11. Orthographic *N* for these texts[g]	−.59	−.30	−.50	−.54	−.27	.31	.18	.03	−.07	.63	—			
12. Orthographic *N*[g]	−.65	−.29	−.56	−.59	−.31	.34	.20	.06	−.01	.60	.89	—		
13. Orthographic transparency	−.20	−.26	−.26	−.19	−.04	.04	.05	−.06	−.10	.14	.10	.10	—	
14. Phonological transparency	−.22	−.18	−.23	−.20	−.08	.10	−.03	−.01	−.02	.10	.08	.11	.30	—

Note. EWFG = *Educator's Word Frequency Guide* (Zeno et al., 1995); OLD = orthographic Levenshtein distance.
[a]Number of morphemes is derived from analysis of CELEX and Unisyn. [b]Age of acquisition is based on an expansion of the Kuperman age-of-acquisition norms using CELEX lemma data. [c]Log frequency is the total of EWFG words in the grades 1–5 corpus. [d]Rank frequency is based on Google Ngram, Hyperspace Analogue to Language, and EWFG standard frequency index. [e]Mean bigram frequency is derived from these data. [f]Hyperspace Analogue to Language frequency for the word's 20 nearest orthographic neighbors using the Levenshtein algorithm. [g]The first orthographic *N* is derived from these data, and the second is from the English Lexicon Project (Balota et al., 2007); both use the Coltheart formula.

Research Question 4: Frequency of Polysyllabic and Polymorphemic Words

Syllable Differences Between Grades

Distributions of frequency of polysyllabic words between first-grade and third-grade texts are presented in Table 7. To determine whether a significantly greater proportion of polysyllabic words was in third-grade texts than in first-grade texts, we performed type and token analyses. On a type basis, the proportion of polysyllabic words was greater, $\chi^2(1) = 54.64$, $p < .001$, Cramér's $V = .11$, a moderate effect indicating that third-grade texts have more polysyllabic words. On a token basis, the direction of the effect was the same but larger, $\chi^2(1) = 302.88$, $p < .001$, Cramér's $V = .16$.

Concerning the distribution of syllable lengths (comparison of exact number of syllables), the type analysis showed a statistically significant difference in the distribution of syllable lengths, $\chi^2(5) = 71.46$, $p < .001$, Cramér's $V = .13$. On a token basis, the difference was larger, $\chi^2(5) = 320.45$, $p < .001$, Cramér's $V = .16$. Both analyses indicated that words in third-grade texts skewed longer; the difference in magnitude compared with the bivariate analysis indicated that words in third-grade texts were not only more likely to be polysyllabic but also more likely to be more than two syllables.

Morpheme Differences Between Grades

Frequencies of the categories of polymorphemic words in first-grade and third-grade texts are provided in Table 8. To determine whether a significantly greater proportion of polymorphemic words was in third-grade than first-grade texts, the first comparison concerned the number of polymorphemic words of any type and monomorphemic words. On a type basis, third-grade words were more likely

TABLE 4
Exploratory Factor Analysis Factor Loadings

Variable	Factor 1	Factor 2	Factor 3	Factor 4
Length				
Letters	−0.120*	0.886*	0.017	−0.019
Morphemes	0.000	0.540*	−0.068	0.530*
Syllables	−0.017	0.827	−0.007	−0.307
Phonemes	−0.032*	0.872*	−0.063*	−0.157*
Familiarity				
Age of acquisition	0.031	−0.041	−0.594*	−0.302*
EWFG log frequency	0.022	−0.069	0.865*	0.010
Rank frequency	0.027	0.075	0.590*	0.183*
Words with first root	0.024	0.146	0.236*	0.482*
Orthographic structure				
Bigram frequency	0.177*	0.364*	0.072	0.052
Orthographic Levenshtein distance frequency	0.345*	−0.454*	0.155*	−0.151*
Orthographic *N*	0.969*	−0.019	−0.008	−0.052
Orthographic *N* for these texts	0.894*	−0.024	−0.027	−0.024
Orthographic transparency	0.167	0.643*	0.071	0.092
Phonological transparency	−0.044	0.590*	0.052	0.029

Note. EWFG = *The Educator's Word Frequency Guide* (Zeno et al., 1995). The variables are organized according to hypothesized construct.
*$p < .05$.

to be polymorphemic, $\chi^2(1) = 35.77$, $p < .001$, Cramér's $V = .09$, indicating a moderate effect. On a token basis, the effect was large, $\chi^2(1) = 277.00$, $p < .001$, Cramér's $V = .14$.

The second comparison concerned the distribution of morpheme types, specifically whether words fit into different morphological categories by grade among polymorphemic words. For the type analysis, the difference was not statistically significant, $\chi^2(6) = 11.86$, $p = .07$, Cramér's $V = .07$. For the token analysis, the difference was moderate in size and statistically significant, $\chi^2(6) = 60.66$, $p < .001$, Cramér's $V = .10$.

For the statistically significant token effect, we conducted a post hoc analysis to identify whether the difference was related to the distribution of words by morphological category (i.e., compound, inflection, derivation). The analysis indicated that compound words were responsible for this difference, $\chi^2(1) = 47.72$, $p < .001$, Cramér's $V = .08$, for the comparison between compound words and all other categories combined. Compound words accounted for 17% of the morphologically complex words in first-grade texts but only 9% in third-grade texts. The compounds in first-grade texts had relatively simple morphological structures: A part-of-speech analysis showed that 90% were nouns, and concreteness analysis using Brysbaert et al.'s (2014) data showed a mean concreteness rating of 4.2 on a 5-point scale for the compounds.

Discussion

In this study, we considered how word features that have been shown to influence word recognition are evident in texts used for instruction over the reading acquisition period. In particular, we were interested in whether there were distinct factors that accounted for the words in texts at different points during this period. Our goal was to understand the demands of third-grade texts and their relation to first-grade texts. To this end, we examined the features of all words in three different reading programs in first- and third-grade texts.

Value of Examining Word Complexity Factors

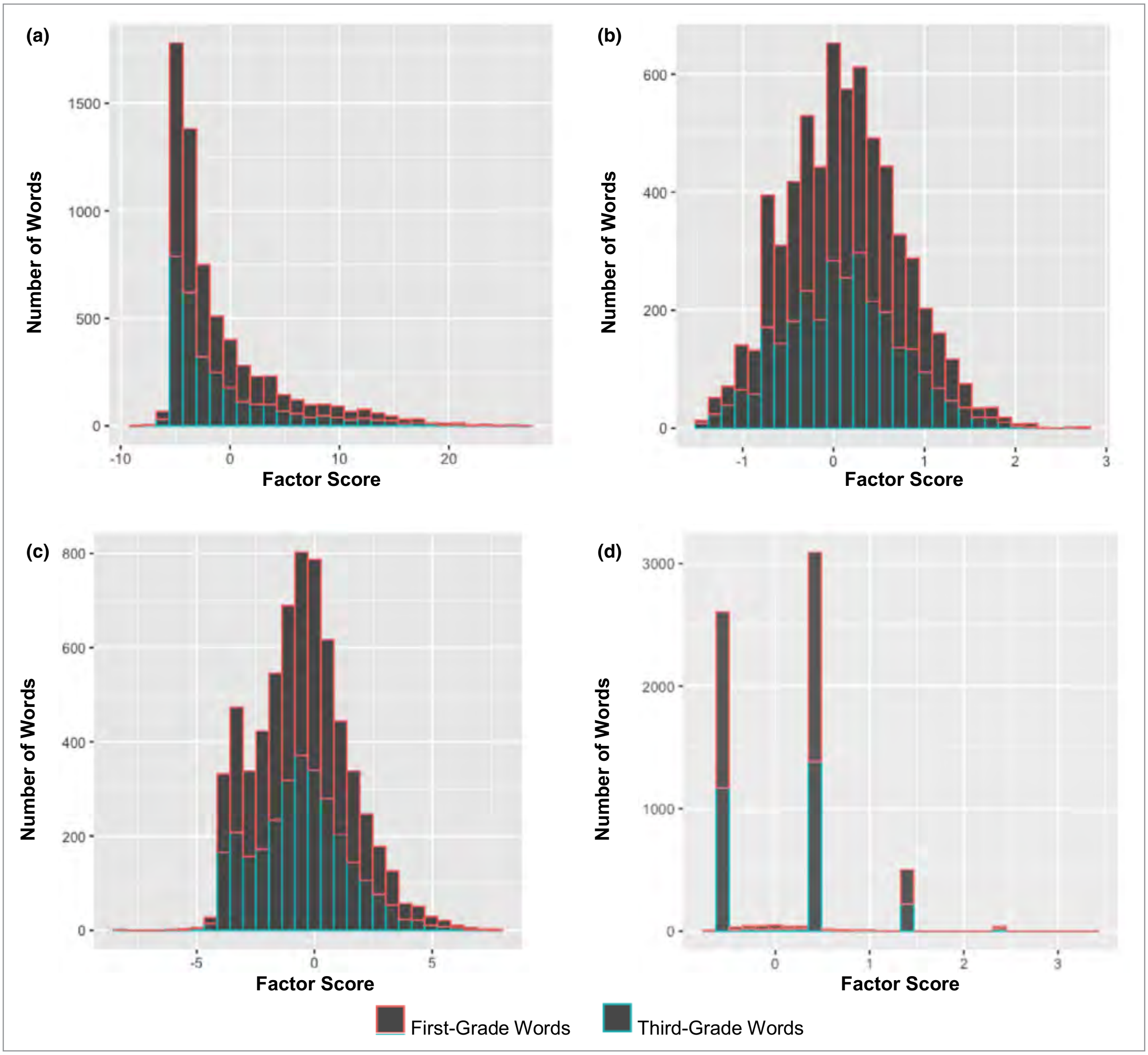
For our first research question, we considered whether a set of theoretically motivated factors could be extracted from a set of manifest variables. The data indicated that 14 manifest variables of word complexity could be characterized as a set of four latent factors. We hypothesized three prominent factors: length, familiarity, and structure; these factors were represented in the final set of four but with several distinctions.

The first of the final factor set consisted of orthographic patterns. More specifically, the variables in the first factor, such as orthographic *N* in a large database and for texts in the given sample, and orthographic Levenshtein distance frequency have to do with the frequency with which orthographic patterns appear across words. The presence of shared orthographic elements across words may be a critical feature of texts when, as is the case in the current instructional texts (Hiebert, 2005), many unique content words appear only once or twice in a grade level's texts.

The next distinguishing factor included measures that fell into the hypothesized length factor: letters, syllables, and phonemes. Three variables related to structure—bigram frequency, orthographic transparency, and phonological transparency—also loaded on this construct. Further, one variable that had been hypothesized as part of the length construct, morphemes, did not load on this factor but formed its own factor.

As we predicted, variables related to familiarity formed a distinctive factor. Within this factor, frequency of appearance in written language dominated. Age of acquisition also

FIGURE 1
Distributions of Factor Scores for First-Grade and Third-Grade Words: (a) Orthographic Structure, (b) Length, (c) Familiarity, and (d) Morphological Structure



Note. The color figure can be viewed in the online version of this article at http://ila.onlinelibrary.wiley.com.

contributed to the variable, although not as substantially as word frequency. The fourth factor confirmed the critical role of morphemes in understanding the word recognition task for primary-level students. At both grade levels, the number of morphemes was a distinguishing feature of words in instructional texts. The number of words that share the first root of a word contributed to this factor in third grade.

Thus, this EFA indicated that 14 manifest variables can be reduced to a set of four constructs that represent relatively distinctive aspects of words: orthography, length and transparency, frequency, and morphology. This finding has one simple implication: Future analyses of word complexity may not need to rely on a set of variables that serve as proxies for broader concepts. Rather, a smaller set of variables representing relevant constructs can be extracted from a set of manifest variables, and researchers can conduct analyses of word complexity using these constructs. This finding has three advantages.

Accounting for the Multidimensional Nature of Word Complexity Constructs

As we described at the outset, multidimensional constructs likely better reflect students' perceptions of word characteristics than specific variables do. The length finding is consistent with

TABLE 5
Factor Loadings From the Confirmatory Factor Analysis for First-Grade and Third-Grade Words Separately

Variable	First-grade words	Third-grade words
Factor 1		
Orthographic *N*	0.979	0.982
Orthographic *N* for these texts	0.906	0.886
Orthographic Levenshtein distance frequency	−0.870	−0.778
Factor 2		
Letters	0.942	0.939
Syllables	0.784	0.767
Phonemes	0.884	0.862
Bigram frequency	0.210	0.159
Orthographic transparency	0.495	0.436
Phonological transparency	0.497	0.561
Factor 3		
Age of acquisition	−0.686	−0.440
EWFG log frequency	1.004	0.894
Rank frequency	0.526	0.605
Factor 4		
Morphemes	0.988	0.989
Words with first toot	0.012	0.319

Note. EWFG = *The Educator's Word Frequency Guide* (Zeno et al., 1995).

that idea; the construct was linked to the number of letters, syllables, and phonemes; bigram frequency; and phonological and orthographic transparency. This aligns with a parallel distributed processing account of word recognition in which readers simultaneously perceive multiple levels of units within words and read using them (e.g., Perry et al., 2010).

Making Analysis of Word Complexity Manageable

Analyses of data from texts are notoriously complex, and researchers have typically addressed this problem by focusing on a small set of variables, such as word length or frequency. Such an approach sacrifices researchers' ability to examine other potentially important characteristics of words, especially structure. We believe that the analytic strategy employed in this analysis allows a consideration of differences in word complexity relative to several multidimensional constructs, but without examining every manifest variable individually.

Providing Data to Support the Theoretical Model

One remarkable outcome of this factor analytic model is that we could test and extend our theoretical model without specifying the factor structure a priori. Moreover, the model created factors specifically linked to morphemes, indicating that these variables had variance specific to those constructs without our building that into the model; this lent credence to our manifest variable analysis for research question 4 because these are salient factors worth further consideration.

Differences Between Grades

For our second question, we examined differences between grades and expected that third-grade words would be more complex than first-grade words along all dimensions of word complexity. We confirmed this prediction, which

TABLE 6
Factor Covariances

	Factor 1		Factor 2		Factor 3		Factor 4	
Factor	Estimate	*SE*	Estimate	*SE*	Estimate	*SE*	Estimate	*SE*
1. Orthographic structure	—		−1.923 −0.664	0.057	1.567 0.231	0.153	−0.692 −0.235	0.079
2. Length	−2.693 −0.714	0.007	—		−0.330 −0.326	0.024	0.225 0.511	0.010
3. Familiarity	5.044 0.376	0.257	−0.683 0.477	0.028	—		−0.260 −0.253	0.019
4. Morphological structure	−1.195 −0.311	0.085	0.247 0.604	0.008	−0.468 −0.321	0.022	—	

Note. *SE* = standard error. All covariances are statistically significant at $p < .001$. Correlations below the diagonal are for first-grade words and those above the diagonal for third-grade words. Top estimates are unstandardized, and bottom estimates are standardized.

TABLE 7
Frequencies of Polysyllabic Words by Grade

Number of syllables	First-grade texts only				Third-grade texts only			
	Types		Tokens		Types		Tokens	
	N	%	*N*	%	*N*	%	*N*	%
1	289	38.9	1,005	49.2	913	25.5	3,005	29.4
2	340	45.8	776	38.0	1,747	48.9	4,925	48.2
3	97	13.1	226	11.1	696	19.5	1,856	18.2
4	16	2.2	34	1.7	188	5.3	384	3.8
5	1	0.1	2	0.1	29	0.8	46	0.5
6	0	0.0	0	0.0	2	0.1	3	0.0
Any	743		2,043		3,575		10,219	

Note. Types = number of unique words; Tokens = total number of uses of words. Percentages do not total 100 because of rounding.

TABLE 8
Frequencies of Types of Polymorphic Words by Grade

Variable	First-grade texts only				Third-grade texts only			
	Types		Tokens		Types		Tokens	
	N	%	*N*	%	*N*	%	*N*	%
Compound	39	4.8	75	3.2	149	3.9	351	3.1
Inflected	216	26.5	383	16.4	1,340	35.0	3,393	30.4
Derived	70	8.6	126	5.4	452	11.8	1,075	9.6
Compound and inflected	19	2.3	28	1.2	78	2.0	116	1.0
Compound and derived	3	0.4	12	0.5	6	0.2	16	0.1
Compound, inflected, and derived	1	0.1	2	0.1	11	0.3	17	0.2
Inflected and derived	27	3.3	39	1.7	166	4.3	320	2.9
Any polymorphemic structure	375	46.0	665	28.5	2,202	57.5	5,288	47.4

is noteworthy because the effect was not limited to length and familiarity. Structure, defined primarily as the number of letters, syllables, and phonemes in words, also shows a difference between grades. Older readers encounter words with more letters, syllables, and phonemes. Further, orthographic units are not as common across words, meaning older readers encounter more complex graphemes. The syllabic effect indicates that a connection exists between the orthographic complexity of words and number of syllables.

Morphological units appear to be relevant word characteristics that should result in a sensitivity to morphology from an early age, as Carlisle and Kearns (2017) suggested. This leaves open the possibility that novice readers may benefit from instruction on examining words' morphology in the primary grades. Certainly, they should receive this instruction in the middle grades (e.g., Goodwin, 2016; Goodwin & Ahn, 2013).

Differences Within Grades

For our third research question, we considered whether texts become more complex within a grade. The words increased in complexity from the beginning to the end of first-grade texts. Such a shift acknowledges the rapid pace of growth that occurs over the first-grade year and the nature of the English lexicon. Hiebert, Goodwin, and Cervetti (2018) reported that approximately 1,307 morphological families accounted for 97% of total words and 96% of unique words in first-grade texts identified as exemplars within the Common Core State Standards (National

Governors Association Center for Best Practices & Council of Chief State School Officers, 2010). The number of word families that account for the majority of words in texts increases beyond first-grade text, but simultaneously, these word families account for a somewhat smaller percentage of unique words. In the third-grade analysis of Common Core exemplar texts, 1,787 word families accounted for 92% of the total words and 89% of unique words (Hiebert et al., 2018). That is, the number of new words in third-grade texts is high, as evident in Table 2, and many of these words are polysyllabic (see Table 7).

Even with many new words that are polysyllabic, third graders are confronted with the same level of word complexity across third-grade texts. This pattern merits attention in light of policies related to third-grade reading achievement. At least 16 states and the District of Columbia have policies mandating students attain proficient levels on summative assessments at the end of third grade or risk consequences such as retention or mandatory summer school attendance (Auck & Atchison, 2016). An underlying assumption of these initiatives is that the third-grade curriculum provides students with experiences that will be unavailable in fourth grade and beyond. Yet, the current analyses suggest that third graders need the same level of word recognition proficiency at the beginning of the year as they do at year's end. For students repeating third grade because of a lack of reading proficiency, repetition of texts with no developmental trajectory and with which they have already been unsuccessful can hardly be expected to serve as a mechanism for bringing them to proficiency.

Syllabic and Morphological Factors

The intersection of syllables and morphemes, the focus of our final question, is important. Nagy and Anderson (1984) identified morphological complexity as becoming increasingly critical as readers move through grades. What has been unclear before this study, however, is the nature of the morphological task posed by first-grade and third-grade texts. First graders may see fewer polysyllabic words than third graders, but because first-grade texts have fewer total words than third grade-texts have, first graders encounter a sizable number of polysyllabic words. Every fifth or sixth word that first graders encounter is polysyllabic (17.6% of total words). The majority of these words will have inflections or are compounds, but about one in the five or six words will be a derived word. Research that has considered how quickly beginning readers acquire inflected patterns and compound words is limited in scope, particularly with English learners, for whom inflected endings may vary from their native languages.

Morphologically complex words are even more frequent in third-grade texts, where every fourth word is morphologically complex, than in first-grade texts. Moreover, the distribution of the morphological categories changes, with fewer compound words in third-grade texts. This indicates a shift toward units that require greater morphological sophistication to pronounce and comprehend. This underscores the importance of additional research on the acquisition of morphological knowledge in beginning readers and English learners.

The level of morphological complexity confronting students at both first grade and third grade must be considered in light of the observation that much of the research on word recognition and most word recognition instruction have addressed monosyllabic words (Roberts et al., 2011). It is therefore remarkable, and perhaps concerning, that few studies have explored word recognition for polysyllabic words.

Value of the Factor Model in Explaining Performance

One criticism of our model might be that we extracted a set of factors and compared texts from two grade levels but do not have any data related to performance. That is, the factors capture meaningful data about the words (and thus texts), but this does not mean these factors have greater relevance for performance than another set of variables. Data on students' performance are not readily available, but the English Lexicon Project (Balota et al., 2007) provides accuracy and response times for adults' pronunciations of words. As a post hoc analysis, we regressed adult performance for each word present in the current data set on a group of variables representing the lexical decision response used by Yap and Balota (2009), our manifest variables, and our factors. We found that for naming accuracy, the Yap and Balota variables predicted 3% of the variance, our manifest variables 6%, and our factors 8%. For naming response times, the percentages were 40%, 42%, and 44%, respectively. These data indicate that our factors captured adult performance better than previous models, supporting the idea that our manifest variables themselves were strong choices and that these factors represent distinct constructs that better explain performance than multiple groups of manifest variables. We consider this post hoc analysis as supporting this model's validity.

Impact

For the reasons just described, we believe that the current results have value for researchers. The results may also have implications for instruction, especially the data showing increasing morphological complexity across the words in first-grade texts. This suggests that first graders might benefit from learning morphological structures within words. The first-grade morphological construct was defined largely by the number of morphemes, suggesting that the increase in morphological complexity is primarily orthographic rather than semantic. As a result, a potential implication for instruction is that novice readers might benefit from learning morphemes within word recognition lessons, possibly teaching them as sound–spelling units in the same way as *m* = /m/.

For third grade, there are fewer implications for instruction given the absence of changes in the complexity of the words in the sequence of texts that each curriculum includes. However, this suggests that teachers could choose texts across an entire curriculum without concern that later texts might have greater word complexity. The absence of increasing complexity has implications for curriculum design. As we already pointed out, the phonics/structural analysis components of these programs are inadequately focused on polymorphemic words given their increasing salience in texts across the elementary and middle school grades (e.g., Kearns et al., 2016). These lessons should align better with the growing need for morphological knowledge, and the texts should match those lessons in increasing the morphological complexity of the words across the year. Curriculum designers should also attend to the types of morphemes they address in their third-grade programs, placing greater emphasis on inflectional and derivational morphemes than on compounds.

Unanswered Questions

Our aim in this study was to describe the complexity of the words in first- and third-grade texts. All the potential analyses required to understand the demands on students' word recognition could not be addressed in a single study. We raise several critical, unanswered questions to illustrate subsequent lines of necessary work. The unanswered questions also point to limitations of and caveats about the current findings.

Word Demands in and out of Text Contexts

Our analyses address features of words in isolation from the contexts of text. Texts have numerous structures that can facilitate word recognition, including but not limited to thematic content, text structure, syntax, semantic connections among words, and linguistic features such as collocation. In early stages of reading, however, word-level variables appear critical. Five of nine variables in Fitzgerald et al.'s (2015) analysis that predicted first- and second-grade students' performance on a silent reading maze task were word-level measures, all of which related to the primary constructs examined in the current study (i.e., age of acquisition, frequency, orthographic structure). Further, all three measures related to discourse structure reflect vocabulary demands, such as the repetition of words across phrases within sentences. When texts had high levels of repetition of vocabulary, young readers were able to choose more correct choices in a maze task. At higher text complexity when these structures were no longer available, less proficient readers did more poorly. Presumably, when the percentage of words unknown to readers reaches a critical threshold, comprehension will be compromised because students' automaticity is affected (LaBerge & Samuels, 1974). Our analyses provide a portrait of the kinds and numbers of words that students will confront in typical instructional texts. Subsequent analyses are needed to establish how students of different proficiency levels navigate the word-level demands in typical texts.

The Role of Semantic Variables

The foundational work that went into the identification of variables for the current analysis was considerable. Even so, we recognize that some word features, such as word polysemy and semantic connections, were not addressed. Large-scale measures of polysemy, such as WordNet (Miller, 1995), have not been proven to have sufficient discriminatory capacity. The development of fine-tuned measures of polysemy can be time-consuming and challenging to validate (see, e.g., Cervetti, Hiebert, Pearson, & McClung, 2015).

For the current project, we examined Durda and Buchanan's (2008) procedure for establishing distance and similarity of semantic neighborhoods. Their method, WINDSORS (Windsor improved norms of distance and similarity of representations of semantics), holds promise in an area that has been troublesome to represent (e.g., Nagy & Hiebert, 2011). However, applying the WINDSORS procedure to a database of 8,550 types that appear in more than 200 unique texts (the database in this study) is currently prohibitive.

The construct of semantic relatedness illustrates one of the challenges in conducting analyses of corpora: the presence of available databases. For variables such as age of acquisition, researchers have used inventive methods to compile norms (Kuperman et al., 2012). We anticipate that additional measures will become available in the future. Whether norms of semantic relatedness can be easily established and applied to large numbers of texts remains to be seen.

Relation Between Typical Student Proficiency and Word Recognition of Current Texts

Differences in word complexity between grade levels were statistically significant, as were most word complexity variables from the beginning to the end of first-grade texts. However, a change in complexity of word features between the grades or across first grade does not provide evidence on students' ability to read the words in texts proficiently. At both first- and third-grade levels, an unaddressed question is how well most students read the current texts.

A recent analysis (Hiebert et al., 2020) described students' performances on assessment texts with word features similar to those in current first-grade texts. In that study, students in the bottom quartile read with 84% accuracy on one-minute curriculum-based-measurement texts. The words that the lowest performing students were unable to read in the spring averaged 5.5 letters, 5 phonemes, 1.5 syllables, and 1.6 morphemes—almost the precise features of the average for words in our first-grade text analysis (see Table 2).

Kilpatrick (2020) questioned why recent intervention projects have not experienced the success of projects (e.g., Torgesen et al., 2001) that motivated the Response to Intervention legislation (Individuals With Disabilities Education Act, 2004). The texts used in one of Torgesen et al.'s (2001) interventions were published in the late 1970s when primary texts had few multisyllabic words and core vocabulary was repeated (Hiebert, 2005). We believe that examinations of the match between student capacity and demands of current texts are critical if classroom instruction and Tier 2 and Tier 3 interventions are to address the needs of the substantial portion of a U.S. grade cohort that fails to attain a proficient level on the National Assessment of Educational Progress (National Center for Education Statistics, 2019).

Views of a Developmental Trajectory

Our attention to first- and third-grade texts reflected the perspective of the Common Core (National Governors Association Center for Best Practices & Council of Chief State School Officers, 2010), which in turn drew heavily on Chall's (1983) stages of reading development. In Chall's model, grades 1 and 2 are viewed as a period of building a decoding foundation, whereas the task of grades 2 and 3 is to solidify word recognition and increase fluency. According to Chall's stage theory, success with these developmental emphases will mean that students are prepared for the breadth of reading content and genres in subsequent grade levels. Whereas policymakers may view third grade as the last bastion for students to build successful reading skills (Auck & Atchison, 2016), the downward push of the reading curriculum (e.g., Bassok, Latham, & Rorem, 2016) may mean that tasks previously associated with third grade are now relegated to second grade.

Even if analyses of second-grade texts show a more clearly defined trajectory in word recognition than was evident in the third-grade texts of this analysis, questions about the match between students' capacity and grade-level expectations remain. The downward push of reading expectations, clearly evidenced by the inclusion of kindergarten in reading standards promoted by the No Child Left Behind Act (2002), has yet to be reflected in the profile of U.S. fourth-grade cohorts on the National Assessment of Educational Progress (National Center for Education Statistics, 2019). To provide an understanding of the expectations and opportunities for word recognition, comprehensive analyses are required of texts across the primary grades.

Relation Between Complex Words and Instruction

Another unanswered question pertains to the instruction that accompanies texts. Although questions remain unaddressed about the lesson-to-text match (LTTM) perspective (Stein, Johnson, & Gutlohn, 1999), a match between instruction and the words in instructional texts would be expected. LTTM was not the focus of this study, but when our findings showed a heavy proportion of polysyllabic words, we examined the scope and sequences of the three programs to determine the role of instruction on polysyllabic word structure. The three programs followed a similar scope and sequence for phonics/structural analysis instruction. Even through the end of third grade, instruction of GPCs in monosyllabic words dominated. The word recognition curriculum and its match to the word recognition task of current texts require urgent attention from researchers and curriculum developers.

Conclusion

We set out to examine the dimensions of word complexity in primary-level texts because of the importance of reading opportunities in the primary grades. The results indicate the promise of a theoretically driven factor analytic approach for characterizing word complexity, possibly with greater clarity than examining manifest variables alone. The results also indicate the need to attend to syllabic and morphemic complexity in instruction and in future research on word complexity in the instructional texts of the primary grades.

We conclude with an urgent appeal for experimental work on texts. Texts, along with readers, are central components of any reading activity (RAND Reading Study Group, 2002). The marketplace is currently filled with thousands of texts claiming a research base that will support the reading development of beginning and struggling readers. Few, if any, texts in these purportedly research-based programs have been validated in trial tests with beginning or struggling readers. Current text programs are more likely to be influenced by state and district policies than by research or theoretical frameworks. In turn, state and district policies reflect advocacy by various groups (Coburn, 2005). In-depth descriptions of the features of words in texts, such as the current study provides, can be the basis for experiments that address critical issues of practice, such as the percentage of words that students must recognize to comprehend a text successfully. The texts in which students are asked to apply their word recognition proficiencies must become a focus of experimental work if grounded solutions are to be identified for students who are currently failing to attain proficient reading levels.

NOTE

We thank Joanne Carlisle for her contributions in framing this article, Wilhelmina van Dijk for her support in conducting the data analyses, and Alia Pugh for her contributions to linguistic coding of the data.

REFERENCES

Andrews, S. (1997). The effect of orthographic similarity on lexical retrieval: Resolving neighborhood conflicts. *Psychonomic Bulletin & Review*, *4*, 439–461. https://doi.org/10.3758/BF03214334

Anglin, J.M. (1993). Vocabulary development: A morphological analysis. *Monographs of the Society for Research in Child Development*, *58*(10). https://doi.org/10.2307/1166112

Arciuli, J., Monaghan, P., & Ševa, N. (2010). Learning to assign lexical stress during reading aloud: Corpus, behavioral, and computational investigations. *Journal of Memory and Language*, *63*(2), 180–196. https://doi.org/10.1016/j.jml.2010.03.005

Auck, A., & Atchison, B. (2016). *50-state comparison: K-3 quality*. Denver, CO: Education Commission of the States.

Baayen, R.H., Milin, P., Đurđević, D.F., Hendrix, P., & Marelli, M. (2011). An amorphous model for morphological processing in visual comprehension based on naive discriminative learning. *Psychological Review*, *118*(3), 438–481. https://doi.org/10.1037/a0023851

Baayen, R.H., Piepenbrock, R., & van Rijn, H. (1995). *The CELEX lexical database* [CD-ROM]. Philadelphia, PA: Linguistic Data Consortium.

Balota, D.A., Yap, M.J., Cortese, M.J., Hutchison, K.A., Kessler, B., Loftis, B., … Treiman, R. (2007). The English Lexicon Project. *Behavior Research Methods*, *39*, 445–459. https://doi.org/10.3758/BF03193014

Bassok, D., Latham, S., & Rorem, A. (2016). Is kindergarten the new first grade? *AERA Open*, *2*(1). https://doi.org/10.1177/2332858415616358

Bauman, K., & Davis, J. (2013). Estimates of school enrollment by grade in the American Community Survey, the Current Population Survey, and the Common Core of Data (SEHSD Working Paper 2014-7). Washington, DC: Social, Economic, and Housing Statistics Division, U.S. Census Bureau.

Beck, I.L., & McCaslin, E.S. (1978). *An analysis of dimensions that affect the development of code-breaking ability in eight beginning reading programs*. Pittsburgh, PA: Learning Research and Development Center, University of Pittsburgh.

Braeken, J., & van Assen, M.A. (2017). An empirical Kaiser criterion. *Psychological Methods*, *22*(3), 450–466. https://doi.org/10.1037/met0000074

Breland, H.M. (1996). Word frequency and word difficulty: A comparison of counts in four corpora. *Psychological Science*, *7*(2), 96–99. https://doi.org/10.1111/j.1467-9280.1996.tb00336.x

Browne, M.W. (2001). An overview of analytic rotation in exploratory factor analysis. *Multivariate Behavioral Research*, *36*(1), 111–150. https://doi.org/10.1207/S15327906MBR3601_05

Brysbaert, M., Warriner, A.B., & Kuperman, V. (2014). Concreteness ratings for 40 thousand generally known English word lemmas. *Behavior Research Methods*, *46*(3), 904–911. https://doi.org/10.3758/s13428-013-0403-5

Cain, K., Compton, D.L., & Parrila, R.K. (Eds.). (2017). *Theories of reading development*. Philadelphia, PA: John Benjamins.

California State Board of Education. (2015). *2015 English language arts/English language development instructional materials adoption (K–8)*. Sacramento: Author.

Carlisle, J.F., & Katz, L.A. (2006). Effects of word and morpheme familiarity on reading of derived words. *Reading and Writing*, *19*, 669–693. https://doi.org/10.1007/s11145-005-5766-2

Carlisle, J.F., & Kearns, D.M. (2017). Learning to read morphologically-complex words. In K. Cain, D.L. Compton, & R.K. Parrila (Eds.), *Theories of reading development* (pp. 191–214). Philadelphia, PA: John Benjamins.

Cervetti, G.N., Hiebert, E.H., Pearson, P.D., & McClung, N.A. (2015). Factors that influence the difficulty of science words. *Journal of Literacy Research*, *47*(2), 153–185. https://doi.org/10.1177/1086296X15615363

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

Chall, J.S. (1983). *Stages of reading development*. New York, NY: McGraw-Hill.

Chen, F.F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling*, *14*(3), 464–504. https://doi.org/10.1080/10705510701301834

Chetail, F., Balota, D., Treiman, R., & Content, A. (2015). What can megastudies tell us about the orthographic structure of English words? *Quarterly Journal of Experimental Psychology*, *68*(8), 1519–1540. https://doi.org/10.1080/17470218.2014.963628

Coburn, C.E. (2005). The role of nonsystem actors in the relationship between policy and practice: The case of reading instruction in California. *Educational Evaluation and Policy Analysis*, *27*(1), 23–52. https://doi.org/10.3102/01623737027001023

Coleman, D., & Pimentel, S. (2012). *Revised publishers' criteria for the Common Core State Standards in English language arts and literacy, grades 3–12*. Retrieved from http://www.corestandards.org/assets/Publishers_Criteria_for_3-12.pdf

Coltheart, M., Davelaar, E., Jonasson, J.T., & Besner, D. (1977). Access to the internal lexicon. In S. Dornič (Ed.), *Attention and performance VI: Proceedings of the sixth International Symposium on Attention and Performance, Stockholm, Sweden, July 28–August 1, 1975* (pp. 535–555). Hillsdale, NJ: Erlbaum.

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. https://doi.org/10.1037/0012-1649.33.6.934

Dawson, N., Rastle, K., & Ricketts, J. (2018). Morphological effects in visual word recognition: Children, adolescents, and adults. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *44*(4), 645–654. https://doi.org/10.1037/xlm0000485

De Jong, N.H., IV, Schreuder, R., & Baayen, R.H. (2000). The morphological family size effect and morphology. *Language and Cognitive Processes*, *15*(4/5), 329–365. https://doi.org/10.1080/01690960050119625

Durda, K., & Buchanan, L. (2008). WINDSORS: Windsor improved norms of distance and similarity of representations of semantics. *Behavior Research Methods*, *40*(3), 705–712. https://doi.org/10.3758/BRM.40.3.705

Education Market Research. (2014). *Reading market 2013*. New York, NY: Author.

Ehri, L.C. (2005). Learning to read words: Theory, findings, and issues. *Scientific Studies of Reading*, *9*(2), 167–188. https://doi.org/10.1207/s1532799xssr0902_4

Ehri, L.C. (2017). Orthographic mapping and literacy development revisited. In K. Cain, D.L. Compton, & R.K. Parrila (Eds.), *Theories of reading development* (pp. 127–146). Philadelphia, PA: John Benjamins.

Fitt, S. (2001). *Unisyn lexicon release, version 1.3* [Data file and codebook]. Edinburgh, Scotland: Centre for Speech Technology Research, University of Edinburgh.

Fitzgerald, J., Elmore, J., Koons, H., Hiebert, E.H., Bowen, K., Sanford-Moore, E.E., & Stenner, A.J. (2015). Important text characteristics for early-grades text complexity. *Journal of Educational Psychology*, *107*(1), 4–29. https://doi.org/10.1037/a0037289

Florida Department of Education. (2013). *Instructional materials approved for adoption 2012–2013*. Tallahassee: Author. Retrieved from http://www.fldoe.org/core/fileparse.php/7536/urlt/1213aim.pdf

Foorman, B.R., Francis, D.J., Davidson, K.C., Harm, M.W., & Griffin, J. (2004). Variability in text features in six grade 1 basal reading programs. *Scientific Studies of Reading*, *8*(2), 167–197. https://doi.org/10.1207/s1532799xssr0802_4

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. https://doi.org/10.1080/10888438.2015.1026969

Gamson, D.A., Lu, X., & Eckert, S.A. (2013). Challenging the research base of the Common Core State Standards: A historical reanalysis of text complexity. *Educational Researcher*, *42*(7), 381–391. https://doi.org/10.3102/0013189X13505684

Glushko, R.J. (1979). The organization and activation of orthographic knowledge in reading aloud. *Journal of Experimental Psychology: Human Perception and Performance*, *5*(4), 674–691. https://doi.org/10.1037/0096-1523.5.4.674

Gonnerman, L.M., Seidenberg, M.S., & Andersen, E.S. (2007). Graded semantic and phonological similarity effects in priming: Evidence for a distributed connectionist approach to morphology. *Journal of Experimental Psychology: General*, *136*(2), 323–345. https://doi.org/10.1037/0096-3445.136.2.323

Goodwin, A.P. (2016). Effectiveness of word solving: Integrating morphological problem-solving within comprehension instruction for middle school students. *Reading and Writing*, *29*, 91–116. https://doi.org/10.1007/s11145-015-9581-0

Goodwin, A.P., & Ahn, S. (2013). A meta-analysis of morphological interventions in English: Effects on literacy outcomes for school-age children. *Scientific Studies of Reading*, *17*(4), 257–285. https://doi.org/10.1080/10888438.2012.689791

Graesser, A.C., McNamara, D.S., & Kulikowich, J.M. (2011). Coh-Metrix: Providing multilevel analyses of text characteristics. *Educational Researcher*, *40*(5), 223–234. https://doi.org/10.3102/0013189X11413260

Graves, M.F., Elmore, J., & Fitzgerald, J. (2019). The vocabulary of core reading programs. *The Elementary School Journal*, *119*(3), 386–416. https://doi.org/10.1086/701653

Harm, M.W., & Seidenberg, M.S. (2004). Computing the meanings of words in reading: Cooperative division of labor between visual and phonological processes. *Psychological Review*, *111*(3), 662–720. https://doi.org/10.1037/0033-295X.111.3.662

Hayes, D.P., Wolfer, L.T., & Wolfe, M.F. (1996). Schoolbook simplification and its relation to the decline in SAT-verbal scores. *American Educational Research Journal*, *33*(2), 489–508. https://doi.org/10.3102/00028312033002489

Hiebert, E.H. (2005). State reform policies and the task textbooks pose for first-grade readers. *The Elementary School Journal*, *105*(3), 245–266. https://doi.org/10.1086/428743

Hiebert, E.H., & Fisher, C.W. (2007). Critical word factor in texts for beginning readers. *The Journal of Educational Research*, *101*(1), 3–11. https://doi.org/10.3200/JOER.101.1.3-11

Hiebert, E.H., Goodwin, A.P., & Cervetti, G.N. (2018). Core vocabulary: Its morphological content and presence in exemplar texts. *Reading Research Quarterly*, *53*(1), 29–49. https://doi.org/10.1002/rrq.183

Hiebert, E.H., Toyama, Y., & Irey, R. (2020). Features of known and unknown words by first graders of different proficiency levels in winter and spring. *Education in Science*, *10*(12), Article 389. https://doi.org/10.3390/educsci10120389

Hoover, W.A., & Gough, P.B. (1990). The simple view of reading. *Reading and Writing*, *2*(2), 127–160. https://doi.org/10.1007/BF00401799

Hu, L.T., & Bentler, P.M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, *6*(1), 1–55. https://doi.org/10.1080/10705519909540118

Individuals With Disabilities Education Act, 20 U.S.C. § 1400 *et seq*. (2004).

Jorgensen, T.D., Pornprasertmanit, S., Schoemann, A.M., Rosseel, Y., Miller, P., & Quick, C. (2018). *semTools: Useful tools for structural equation modeling* (R package version 0.5-1) [Computer software].

Juel, C., & Minden-Cupp, C. (2000). Learning to read words: Linguistic units and instructional strategies. *Reading Research Quarterly*, *35*(4), 458–492. https://doi.org/10.1598/RRQ.35.4.2

Juel, C., & Roper/Schneider, D. (1985). The influence of basal readers on first grade reading. *Reading Research Quarterly*, *20*(2), 134–152. https://doi.org/10.2307/747751

Kearns, D.M. (2015). How elementary-age children read polysyllabic polymorphemic words. *Journal of Educational Psychology*, *107*(2), 364–390. https://doi.org/10.1037/a0037518

Kearns, D.M. (2020). Does English have useful syllable division patterns? *Reading Research Quarterly*, *55*(S1), S145–S160. https://doi.org/10.1002/rrq.342

Kearns, D.M., Steacy, L.M., Compton, D.L., Gilbert, J.K., Goodwin, A.P., Cho, E., … Collins, A.A. (2016). Modeling polymorphemic word recognition: Exploring differences among children with early-emerging and late-emerging word reading difficulty. *Journal of Learning Disabilities*, *49*(4), 368–394. https://doi.org/10.1177/0022219414554229

Kilpatrick, D. (2020). The study that prompted Tier 2 of RTI: *Why aren't our Tier 2 results as good? The Reading League Journal*, *1*(1), 28–32.

Kline, R.B. (2005). *Principles and practice of structural equation modeling* (2nd ed.). New York, NY: Guilford.

Kuperman, V., Stadthagen-Gonzalez, H., & Brysbaert, M. (2012). Age-of-acquisition ratings for 30,000 English words. *Behavior Research Methods*, *44*(4), 978–990. https://doi.org/10.3758/s13428-012-0210-4

LaBerge, D., & Samuels, S.J. (1974). Toward a theory of automatic information processing in reading. *Cognitive Psychology*, *6*(2), 293–323. https://doi.org/10.1016/0010-0285(74)90015-2

Lesgold, A., Resnick, L.B., & Hammond, K. (1985). Learning to read: A longitudinal study of word skill development in two curricula. In G.E. MacKinnon & T.G. Waller (Eds.), *Reading research: Advances in theory and practice* (Vol. 4, pp. 107–138). New York, NY: Academic.

Levenshtein, V.I. (1966). Binary codes capable of correcting deletions, insertions, and reversals. *Soviet Physics Doklady*, *10*(8), 707–710.

Levesque, K.C., Kieffer, M.J., & Deacon, S.H. (2017). Morphological awareness and reading comprehension: Examining mediating factors. *Journal of Experimental Child Psychology*, *160*, 1–20. https://doi.org/10.1016/j.jecp.2017.02.015

MacCallum, R.C., Browne, M.W., & Sugawara, H.M. (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods*, *1*(2), 130–149. https://doi.org/10.1037/1082-989X.1.2.130

Marinus, E., & de Jong, P.F. (2010). Variability in the word-reading performance of dyslexic readers: Effects of letter length, phoneme length and digraph presence. *Cortex*, *46*(10), 1259–1271. https://doi.org/10.1016/j.cortex.2010.06.005

Masterson, J., Stuart, M., Dixon, M., & Lovejoy, S. (2010). Children's printed word database: Continuities and changes over time in children's early reading vocabulary. *British Journal of Psychology*, *101*(2), 221–242. https://doi.org/10.1348/000712608X371744

Menon, S., & Hiebert, E.H. (2005). A comparison of first graders' reading with little books or literature-based basal anthologies. *Reading Research Quarterly*, *40*(1), 12–38. https://doi.org/10.1598/RRQ.40.1.2

Miller, G.A. (1995). WordNet: A lexical database for English. *Communications of the ACM*, *38*(11), 39–41. https://doi.org/10.1145/219717.219748

Moats, L.C. (1999). *Teaching reading is rocket science: What expert teachers of reading should know and be able to do*. Washington, DC: American Federation of Teachers.

Murray, M.S., Munger, K.A., & Hiebert, E.H. (2014). An analysis of two reading intervention programs: How do the words, texts, and programs compare? *The Elementary School Journal*, *114*(4), 479–500. https://doi.org/10.1086/675635

Nagy, W.E., & Anderson, R.C. (1984). How many words are there in printed school English? *Reading Research Quarterly*, *19*(3), 304–330. https://doi.org/10.2307/747823

Nagy, W.E., & Hiebert, E.H. (2011). Toward a theory of word selection. In M.L. Kamil, P.D. Pearson, E.B. Moje, & P.P. Afflerbach (Eds.), *Handbook of reading research* (Vol. 4, pp. 388–404). New York, NY: Routledge.

National Center for Education Statistics. (2019). *The Nation's Report Card: NAEP reading assessment*. Washington, DC: National Center

for Education Statistics, Institute of Education Sciences, U.S. Department of Education.
National Governors Association Center for Best Practices & Council of Chief State School Officers. (2010). *Common Core State Standards for English language arts and literacy in history/social studies, science, and technical subjects*. Washington, DC: Authors.
National Institute of Child Health and Human Development. (2000). *Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction: Reports of the subgroups* (NIH Publication No. 00-4754). Washington, DC: U.S. Government Printing Office.
No Child Left Behind Act of 2001, 20 U.S.C. § 6301 *et seq*. (2002).
Perfetti, C.A. (2003). The universal grammar of reading. *Scientific Studies of Reading*, *7*(1), 3–24. https://doi.org/10.1207/S1532799XSSR0701_02
Perfetti, C.A. (2007). Reading ability: Lexical quality to comprehension. *Scientific Studies of Reading*, *11*(4), 357–383. https://doi.org/10.1080/10888430701530730
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). Amsterdam, Netherlands: John Benjamins.
Perry, C., Ziegler, J.C., & Zorzi, M. (2010). Beyond single syllables: Large-scale modeling of reading aloud with the connectionist dual process (CDP++) model. *Cognitive Psychology*, *61*(2), 106–151. https://doi.org/10.1016/j.cogpsych.2010.04.001
Plaut, D.C., McClelland, J.L., Seidenberg, M.S., & Patterson, K. (1996). Understanding normal and impaired word reading: Computational principles in quasi-regular domains. *Psychological Review*, *103*(1), 56–115. https://doi.org/10.1037/0033-295X.103.1.56
RAND Reading Study Group. (2002). *Reading for understanding: Toward an R&D program in reading comprehension*. Santa Monica, CA: Rand.
R Core Team. (2019). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.
Roberts, T.A., Christo, C., & Shefelbine, J.A. (2011). Word recognition. In M.L. Kamil, P.D. Pearson, E.B. Moje, & P.P. Afflerbach (Eds.), *Handbook of reading research* (Vol. 4, pp. 229–258). New York, NY: Routledge.
Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, *48*(2), 1–36. https://doi.org/10.18637/jss.v048.i02
Rueschemeyer, S., & Gaskell, M.G. (Eds.). (2018). *The Oxford handbook of psycholinguistics* (2nd ed.). New York, NY: Oxford University Press.
Seidenberg, M.S. (1987). Sublexical structures in visual word recognition: Access units or orthographic redundancy? In M. Coltheart (Ed.), *Attention and performance XII: The psychology of reading* (pp. 245–263). Hillsdale, NJ: Erlbaum.
Seymour, P.H.K. (1997). Foundations of orthographic development. In C.A. Perfetti, L. Rieben, & M. Fayol (Eds.), *Learning to spell: Research, theory, and practice across languages* (pp. 319–338). Mahwah, NJ: Erlbaum.
Seymour, P.H.K., Aro, M., & Erskine, J.M. (2003). Foundation literacy acquisition in European orthographies. *British Journal of Psychology*, *94*(2), 143–174. https://doi.org/10.1348/000712603321661859
Share, D.L. (1995). Phonological recoding and self-teaching: *Sine qua non* of reading acquisition. *Cognition*, *55*(2), 151–218. https://doi.org/10.1016/0010-0277(94)00645-2
Statacorp. (2017). *Stata statistical software: Release 15* [Computer software]. College Station, TX: Author.
Stein, M., Johnson, B., & Gutlohn, L. (1999). Analyzing beginning reading programs: The relationship between decoding instruction and text. *Remedial and Special Education*, *20*(5), 275–287. https://doi.org/10.1177/074193259902000503
Stenner, A.J., Burdick, H., Sanford, E.E., & Burdick, D.S. (2007). *The Lexile framework for reading technical report*. Durham, NC: MetaMetrics.
Texas Education Agency. (2015). *Instructional materials current adoption bulletin (current year = 2015–2016)*. Retrieved from https://repositories.lib.utexas.edu/bitstream/handle/2152/31600/TEA%20adoption%20bulletin%202015.pdf?sequence=3&isAllowed=y
Thurstone, L.L. (1947). *Multiple factor analysis*. Chicago, IL: University of Chicago Press.
Torgesen, J.K., Alexander, A.W., Wagner, R.K., Rashotte, C.A., Voeller, K.K., & Conway, T. (2001). Intensive remedial instruction for children with severe reading disabilities: Immediate and long-term outcomes from two instructional approaches. *Journal of Learning Disabilities*, *34*(1), 33–58. https://doi.org/10.1177/002221940103400104
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. https://doi.org/10.1037/0096-3445.124.2.107
University of Essex, Department of Psychology. (2003). *Children's Printed Word Database* (Rev. ed.). Retrieved from https://www1.essex.ac.uk/psychology/cpwd/
Wells, J.C. (1982). *Accents of English*. Cambridge, UK: Cambridge University Press.
Yap, M.J., & Balota, D.A. (2009). Visual word recognition of multisyllabic words. *Journal of Memory and Language*, *60*(4), 502–529. https://doi.org/10.1016/j.jml.2009.02.001
Yap, M.J., & Balota, D.A. (2015). Visual word recognition. In A. Pollatsek & R. Treiman (Eds.), *The Oxford handbook of reading* (pp. 26–43). Oxford, UK: Oxford University Press.
Yarkoni, T., Balota, D., & Yap, M. (2008). Moving beyond Coltheart's *N*: A new measure of orthographic similarity. *Psychonomic Bulletin & Review*, *15*, 971–979. https://doi.org/10.3758/PBR.15.5.971
Zeno, S.M., Ivens, S.H., Millard, R.T., & Duvvuri, R. (1995). *The educator's word frequency guide*. Brewster, NY: Touchstone Applied Science Associates.
Zoccolotti, P., De Luca, M., Di Filippo, G., Judica, A., & Martelli, M. (2009). Reading development in an orthographically regular language: Effects of length, frequency, lexicality and global processing ability. *Reading and Writing*, *22*, Article 1053. https://doi.org/10.1007/s11145-008-9144-8

PROGRAMS IN THE DATABASE

Afflerbach, P., Blachowicz, C., Boyd, C.D., Izquierdo, E., Juel, C., Kame'enui, E., … Wixson, K.K. (2013). *Scott Foresman Reading Street: Common Core*. Glenview, IL: Pearson.
August, D., Bear, D.R., Dole, J.A., Echevarria, J., Fisher, D., Francis, D., … Tinajero, J.V. (2014). *McGraw-Hill Reading Wonders: CCSS reading language arts program*. New York, NY: McGraw-Hill.
Baumann, J.F., Chard, D.J., Cooks, J., Cooper, J.D., Gersten, R., Lipson, M., … Vogt, M. (2014). *Journeys: Common Core*. Orlando, FL: Houghton Mifflin Harcourt.

Submitted April 19, 2018
Final revision received February 2, 2021
Accepted February 3, 2021

DEVIN M. KEARNS (corresponding author) is an associate professor in the Department of Educational Psychology at the University of Connecticut, Storrs, USA; email devin.kearns@uconn.edu. He researches reading disability and focuses on designing and testing reading interventions and studying how educational practice, cognitive science, and neuroscience can be linked.

ELFRIEDA H. HIEBERT is the CEO and president of TextProject, Santa Cruz, California, USA; email hiebert@textproject.org. Her research addresses how fluency, vocabulary, and knowledge can be fostered through appropriate texts.

Research Review of Word Features in the Texts of the Primary-Grade Period

Studies of the word features of primary-level texts fall into four categories: unique words relative to total words, portions of content or rare words, specific word features, and connections between words in student texts and lessons in teacher's guides. The diversity in focus of studies makes it difficult to compare contributions of the analyses in understanding word complexity for primary-level readers. Consequently, we applied analytic schemes to the same set of texts: the last first- and third-grade text of three core reading programs (see Table A1).

Not included in this review are projects that provide databases of unique words in beginning texts, such as that of Graves, Elmore, and Fitzgerald (2019) for U.S. texts and of Masterson, Stuart, Dixon, and Lovejoy (2010) for U.K. texts. Words in these lists are available for research or curricular purposes, but the complexity of the word recognition task within and across programs is difficult to establish because these lists fail to indicate where and with what frequency words appear in specific texts or programs.

Type/Token Ratio and Descriptions of Types of Unique Words

Chall (1967), in one of the initial studies of the demands of beginning reading texts, used a measure that she described as vocabulary load but, more typically, is referred to as type/token ratio. *Types* refers to the different or unique words, and *tokens* refers to the total number of words. Chall concluded that prominent core reading programs of the era had one to two unique, new words per 100 running words of text (i.e., a type/token ratio of 1:100 to 2:100), a number that, according to Chall, remained constant from the beginning of first grade to the end of third grade.

Similar to Chall (1967), Hiebert (2005) used type/token ratio to describe the shifts in first-grade texts from 1960 to the early 2000s. Using digital technology, these analyses showed that Chall's manual counts had produced underestimates of the unique words per 100 for the early 1960s copyrights: First-grade texts had a 8:100 type/token ratio at the beginning of grade 1 and a 10:100 ratio at year's end. Analyses of type/token ratio in Table A1 show that type/token ratios in texts of a similar level in the 2010s were more than double those of texts of the 1960s. Numerous different words are likely to pose a more challenging task for beginning readers than fewer different words are. Yet, features of individual words, such as length, frequency, and consistency of grapheme–phoneme patterns, also need to be considered.

Vocabulary

Recognizing that Chall's (1967) vocabulary load measure did not distinguish grammatical vocabulary (e.g., *of*, *the*) from content vocabulary (e.g., *balloon*, *horse*), Hayes et al. (1996) created the LEX measure, a logarithm of average frequency of words in a text with grammatical words eliminated. Applying this measure to texts from 1946–1962 and 1963–1991, Hayes et al. reported a decline in the LEX (i.e., vocabulary had gotten harder) for both first- and third-grade texts. In a subsequent analysis using the LEX measure but with a substantially larger sample of third-grade texts over the entire 20th century, Gamson et al. (2013) concluded that LEX scores for third-grade texts were more challenging for the 1990s than for all earlier decades.

The application of this measure of content vocabulary in Table A1 indicates that, as would be expected, first-grade texts are easier than third-grade texts. What these scores mean for developing readers, however, remains uncertain. For example, two words in a first-grade text, *declaration* and *bang*, have similar ratings for the standard frequency index (50.6 and 49.6, respectively), even though recognizing these words likely involves different levels of proficiency.

An alternative perspective on vocabulary frequency was adopted by Hiebert and Fisher (2007), who established the percentage of unique words with a frequency of fewer than nine appearances per million in the Zeno et al. (1995) database, a measure described as the critical word factor—words that appear infrequently in texts. As shown in Table A1, first-grade texts have fewer rare words than third-grade texts do, but similar to the LEX measure, the critical word factor fails to distinguish between the complexity of words in the rare category, such as *declaration* and *bang*.

Text-Specific Features

Based on first and second graders' performances on a maze assessment, Fitzgerald et al. (2015) created the Early Literacy

TABLE A1
Application of Text Complexity Systems to Illustrative First- and Third-Grade Texts

	New words/100[a]				Lexile and early literacy indicators[d]						
Text (reading program)	Single text	All texts	Frequency of content vocabulary[b]	Critical word factor[c]	Lexile level	Mean sentence length	Mean log word frequency	Decodability	Semantics	Text structure	Syntax
First grade											
"Winners Never Quit!" (HMH)	43	16	2.87	11.60	470	5.67	3.50	3	3	4	4
"Happy Birthday, USA" (MH)	49	16	2.80	5.03	550	8.89	3.71	4	5	5	4
"A Stone Garden" (SF)	48	16	2.77	9.45	560	8.53	3.57	4	4	5	4
Third grade											
"Climbing Mount Everest" (HMH)	42	14	2.50	10.96	680	9.05	3.34				
"Alligators & Crocodiles" (MH)	42	14	2.52	12.81	900	12.16	3.37				
"Atlantis" (SF)	43	14	2.55	11.19	960	14.66	3.47				

Note.. HMH = Houghton Mifflin Harcourt's Journeys (Baumann et al., 2014); MH = McGraw-Hill's Wonders (August et al., 2014); SF = Scott Foresman's Reading Street (Afflerbach et al., 2013).
[a]Based on Chall (1967) and Hiebert (2005). [b]Based on Hayes et al. (1996) and Gamson, Lu, and Eckert (2013) using mean log word frequency (Stenner, Burdick, Sanford, & Burdick, 2007). [c]Based on Hiebert and Fisher (2007). [d]Based on Fitzgerald et al. (2015).

Indicator (ELI) system that consists of four constructs: decoding, semantics, structure, and syntax. All constructs except syntax reflect a cluster of variables: decoding (complexity of vowel patterns; Menon & Hiebert, 2005) and word length in syllables, semantic load (age of acquisition, word abstractness, and word rareness), and discourse structure (phrase diversity, information load/density, and noncompressibility). In the typical reporting (Fitzgerald et al., 2015), however, data are only provided for the four constructs in the form of a scale from 1 (*very low*) to 5 (*very high*).

The ELI data are provided through approximately 650 Lexile levels, which corresponds to the end-of-second-grade level (Coleman & Pimentel, 2012). Hence, third-grade texts cannot be compared with first-grade texts using the ELI measures, but they can be compared on the Lexile components of mean log word frequency and mean sentence length (Stenner et al., 2007). The application of ELI and Lexile systems in Table A1 shows that third-grade texts have less frequent vocabulary than first-grade texts have. Neither system, however, indicates the kinds of words that students will need to be able to recognize in texts.

LTTM

Following Chall's (1967) criticism of the lack of congruity between lessons in teacher's guides and student texts of core reading programs, Beck and McCaslin (1978) analyzed the match between instruction in GPCs and presence of words with those patterns in student texts from eight reading programs. Beck and McCaslin concluded that the match between lessons and the words in student texts of code emphasis programs would support students in reading development; however, the researchers deemed the phonics instruction offered by four mainstream programs as failing to provide the foundation for word recognition in texts.

Stein et al. (1999) extended Beck and McCaslin's (1978) work by establishing a metric called LTTM, which was defined as the match between lessons on GPCs in teacher's guides and the words in student texts. For example, an 80% decodable text meant that 80% of words had GPCs or were high-frequency words that had been taught in lessons. Foorman, Francis, Davidson, Harm, and Griffin (2004) applied this metric to determine whether texts adopted for use in Texas fit the state's requirement for a 75% LTTM. Potential decoding accuracy rates varied widely across the six programs and often depended heavily on holistically taught words.

This model continues to be applied to the texts of early reading programs (e.g., Murray, Munger, & Hiebert, 2014). However, we do not provide data on LTTM in Table A1 because, with few exceptions (typically proper names), GPCs in words have been covered in lessons. For example, the first-grade curriculum of programs in Table A1 covers all phonemes for vowels (Moats, 1999) except schwa, which is covered in grade 2. Further, words with rare GPCs are taught as sight words. That is, the LTTM for the final text of a grade is invariably 100%. Further, no research has yet confirmed that a single lesson (which is the case for many vowel GPCs in the core programs) is sufficient for acquiring a pattern. The LTTM measure provides little information on the word recognition proficiencies required by readers to read texts successfully.

APPENDIX B

Comparison of the Current Corpus With Other U.S. and U.K. Corpora

The analysis shown in Table B1 indicates that corpora (Graves et al., 2019) for texts used in grades 1 and 3 for reading instruction in the United States are similar in most features to texts used in levels 1 and 3 in the United Kingdom from the University of Essex (2003; see also Masterson et al., 2010). Differences are apparent in the Graves et al. (2019) corpus in having words that are somewhat less frequent and having a higher age of acquisition. As a corpus becomes larger, the expectation is that the number of rare words increases substantially. From that perspective, the distributions of the Essex and Graves et al. databases would suggest that the typical U.S. program has substantially more rare words than a U.K. program.

TABLE B1
Distribution of Words According to Frequency

Corpus	Unique	Total	Length	Frequency	Age of acquisition	Dispersion	Concreteness
U.K. levels 1 and 3[a]	9,603	593,215	6.5	126.3	6.8	.63	3.7
Graves et al. (2019) corpus for grades 1 and 3	10,342	417,054	6.2	98.9	7.7	.63	3.7

[a]Sources: Masterson et al. (2010) and University of Essex (2003).

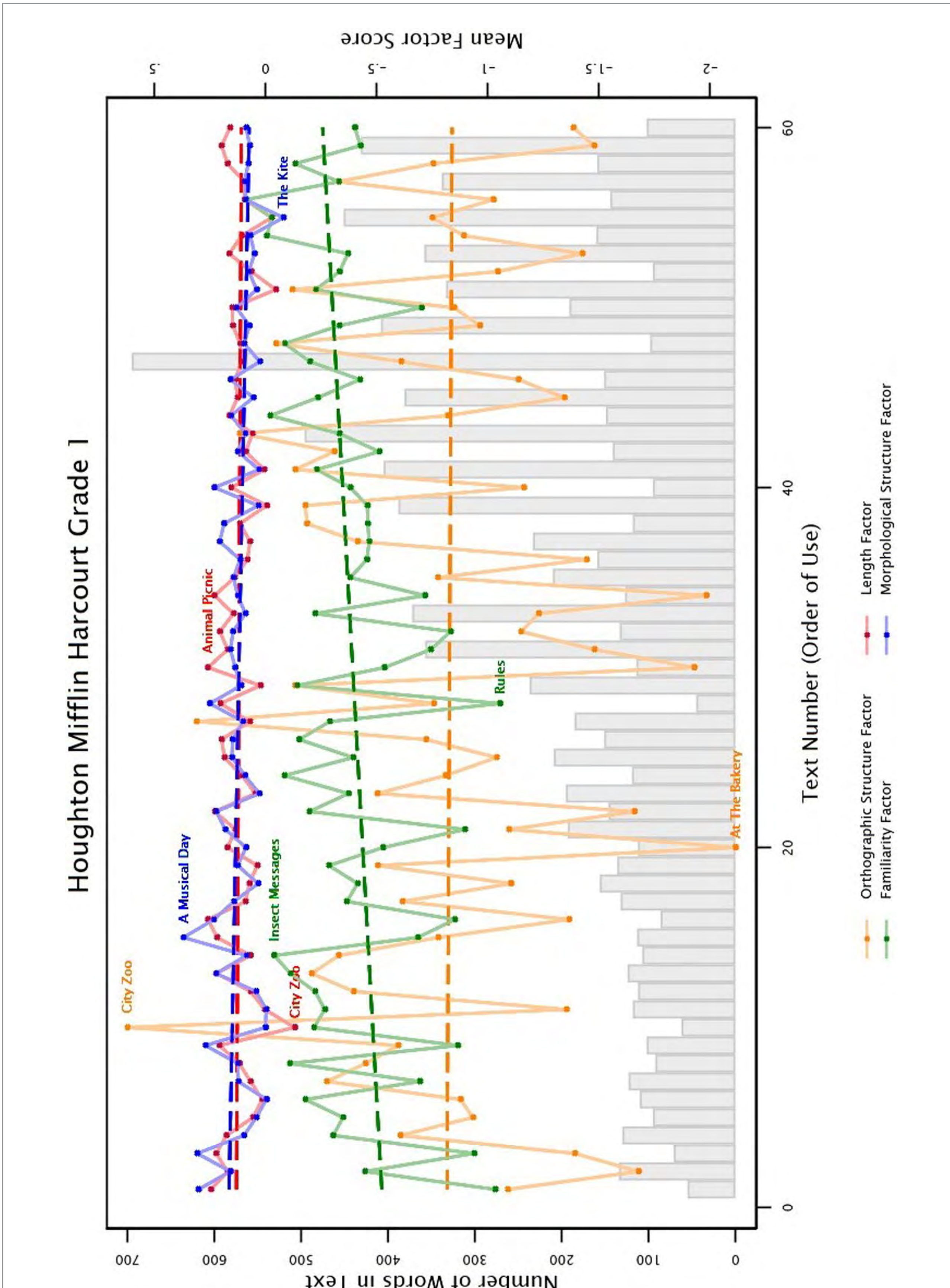
APPENDIX C

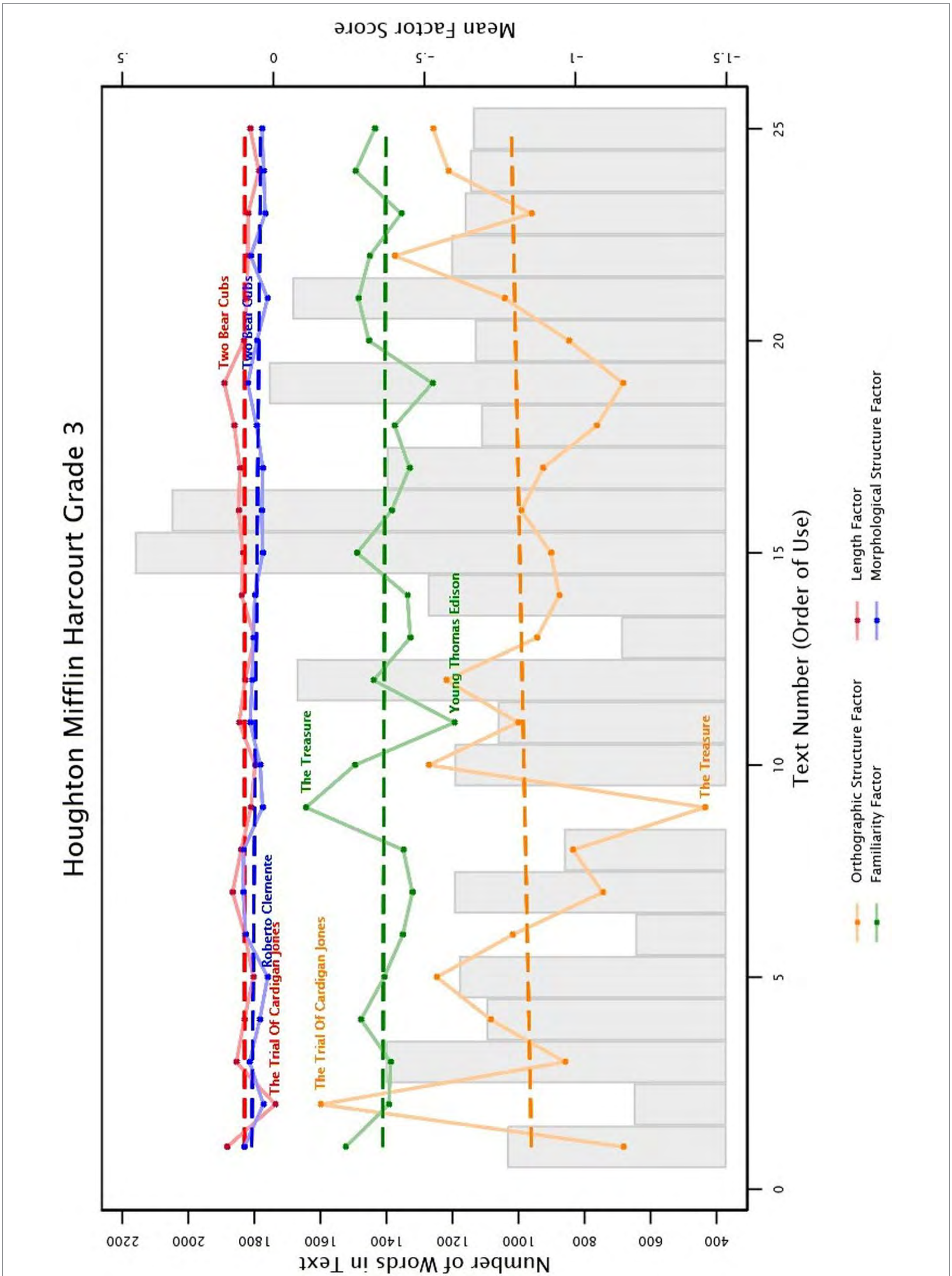
Criteria for Inclusion and Exclusion of Characters in Text

All nonword formatting was removed (e.g., ellipses, quotation marks). Words that contained internal hyphens (e.g., *extra-long*) were split into two separate words. Numerals were not counted as words (e.g., *11* from *June 11*), but associated words (i.e., *June*) were retained. Numbers written as words (e.g., *twenty*) were counted as words. All other words were retained, such as proper names (e.g., *Makoto*), foreign words (e.g., *mariachi*), and colloquial expressions (e.g., *thingamajigs*).

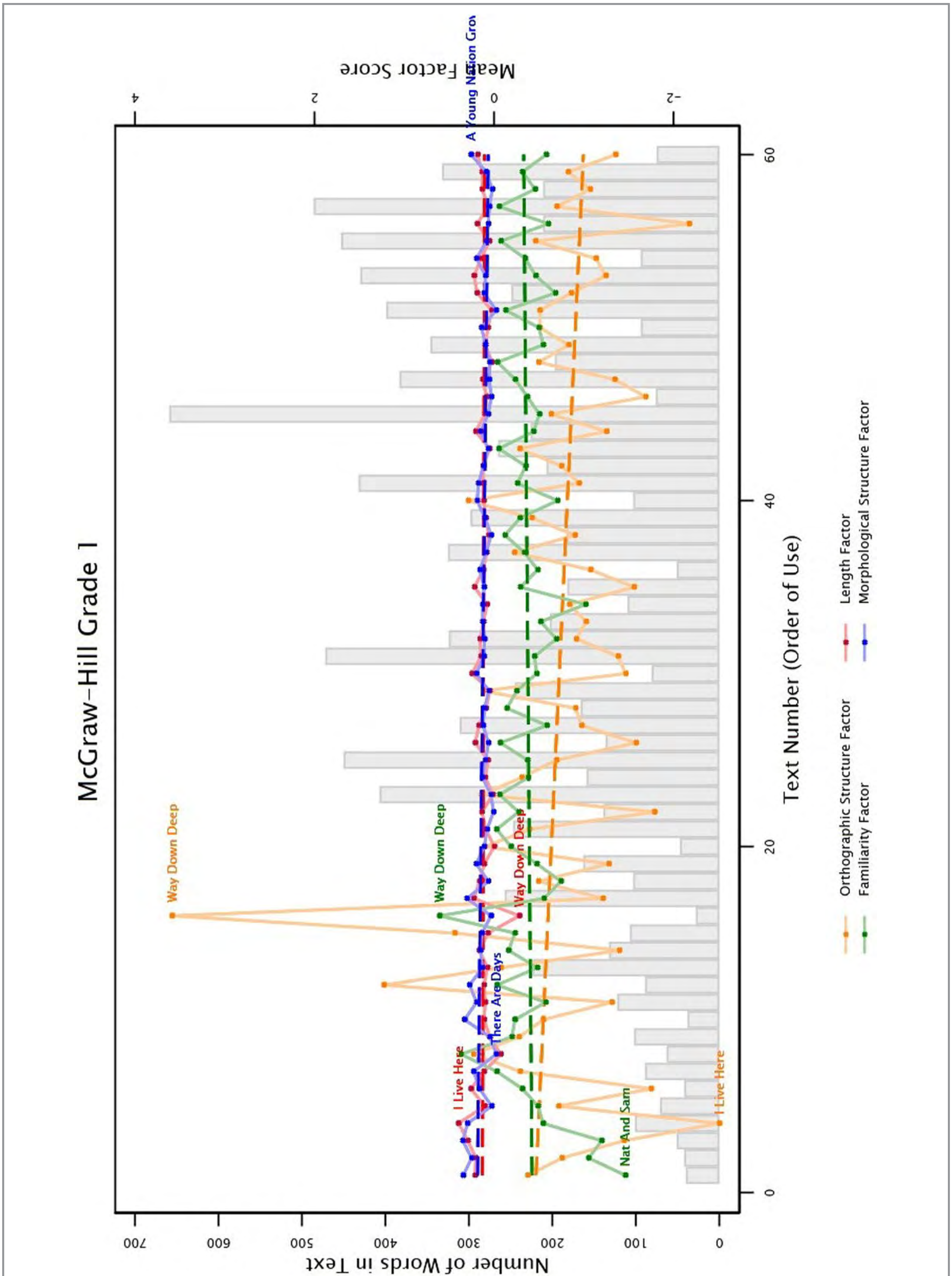
APPENDIX D

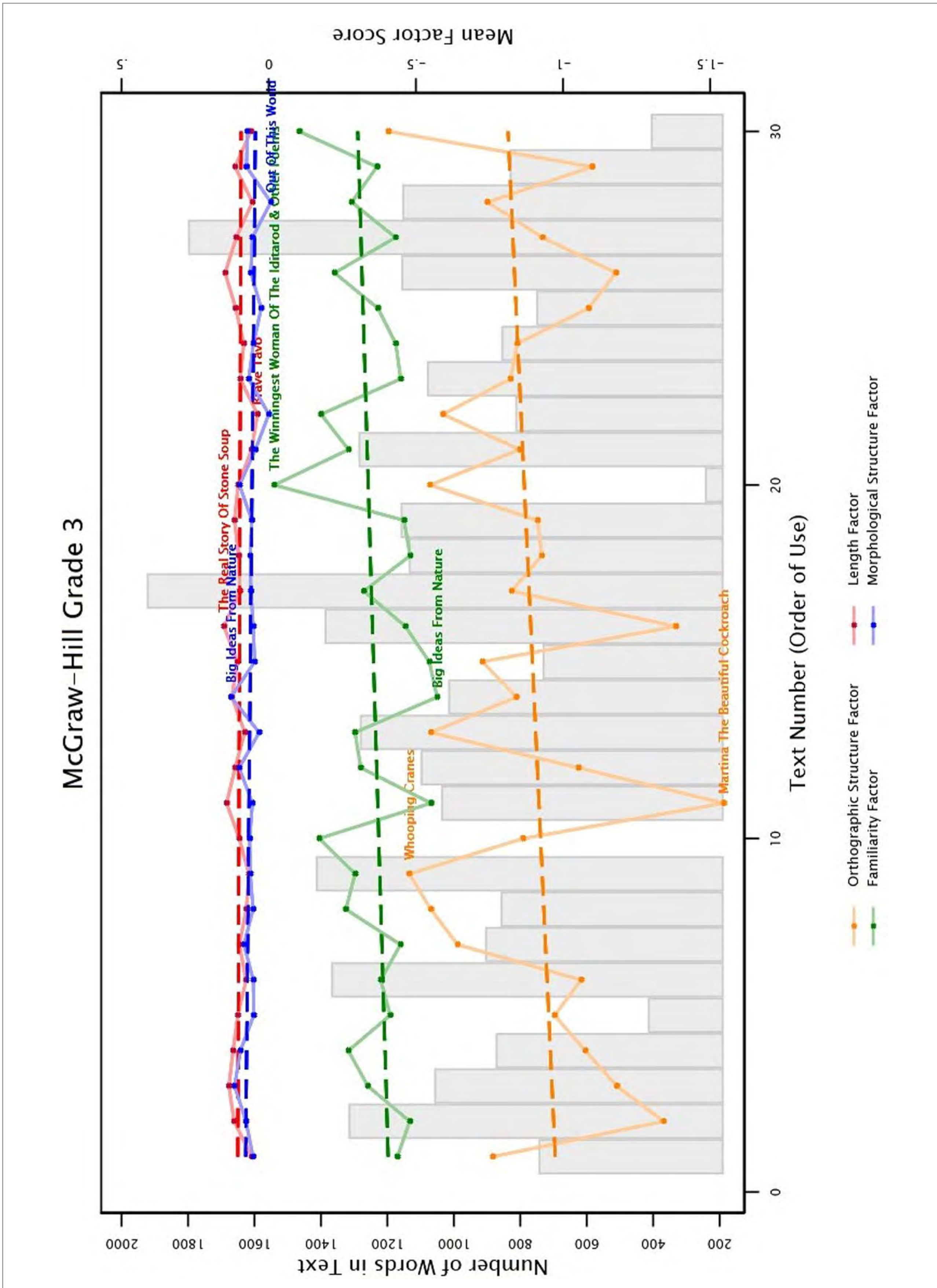
Trajectories of Factor-Level Change for Each Program


Houghton Mifflin Harcourt Grade 1
Mean Factor Score
Number of Words in Text
Text Number (Order of Use)
City Zoo
A Musical Day
Insect Messages
City Zoo
Animal Picnic
Rules
At The Bakery
The Kite
Orthographic Structure Factor
Familiarity Factor
Length Factor
Morphological Structure Factor

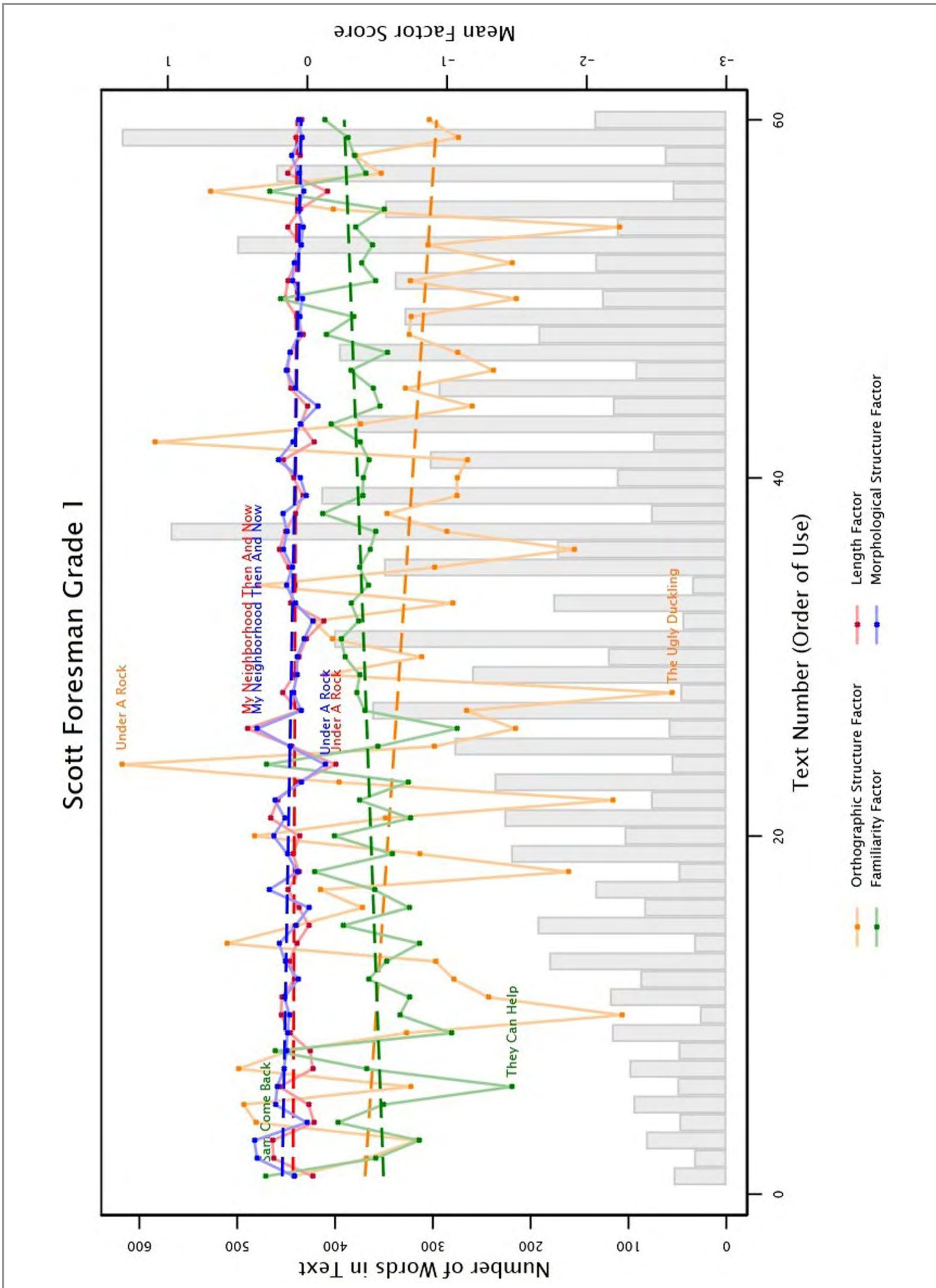


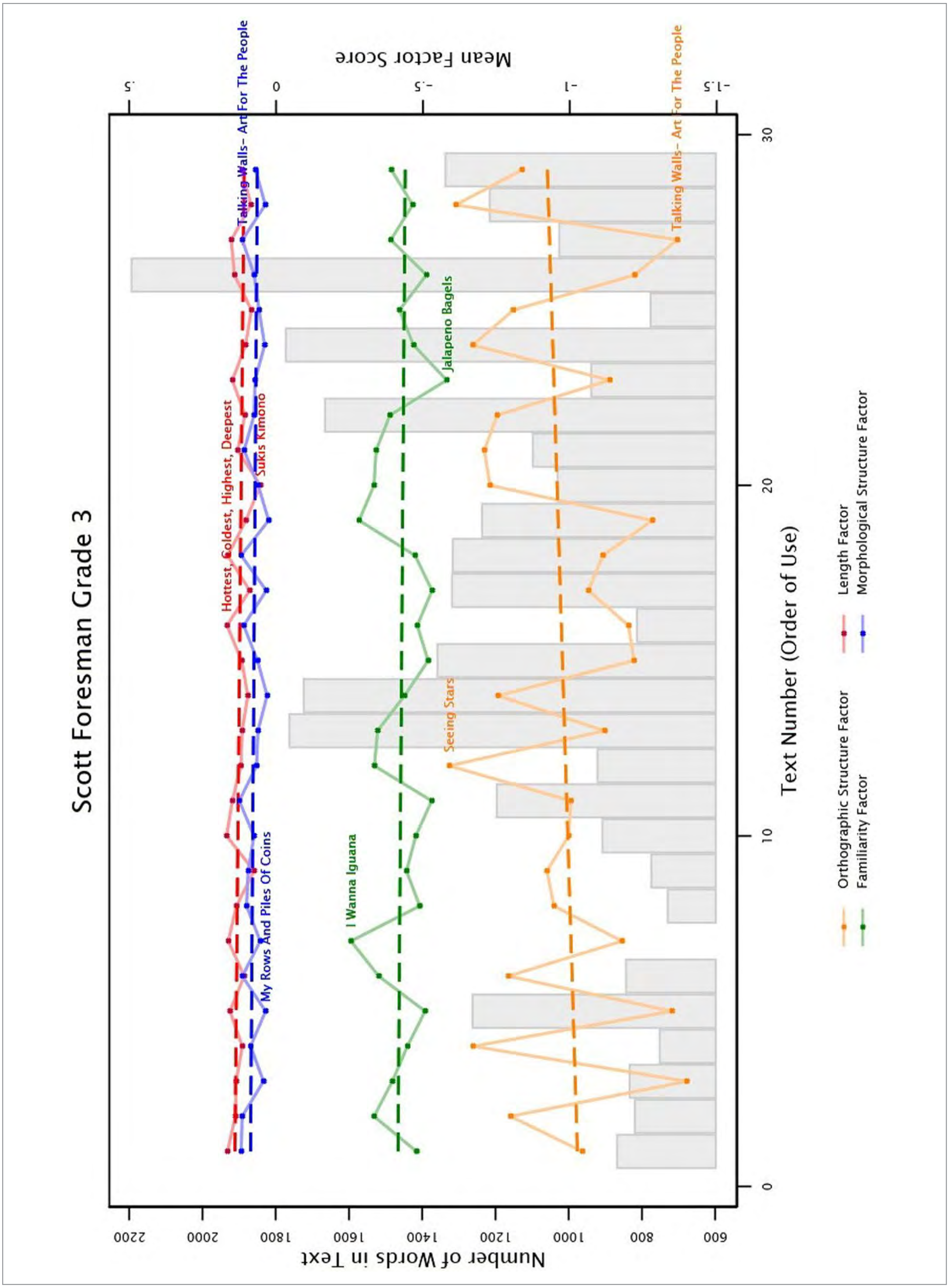
Houghton Mifflin Harcourt Grade 3
Mean Factor Score
Number of Words in Text
Text Number (Order of Use)
Two Bear Cubs
Two Bear Cubs
Roberto Clemente
The Trial Of Cardigan Jones
The Trial Of Cardigan Jones
The Treasure
The Treasure
Young Thomas Edison
Orthographic Structure Factor
Familiarity Factor
Length Factor
Morphological Structure Factor


McGraw−Hill Grade 1
Mean Factor Score
Number of Words in Text
Text Number (Order of Use)
Orthographic Structure Factor
Familiarity Factor
Length Factor
Morphological Structure Factor
Way Down Deep
I Live Here
Nat And Sam
There Are Days
A Young Nation Grov



McGraw–Hill Grade 3
Mean Factor Score
Number of Words in Text
Text Number (Order of Use)
The Real Story Of Stone Soup
Big Ideas From Nature
Brave Tavo
Out Of This World
The Winningest Woman Of The Iditarod & Other Poems
Big Ideas From Nature
Whooping Cranes
Martina The Beautiful Cockroach
Orthographic Structure Factor
Familiarity Factor
Length Factor
Morphological Structure Factor


Scott Foresman Grade 1
Mean Factor Score
Number of Words in Text
Text Number (Order of Use)
Under A Rock
My Neighborhood Then And Now
My Neighborhood Then And Now
Under A Rock
Under A Rock
Sam Come Back
They Can Help
The Ugly Duckling
Orthographic Structure Factor
Familiarity Factor
Length Factor
Morphological Structure Factor


Scott Foresman Grade 3
Mean Factor Score
Number of Words in Text
Text Number (Order of Use)
Orthographic Structure Factor
Familiarity Factor
Length Factor
Morphological Structure Factor
Talking Walls– Art For The People
Hottest, Coldest, Highest, Deepest
Sukis Kimono
My Rows And Piles Of Coins
I Wanna Iguana
Jalapeno Bagels
Seeing Stars