ASHA American Speech-Language-Hearing Association

JSLHR

Research Article

Using Multivariate Adaptive Regression Splines to Predict Lexical Characteristics' Influence on Word Learning in First Through Third Graders

Lindsey Peters-Sanders,[a] Houston Sanders,[a] Howard Goldstein,[a] and Kandethody Ramachandran[a]

[a] University of South Florida, Tampa

ARTICLE INFO

Article History:
Received March 15, 2022
Revision received June 27, 2022
Accepted October 14, 2022

Editor-in-Chief: Stephen M. Camarata
Editor: Emily Lund

https://doi.org/10.1044/2022_JSLHR-22-00165

ABSTRACT

Purpose: Identifying appropriate targets for vocabulary instruction and determining the optimal sequence for instruction continue to be a challenge. The purpose of this study is to investigate how previously studied lexical characteristics collectively influence children's word learning.
Method: A secondary data analysis was conducted using the word learning results of 350 first-, second-, and third-grade students who participated in an investigation examining the effects of a supplemental vocabulary intervention. We investigated the influence of the following lexical characteristics on the learning of 377 words: word frequency, level of concreteness, phonotactic probabilities, neighborhood density, and age of acquisition using multivariate adaptive regression splines (MARS).
Results: MARS modeled the influence lexical characteristics had on word learning and determined the relative importance of each variable for each grade-level model. Results revealed age of acquisition was the most important factor related to word learning in all grades, but contributions of other lexical characteristics and their level of importance differed across models. All respective models fit well, with root-mean-square error values ranging from 0.11 to 0.15 and generalized cross validation scores of 0.01 and 0.03.
Conclusions: Nuanced information from the MARS analysis provides insights into how lexical characteristics affect word learning differently for children in different grade levels. This information is key to understanding the vocabulary acquisition of school-aged children. The findings from this research have the potential to inform the development of a word selection framework that will organize vocabulary targets into an appropriate sequence based on relevant predictors.
Supplemental Material: https://doi.org/10.23641/asha.21899529

Vocabulary knowledge impacts social interaction, participation in classroom routines, and learning in academic content areas. Unfortunately, there is no established method of teaching vocabulary in the early primary grades (National Reading Panel, 2000). Despite the well-established role of vocabulary instruction in children's development of oral language and reading skills, little is known about what words to teach and when. It is impossible to teach all the words children will need to learn (Stahl & Nagy, 2007). To facilitate vocabulary instruction, several researchers developed word lists or guidelines to help teachers identify target words to teach (Beck et al., 2002; Biemiller, 2006; Biemiller & Boote, 2006).

Beck et al. (2002) developed the concept of word tiers. This tier system classifies words based on their utility, frequency of use, and specificity. Tier 1 words are basic, familiar words used on a daily basis; children tend to learn these words because of their frequent exposure, so they need not be directly targeted for instruction (e.g., good, pretty, big, sad). Tier 2 words are more sophisticated than Tier 1 words and are important to literacy development, because they occur in multiple contexts

Correspondence to Lindsey Peters-Sanders: lapeters@usf.edu. ***Disclosure:*** *The authors have declared that no competing financial or nonfinancial interests existed at the time of publication.*

Downloaded from: https://pubs.asha.org ERIC on 03/03/2023, Terms of Use: https://pubs.asha.org/pubs/rights_and_permissions

(e.g., complex, verify, coincide). Tier 3 words tend to be less frequent and domain-specific. Tier 3 words are not necessarily harder to learn; they just have very specific utility (e.g., environment, radius, piano). Children are not likely to encounter Tier 2 words often in everyday conversation, but they will encounter them in academic texts. Due to the lack of transparency of Tier 2 words, it would be difficult for children to derive meaning from print alone. Because of individual factors relating to a child's language experiences, it is important to focus vocabulary instruction on Tier 2 words to maximize reading comprehension. Tier 2 instruction will vary across grade levels, affording educators flexibility in instruction by targeting words in current instructional texts, for example. This has the potential to increase comprehension and generalization skills (Gray & Yang, 2015).

Biemiller agrees with Beck et al.'s (2002) principle of word tiers, but he defines them differently; there are groups of words that are learned without instruction (Tier 1), words with meanings worth teaching (Tier 2), and words with meanings to be learned later (Tier 3). Additionally, Biemiller (2010) notes the importance of distinguishing between sets of tiers for the primary grades (kindergarten to second grade) and upper elementary grades (third to sixth grade), a distinction Beck et al. do not make.

Although Biemiller (2010) notes word selection is at the teacher's discretion, he has found that words are learned in a similar sequence even when examining the learning of children from different populations (e.g., economically disadvantaged, second language learners) or when a variety of assessment methods are used (Biemiller, 2005). Biemiller and colleagues presented children in Grades 1 through 5 with a series of sentences that provide a context for a vocabulary word and then ask children to define the word (Biemiller & Boote, 2006; Biemiller & Slonim, 2001). Results were used to derive a sequence of vocabulary acquisition across children in elementary school.

Although words may be acquired sequentially, learning is likely to relate more to the size of a child's vocabulary rather than the grade they are in. The extent to which a child knows a word depends upon their developmental stage, language status, and individual experiences. For children, the vocabulary opportunities, linguistic support, and literacy-related learning experiences greatly affect their cognitive and language development and academic success (Crow & O'Leary, 2015; McLoyd, 1998). These experiences begin in infancy and help to build a foundation for language acquisition. The knowledge from their existing lexicon is used to acquire new vocabulary words, making connections between known and new concepts. Because the number of words known by children within any grade level will differ greatly, it is difficult to assign a group of words to just one grade level.

Despite recognition of the need to inform the selection of words to teach, these frameworks lack a systematic method to identify appropriate instructional targets. In a systematic review of word selection in early childhood vocabulary instruction, Hadley and Mendez (2021) found that among studies that used Beck and colleagues' tiered system for word selection, only 41% of the words were categorized as Tier 2 based on their coding criteria. They also found that several words fell within gray areas, fitting into more than one tier and that the application of the tiered system varied greatly. To this point Hadley and Mendez posed an interesting question: How do word tiers vary by age? It is unlikely that a Tier 2 word for a preschooler also would be a Tier 2 word for a child in fifth grade. A word's tier cannot be the only deciding factor used to select targets for instruction. There may be other factors that influence learning words and their meanings outside of instructional methods that could be used to organize words into an appropriate sequence for instruction.

Lexical Characteristics Influence on Word Learning

Individual lexical characteristics like frequency of use, similarity to other words, imageability, and the age at which words are typically acquired may contribute to the learnability of a word. As children build their lexicon, words acquired first tend to have higher word frequencies and come from more phonetically similar neighbors (Dollaghan, 1994; Hoover et al., 2010; Maekawa & Storkel, 2006; Storkel, 2004). Word frequency measures a word's frequency of use in a given language, in this case American English. Words with a high frequency are used more often than words with a lower frequency. Neighborhood density describes the organization of phonetically similar words in the mental lexicon. The neighborhood for a word is made up of a group of words that differ by one sound substitution, deletion, or addition. For example, the word "aid" has a neighborhood density of 21,634.85 meaning that it has over 20,000 phonetically similar neighbors (i.e., aim, paid, maid) whereas the word "appearance" had a neighborhood density of 0 meaning that it does not have any other phonetically similar neighbors.

According to Luce and Pisoni's (1998) Neighborhood Activation Model, the frequency with which words are used and the density of the neighborhood affect spoken word recognition, discrimination, and the amount of time needed to find and produce a word. Stokes (2010) found that neighborhood density and word frequency accounted for 61% of the variance in vocabulary size of 24- to 30-month-old children with a range of vocabulary sizes. She found that smaller vocabularies are made up of words high in neighborhood density and low in word frequency. The phonotactic probability of a word also can

Downloaded from: https://pubs.asha.org ERIC on 03/03/2023, Terms of Use: https://pubs.asha.org/pubs/rights_and_permissions

contribute to learnability. Phonotactic probability is the frequency of phonological segments and the sequences of phonological segments that occur in words in a given language (Vitevitch & Luce, 2004). Words made up of common sound sequences have higher phonotactic probabilities than words made up of rare sound sequences.

Most research examining the impact of phonotactic probability and neighborhood density focus on word recall and recognition, and not on learning words and their associated meanings. So, why would phonotactic probability and neighborhood density relate to children's ability to learn the meanings of sophisticated vocabulary words? Storkel et al. (2013) define three phases of word learning: triggering, configuration, and engagement. In the process of learning a word, first a child recognizes it as known or not, and then, if known, recalls what they already know about that concept. This first step activates prior knowledge and facilitates the connections made between known and novel words and concepts.

Phonotactic probability and neighborhood density play important, but different roles in each phase of word learning. It is easier to recognize new words when they are composed of low probability sound sequences (Frisch et al., 2000; Vitevitch et al., 1997). Storkel et al. (2013) found that phonotactic probability contributed to a child's ability to recognize sound sequences as acceptable and novel, and that neighborhood density contributed to the configuration of novel concepts. Storkel and Lee (2011) found that children were more precise naming objects when learning rare sound sequences than when learning common sound sequences. Neighborhood density contributed to the integration of new and existing lexical representations in adults and young children (Storkel et al., 2006; Storkel & Lee, 2011). In another study, Storkel (2009) investigated word learning in infants and found that they knew more words that were made up of low probability sound sequences than those made up of high probability sound sequences. In this same study, she also found that infants knew more words from dense neighborhoods than sparse ones. Similarly, Hoover et al. (2010) found a facilitative interaction between phonotactic probability and neighborhood density in preschoolers. Words were easier to learn when they contained common sound sequences in dense neighborhoods, and when they contained rare sound sequences in sparse neighborhoods. If vocabulary instruction in school focuses on sophisticated, Tier 2 words, understanding the impact these phonological and lexical characteristics have on word learning could provide additional insight into how these words are learned.

Research has demonstrated that young children learn words that are concrete, or highly imageable more so than words that are abstract (Hadley et al., 2021; McDonough et al., 2011; Ponari et al., 2018). Concrete words are things that can be experienced through the five senses (e.g., rock, jump). Abstract words cannot be experienced, and their meanings must be defined by other words (i.e., freedom, justice). McDonough et al. (2011) found a relation between age of acquisition and word imageability (or concreteness). Words that were more concrete were learned earlier. Similarly, Hadley et al. (2021) found a word's imageability explained 34% of variance in word learning. Ponari et al. (2018) found that young children rely heavily on emotional capacity to learn abstract words until approximately ages 8 or 9 years; then, they rely more on linguistic information. As children age, their capacity to learn more abstract words increases, utilizing earlier acquired words that may be more concrete as a foundation to build upon. Because Tier 2 words are often described as sophisticated and are not specific to a theme or discipline, one may assume that these words would be more abstract; however, that has not been fully investigated. If concreteness is highly predictive of children's ability to learn and define Tier 2 words, it would be a useful metric to use when selecting vocabulary targets.

The age at which a person learns a word is referred to as the age of acquisition (AoA). Important factors in word recognition include word frequency, length, and word similarity; however, Kuperman et al. (2012) argues that AoA is an equally important variable for two reasons. First, word frequency measures do not account for individual differences in word exposures and may underestimate the frequency for words typically used in childhood. Second, the time when words are learned influences the ease of use and recall. Words learned earlier are easier to use than those learned later. Bloom and Markson (1998) note that children gradually become better at learning words due to factors like maturing memory and attention, a heightened awareness of a word's shades of meaning, and most importantly, exposure to varied texts and literature. Research has demonstrated the influence age of acquisition has on word retrieval. Meschyan and Hernandez (2002) found that age of acquisition and word frequency affected the speed and accuracy of word retrieval, making it easier for young adults to access the phonological word form. Newman and German (2002) found words with typical stress patterns, high in frequency, and low in neighborhood density and age of acquisition (words learned at a younger age) were easier for children to name. Early elementary school is a pivotal time for word learning, especially as children shift from learning to reading to reading to learn. Understanding the influence of age of acquisition on word learning would provide insight into creating a developmental sequence of vocabulary words.

There are several limitations in this body of research deserving attention. First, most studies examined word retrieval, which differs from word learning; they do not require children to provide definitions for words but name the word after exposure to an illustration or a sentence.

Downloaded from: https://pubs.asha.org ERIC on 03/03/2023, Terms of Use: https://pubs.asha.org/pubs/rights_and_permissions

Examining different word learning experimental tasks may provide insight into the processes of lexical access that could have implications for tasks like retrieval of words and their meanings. However, examining word learning as a result of intentional instructional practices would allow us to better understand how lexical characteristics relate to students' vocabulary acquisition.

Moreover, these studies are conducted with a range of participants, from infants to adults. Research focused on school-age children is especially relevant to efforts to help young children make up for experiential deficiencies that relate to vocabulary acquisition as they are beginning to learn to read (Hart & Risley, 1995; Walker et al., 1994). Additionally, the stimuli used in these studies are not typically the words teachers would use for instruction. McDonough et al. (2011) used nouns and verbs that Beck, Biemiller, and others would consider basic, foundational words easily learned without much explicit instruction. Hoover et al. (2010) and Storkel et al. (2006) used pseudowords created to control for phonotactic probability and neighborhood density. Consequently, results from these studies may overgeneralize the effects of neighborhood density and phonotactic probability on word learning when compared with more authentic word learning tasks.

Although these findings demonstrate that lexical characteristics influence word learning, we do not know the relative contributions that lexical characteristics have on children's ability to learn meanings of sophisticated vocabulary words. To facilitate vocabulary instruction and to better understand the developmental sequence of vocabulary acquisition, we must examine the relation between lexical characteristics and learning of words and their meanings. This study focuses on the first step in this line of research by addressing the following question:

> To what extent do lexical characteristics influence learning of words and their meanings in first, second, and third grade students who received a supplemental vocabulary intervention?

It is hypothesized that word frequency, age of acquisition, level of concreteness, neighborhood density, and phonotactic probability will influence children's learning of word meanings taught in a supplemental vocabulary program. If children are exposed to more frequently used words and their meanings, then words with higher frequencies will yield greater learning gains. If age of acquisition is related to word learning, then words with an age of acquisition younger than, or matching that of, a student's current age will have a higher rate of learning success than those words with older age of acquisition ratings. One would expect children to have more opportunities for exposure to words acquired at an earlier age than those words learned later in childhood. If the level of concreteness can facilitate word learning, then it will be easier for children to learn words and their meanings when words are more concrete compared with more abstract terms. If words are from sparser neighborhoods, then it will be easier for children to recall words and their meanings because there will be fewer similar words competing for access in the lexicon. If words are made up of more common sounds and sound combinations, it will be easier for children to recognize the word and recall its meaning.

If relations exist between lexical characteristics and learning words and their meanings, it will be possible to determine which characteristics best predict word learning. By identifying the lexical characteristics that best predict word learning, we can better organize academic vocabulary targets for instruction based on relevant predictors that would follow a scope and sequence of acquisition to optimize the developmental process of word learning.

Method

Participants

A secondary data analysis was conducted using word learning outcomes of 377 words from 350 first-, second-, and third-grade students who took part in a longitudinal study investigating the effects of a supplemental intervention that taught academic vocabulary words (Goldstein et al., 2017). Students attended two elementary schools that served primarily low-income families. Over 90% of students qualified for free or reduced lunch. The sample was made up of 54% males and 46% females with less than 0.5% were identified with limited English proficiency. The ethnic breakdown of participants was 75% African American, 15% Caucasian, 5% Hispanic, and 5% more than one race (Goldstein et al., 2017).

Supplemental Vocabulary Intervention

The Independent Lexical Instruction and Development (ILIAD) supplementary Tier 2 vocabulary program (Goldstein et al., 2017) was a longitudinal study spanning 3 years. The intervention occurred 4 days a week and was set up as a listening center with the instructional script shown in Supplemental Material S1. Each session lasted 20 min. There was a total of 18 weeks of instruction per school year broken up into four 4- to 5-week waves of instruction. Wireless headphones and a receiver were used allowing students to remain at their seats to complete the intervention sessions. Research assistants ran the listening centers while teachers monitored students during the activity. Each week, one decodable book from the core curriculum was used for all four intervention sessions. Students would follow along with a prerecorded read aloud. This

Downloaded from: https://pubs.asha.org ERIC on 03/03/2023, Terms of Use: https://pubs.asha.org/pubs/rights_and_permissions

allowed for decoding practice in first grade and fluency practice in second and third grades. After the read aloud, children received vocabulary instruction. These lessons were scripted and included opportunities to respond to the narrator and provided multiple opportunities for students to interact with the target words using verbal responses and perform actions or gestures. At the end of each lesson, there was a brief review of the targeted vocabulary words. Students had worksheets with pictures that corresponded to the different target words. Students would complete activities on the worksheets during these lessons to reinforce learning. By third-grade, students were also writing the target words and completing cloze sentences. A sample of the vocabulary lesson can be found in Supplemental Material S1.

Across each grade level, a total of six Tier 2 words—two nouns, two verbs, and two adjectives—were taught each week. In first and second grades, an additional Tier 1 anchor word from the Open Court series was included to connect the lesson to the classroom curriculum. For example, for the story *Cinderella*, the anchor word was "castle." Words taught in one grade level were not taught again in subsequent grades. Students heard the targeted vocabulary words 11–12 times and the corresponding definitions 6–7 times each session, including the review. Tier 2 word selection for the intervention followed the criteria set forth by Beck et al. (2002). The Academic Word List (Coxhead, 2000) included a list of words derived from a variety of college-level texts that could be categorized as Tier 2 words. Each of the words chosen for instruction had to be illustrated, defined, and fit into existing stories. Because of these constraints, researchers were running out of words for third grade. The Living Word List (Dale & O'Rourke, 1976) was used to supplement word selection. A total of 377 vocabulary words were included in our analysis.

The learning outcomes were derived from the decontextualized definition subtest of a researcher-made measure that was administered by research assistants in person every 4–5 weeks. Students had to provide the meaning of the target word without additional contextual support (Goldstein et al., 2017). Students were prompted with the phrase, "Tell me everything you know about ___." If they responded with one attribute of the word, they were prompted a second time with "Tell me something else about ___." A correct response included a definition, a synonym, or a brief description of the word. Internal consistency and interrater reliability were high across grade levels (Cronbach's alpha ranged from .82 to .94; kappa ranged from .83 to 1; Goldstein et al., 2017). Originally, this measure was scored on a 3-point scale, 0 for *not learned*, 1 for *partial knowledge*, and 2 for *full knowledge*. Because we were interested in the number of children who learned each word, we combined partial and full knowledge into one category: learned. The revised scale for this secondary analysis was 0 for not learned and 1 for learned.

Coding of Lexical Characteristics of Words

A total of 377 target vocabulary words were characterized for analysis based on available database estimates of their individual word frequency, neighborhood density, level of concreteness, age of acquisition, and phonological phonotactic probability (see Table 1 for mean, standard deviations, and ranges for each lexical characteristic). In some instances, the targeted vocabulary word was a derivation and not included in the databases. When this occurred, the values for the base or root word were used instead. On average, 26% of students learned words in first grade, 38% of students learned words in second grade, and 22% of students learned words in third grade. Goldstein et al. (2017) noted an overall decrement in learning by third grade. One reason for this could be the difficulty researchers faced when selecting words for instruction in third grade; they were running out of words by third grade and had to consult additional word lists to supplement.

Word Frequency

There are several measures of American English word frequency. The $\text{SUBTLEX}_{\text{US}}$ word frequency measures are based on American English subtitles from movies and television shows and include a corpus of 51 million words. This corpus is available and easily accessed online and provides frequencies for spoken language that approximates everyday language use (Brysbaert & New, 2009). This is a departure from other frequency measures that rely on language found in texts, which may not yield frequency values that best represent the sample of words used for this analysis.

The Kučera and Francis corpus compiled word frequencies for 1.014 million words. These frequency counts are based on 500 samples of text including editorials, essays, technical writings, and various types of fiction printed in 1961 (Francis & Kučera, 1982). Thus, the words selected may better represent the lexicon of an adult than that of a child (Gierut & Dale, 2007). Although the Francis and Kučera metrics have been considered the norm for quite some time, they are dated and do not estimate raw frequency well due to its relatively small corpus size (Balota et al., 2007). Brysbaert and New (2009) note other frequency norms that are not readily available or released due to copyright protection (i.e., Zeno, MetaMetrics, and Celex), and as such were not considered for use in this study. Word frequency values were obtained from the Irvine Phonotactic Online Dictionary version 2.0 (Vaden et al., 2009), which reports frequency measures from the SUBTLEXus corpus. Word frequency ratings for our data ranged from 0.27 to 509.37 in first grade and 0.02 to 801.82 in second grade. Third grade had a much more restricted range of frequencies, from 0.08 to 35.65.

Downloaded from: https://pubs.asha.org ERIC on 03/03/2023, Terms of Use: https://pubs.asha.org/pubs/rights_and_permissions

Table 1. Descriptive statistics for independent lexical instruction and development model variables.

Variable	*M*	*SD*	*Mdn*	Minimum	Maximum	Skew
First grade (*n* = 143)						
Word learning	26%	29%	14%	1%	99%	—
Age of acquisition	8.80	2.17	9.06	3.25	13.61	−0.46
Neighborhood density	1845.32	8474.12	6.49	0	69210.62	6.23
Level of concreteness	2.96	0.97	2.76	1.50	5	0.64
Phonotactic probability	0.22	0.12	0.21	0.03	0.52	0.57
Word frequency	19.29	57.94	6.90	0.27	509.37	6.67
Second grade (*n* = 126)						
Word learning	38%	28%	27%	3%	97%	—
Age of acquisition	8.63	2.29	8.63	3	13.41	−.23
Neighborhood density	1106.11	5560.46	8.53	0	45721.92	6.92
Level of concreteness	2.89	1.00	2.63	1.46	4.97	0.72
Phonotactic probability	0.24	0.15	0.21	0.02	0.66	0.76
Word frequency	31.37	105.91	7.63	0.02	801.82	6.07
Third grade (*n* = 108)						
Word learning	22%	15%	18%	3%	74%	—
Age of acquisition	10.30	1.41	10.25	6.75	14.5	0.15
Neighborhood density	771.50	6666.04	1.41	0	69210.62	10.02
Level of concreteness	2.39	0.62	2.29	1.43	4.15	0.81
Phonotactic probability	0.24	0.14	0.21	0.03	0.72	1.37
Word frequency	4.82	6.39	2.46	0.08	35.65	2.57

Note. *n* = number of words; skew is not reported for word learning (as noted by the em dashes).

Neighborhood Density

Neighborhood density describes the organization of phonetically similar words in the mental lexicon. Words in a neighborhood differ by one sound substitution, deletion, or addition. Depending on the number of possible neighbors, words can be identified as either high or low density. A high density word has many neighbors, whereas a low density word has few phonetically similar words. Neighborhood density counts were also obtained from the Irvine Phonotactic Online Dictionary Version 2.0 (Vaden et al., 2009) and were calculated for words in the SUBTLEXus corpus. Neighborhood densities for our data ranged from 0 (words with no phonologically similar neighbors) up to 69,210.62 neighbors.

Level of Concreteness

Concreteness level ratings were derived from a database of 37,058 English words developed by Brysbaert et al. (2014). This database contains a larger sample size and ratings consistent with norms from ratings gathered in the past by Spreen and Schulz (1966) and Paivio et al. (1968). People rated words on their level of concreteness using a 5-point scale with 1 being *abstract* and 5 being *concrete*. The words used in our analyses had concreteness ratings ranging from 1.5 to 5 in first grade, 1.46 to 4.97 in second grade, and 1.43 to 4.15 in third grade.

Age of Acquisition

Age of acquisition norms were obtained from a corpus created by Kuperman et al. (2012). They compiled age of acquisition ratings for 30,000 words selected from the $\mathrm{SUBTLEX_{US}}$ corpus. These ratings were obtained by asking 1,960 people to rate the age at which they understood a word. While this seems very subjective, other researchers have found that these ratings are a reliable measure of age of acquisition (Gilhooly & Gilhooly, 1980; Gilhooly & Logie, 1980). Age of acquisition ratings for our data set ranged from approximately 4 to 13.5 years old in first and second grades and from 7 to 14.5 years old in third grade.

Phonotactic Probability

Phonotactic probability was calculated using a web-based interface developed by Vitevitch and Luce (2004). Phonotactic probability refers to the frequency with which phonological segments and sequences of phonological segments occur in words in a given language (Vitevitch & Luce, 2004). For first grade, probabilities for the words in our data set ranged from .03 to .52, in second grade, they ranged from .02 to .66, and .03 to .72 in third grade.

Lexical characteristics were not controlled for in this analysis. The individual lexical characteristics for the sophisticated vocabulary words do change across grade levels indicating an increase in level of difficulty. For instance, mean age of acquisition ratings move from 8.80 and 8.63 years of age in first and second grades to 10.30 in third grade. Concreteness ratings move from more concrete words in first grade to more abstract word by third grade (mean of 2.96–2.39). While small, the incremental change across grades may impact learning. The average word frequency in third grade was 4.82 that is significantly lower than first and second grades, where average frequencies were 19.29 and 31.37, respectively.

Downloaded from: https://pubs.asha.org ERIC on 03/03/2023, Terms of Use: https://pubs.asha.org/pubs/rights_and_permissions

Each database was either available for download or was converted to Excel files to streamline data collection. Using the search and retrieval functions in Excel, the various databases were searched for all 377 target words. A secondary matching function and random searches were done to ensure correct words and values were reported from each database.

Data Analysis

To examine the joint and unique predictive variance of lexical characteristics on vocabulary learning, multivariate adaptive regression splines (MARS) was used to analyze the data. MARS has been used to model students' achievement and development (Kilic Depren, 2018; Kolyshkina et al., 2013; Martís et al., 2015). Although use of this technique is slowly emerging in educational research, MARS is a robust, adaptable method for regression that has a strong balance of precision while being interpretable and relatively simple to implement. MARS uses an iterative process that does not rely on assumptions about the data. It handles correlated variables relatively well, works well with a large number of predictor variables, and can work with both linear and nonlinear data.

MARS is an additive model that adaptively uses hinges to balance precision with simplicity to create robust models. Rather than fitting a regression line with one general trend dictated by the entire domain of the data, hinges are inserted to allow a change in the direction of the regression. The possible hinges, direction of the trend line, and corresponding coefficient make up the basis functions used to build the model. This provides a nuanced and dynamic picture of how the trends in word learning change based on the relative influence of lexical characteristics.

The RStudio environment (RStudio Team, 2020) was used to create the models in R using the earth (enhanced adaptive regression through hinges) package (Milborrow & Milborrow, 2007). Several models were created and compared (Sanders, 2021). Each model consists of word learning data for the target words and their respective lexical characteristics across grade levels (Sanders et al., 2021).

First, word-level data were compiled for all vocabulary words and included the grade it was taught and within what unit, the percentage of students who learned each word, and the values for each lexical characteristic. The mean, standard deviation, and range for each of these variables are listed in Table 1. An exploratory data analysis was completed for all variables to check multicollinearity, nonlinear relationships between variables, rank correlations, piecewise normality, and multivariate normality. Preliminary data analyses were performed. Results indicate the data were not normally distributed, contained multicollinearity and correlation, and were heteroskedastic.

Next, MARS was used to model the data. MARS uses forward model creation and backward deletion. Initially, an overfitted model is created and backward creates a simpler model. Number of subsets is how often each variable appears during this process. Variables are ranked by how often they appear. Percent of relative impact is calculated using general cross validation (GCV) and residual sum of squares (RSS). The impact of included variables is relative to the most important variable whose impact is always 100%.

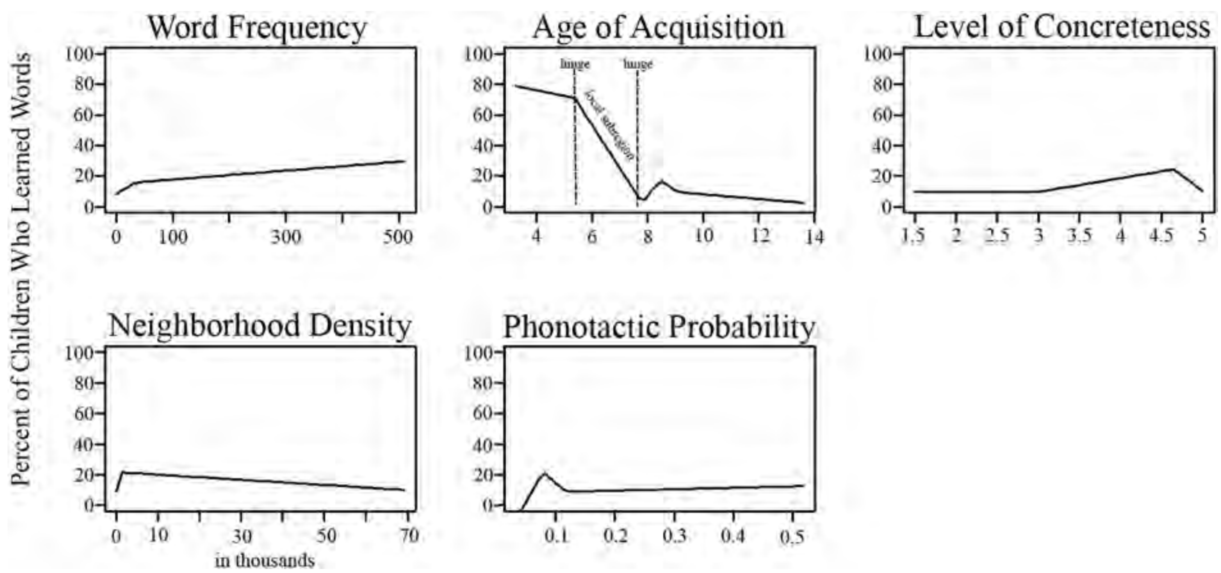
Finally, goodness of fit metrics were calculated to examine how well the resulting models fit the data. General cross validation (GCV) was calculated while building the MARS models to prevent overfitting or underfitting (Friedman et al., 2001). GCV is used for level of importance as well as goodness of fit. For model fit, a lower GCV indicates a better model. Finally, models were compared for accuracy to find the best fit using R^2, root-mean-square error (RMSE), mean square error (MSE), and mean absolute error (MAE). For these measurements, R^2 explains how closely the data are modeled to the regression with it becoming stronger as it converges to 1. However, R^2 is not designed to work with nonlinear data so it is not the most reliable metric for model fit but was included because this metric is a commonly reported statistic (Willett & Singer, 1988). MAE, MSE, and RMSE error are metrics used to determine how much a model differs from the data and are closely related. Indices closer to zero show a better model fit for the data.

Results

For each grade-level model, the following results are presented: identification of predictor variables and those variables that were trimmed from the model, variable importance and selection criteria, the final model including basis functions that are made up of hinge location, direction, and the associated coefficient for each local subregion. Hinges split the data into local subregions. Direction indicates whether the subregion is to the right or left of the associated hinge. The trends for each local subregion can vary from hinge to hinge allowing for a dynamic representation of relations among word learning and lexical characteristics. The graphs associated with each model are also presented. Each graph represents the marginal effect of each variable on word learning, whereas all others were held constant. On the graphs, the scale for each x-axis differs based upon the lexical characteristic values. The variables are organized by the order in which they were entered into the model, not by level of importance. To orient readers to the graphs, we have identified the location of two hinges and the resulting local subregion in Figure 1.

Downloaded from: https://pubs.asha.org ERIC on 03/03/2023, Terms of Use: https://pubs.asha.org/pubs/rights_and_permissions

Figure 1. First grade multivariate adaptive regression splines model. Variables are not ordered by importance. Scale for each *x*-axis differs based on lexical characteristic values. On the age of acquisition graph, two hinges and the associated local subregion are labeled.



First Grade

For the first-grade model, word learning data for 143 words taught in first grade and the lexical characteristics describing those words were entered into the model. The resulting model identified age of acquisition, level of concreteness, word frequency, neighborhood density, and phonotactic probability as relevant predictors of word learning. The relative level of importance of each variable is listed in Table 2. The most important variable was age of acquisition, with level of concreteness 22%–26% as important, neighborhood density 16%–20% as important, word frequency 8% –12% as important, and phonotactic probability 6%–8% as important. The final model included seven basis functions listed in Table 3 and depicted in Figure 1. The figure displays a regression for each lexical characteristic included in the model. Hinges are located in places where the regression line changes direction. For example, age of acquisition remained steady until the hinge at 5.37 years old where the percentage of children who learned words decreased rapidly until the age of 7.81 years as noted on the graph. There was a slight jump in learning between age of acquisition ratings of 7.81 and 8.45 years old, and then slowly decreased as age of acquisition goes up. The percentage of children who learned words remained steady for words with concreteness ratings from 1.5 to 3, and then slowly increased after the hinge at 3. Words that were more concrete were learned by more first graders. Trends for neighborhood density, word frequency, and phonotactic probability remained neutral.

Second Grade

For the second-grade model, word learning data for 126 words taught in second grade and the lexical characteristics describing those words were entered into the model. In Table 2, the most important variable was age of acquisition followed by level of concreteness, word frequency,

Table 2. Importance of explanatory variables and selection criteria for multivariate adaptive regression splines models.

		Percent of relative impact	
Variable	**nsubsets**	**GCV**	**RSS**
First grade			
Age of acquisition	7	100	100
Level of concreteness	5	22.1	26
Neighborhood density	4	16	20.3
Word frequency	2	8.1	12.1
Phonotactic probability	1	5.5	8.3
Second grade			
Age of acquisition	9	100	100
Level of concreteness	8	35.8	44.8
Word frequency	7	26.8	37.2
Neighborhood density	6	16.3	29.6
Phonotactic probability	5	16.9	27.8
Third grade			
Age of acquisition	7	100	100
Neighborhood density	5	62.4	68.8
Word frequency	4	55.5	61.8
Level of concreteness	4	53	59.3

Note. nsubsets = number of subsets, percent of relative impact calculated with general cross validation (GCV), or residual sum of squares (RSS).

Downloaded from: https://pubs.asha.org ERIC on 03/03/2023, Terms of Use: https://pubs.asha.org/pubs/rights_and_permissions

Table 3. Multivariate adaptive regression splines models for word learning in first, second, and third grade.

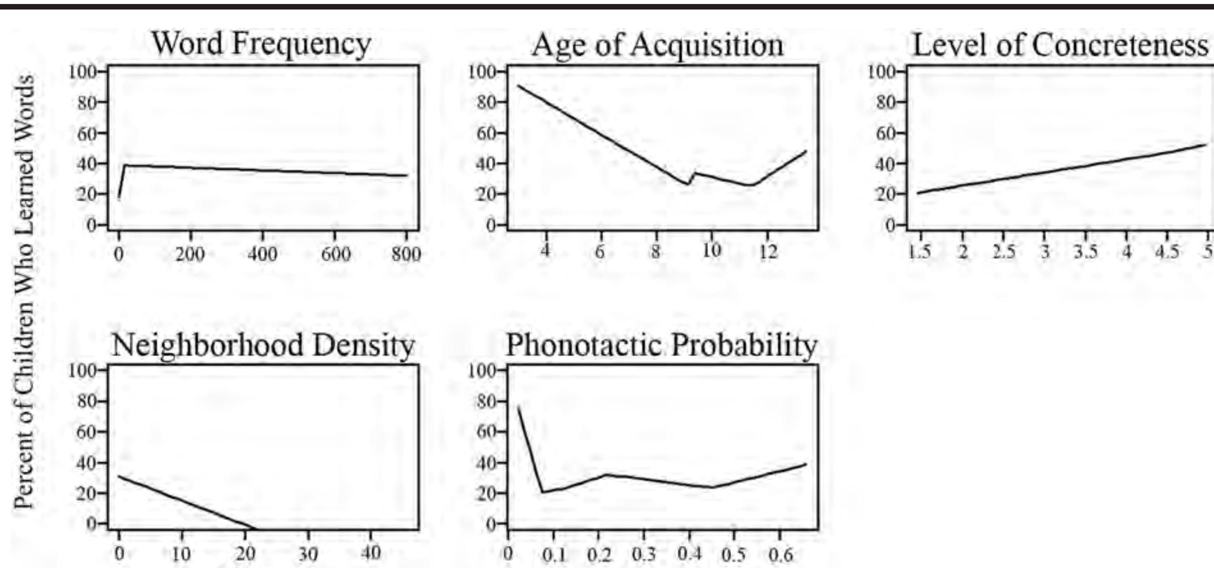
Predictor Variable	Hinge location	Coefficient	Direction
First grade			
(intercept)		0.91	
Age of acquisition	5.37	−0.27	Right
Age of acquisition	7.81	0.42	Right
Age of acquisition	8.45	−0.17	Right
Level of concreteness	3.00	0.07	Right
Neighborhood density	126.04	−0.001	Left
Word frequency	32.22	−0.003	Left
Phonotactic probability	0.08	−3.05	Left
Second grade			
(intercept)		−3.62	
Age of acquisition	9.35	0.09	Left
Age of acquisition	11.44	0.10	Right
Level of concreteness	4.44	−0.09	Left
Word frequency	12.35	−0.02	Left
Neighborhood density		−0.00001	Right
Phonotactic probability	0.07	11.60	Right
Phonotactic probability	0.22	−1.41	Right
Phonotactic probability	0.45	10.62	Left
Phonotactic probability	0.45	−9.48	Right
Third grade			
(intercept)		0.12	
Age of acquisition	9.67	−0.05	Right
Neighborhood density		0.001	Right
Neighborhood density	82.79	−0.001	Right
Word frequency	11.84	0.007	Left
Word frequency	11.84	0.01	Right
Level of concreteness	2	0.31	Right
Level of concreteness	2.3	−0.30	Right

neighborhood density, and finally phonotactic probability. The final model included nine basis functions listed in Table 3 and depicted in Figure 2. Based on age of acquisition, the percentage of children who learned words decreased slowly until age 9.35 years where learning seemed to remain neutral until 11.44 years old. Learning began to increase for words with age of acquisition ratings older than 11.44 years. Learning steadily increased as words became more concrete (values closer to 5). Word frequency remained neutral; learning did not seem to vary for words as frequency rates increased. The percentage of children who learned words steadily declined as neighborhood density values increased, that is why there was no hinge present in the figure. Words in denser neighborhoods were more difficult for children to learn compared with words in sparser neighborhoods. Phonotactic probability had a slightly varied impact on the percent of children who learned words; learning drops rapidly as phonotactic probability values increased to .07 and then remained mostly neutral with minimal increases and decreases between hinges at phonotactic probability values .22 and .45.

Third Grade

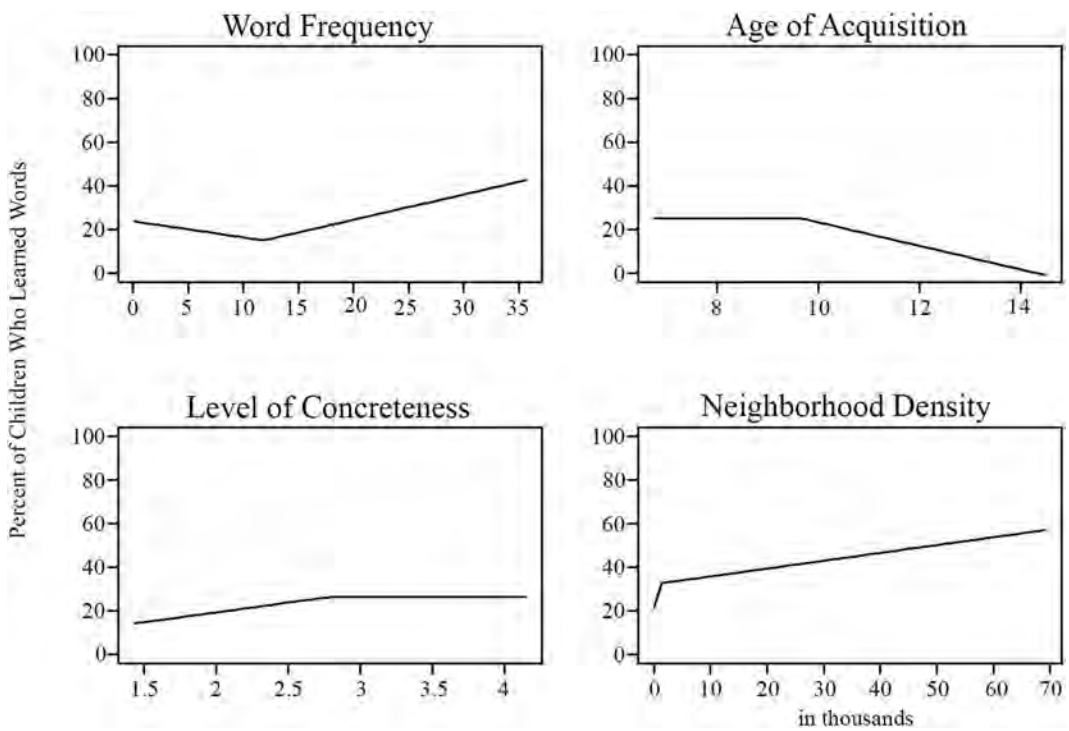
For the third-grade model, word learning data for 108 words taught in third grade and the lexical characteristics describing those words were entered into the model. MARS identified age of acquisition as the most important variable, followed by neighborhood density, word frequency, and finally level of concreteness, listed in Table 2. Phonotactic probability was trimmed from the model. The final model included seven basis functions listed in Table 3 and depicted in Figure 3. The percentage of children who learned words remained constant until an AoA rating of 9.67 and then slowly decreased as AoA increased. Learning steadily increased as neighborhood density values increase,

Figure 2. Second grade multivariate adaptive regression splines model. Variables are not ordered by importance. Scale for each *x*-axis differs based on lexical characteristic values.



Downloaded from: https://pubs.asha.org ERIC on 03/03/2023, Terms of Use: https://pubs.asha.org/pubs/rights_and_permissions

Figure 3. Third grade multivariate adaptive regression splines model. Variables are not ordered by importance. Scale for each *x*-axis differs based on lexical characteristic values.



as words in denser neighborhoods were easier for children to learn compared with words in sparser neighborhoods. As word frequency increased, so did the percentage of children who learned words. Level of concreteness had a small impact on third-grade students' word learning. Words that were more abstract were slightly more difficult for children to learn, but once words reached a concreteness rating of 2.3, learning remained neutral.

Variable Importance

Variable importance for each grade-level model can be found in Table 4. Age of acquisition was the most important lexical characteristic related to word learning for all grade levels. Variable importance for the first- and second-grade models was identical. Level of concreteness was the second most important variable related to word learning, followed by neighborhood density, word frequency, and finally phonotactic probability. Interestingly, the variable importance for the third-grade model differed from the first- and second-grade models. Neighborhood density was the second most important variable, followed by word frequency and level of concreteness. Phonotactic probability was not included in the third-grade model.

Model Fit

The outcome for each MARS model was critiqued using goodness of fit measures found in Table 5. Results from the error measures indicated that each of the grade level models had minimal error, with values for MAE, MSE, RMSE, and GCV close to zero. While the second-grade model fit well, error values were higher compared with the first- and third-grade models. The coefficient of determination (R^2) indicated the first- and second-grade level models explained more of the variance compared

Table 4. Variable importance for first, second, and third grade.

Level of importance	First grade	Second grade	Third grade
1	Age of acquisition	Age of acquisition	Age of acquisition
2	Level of concreteness	Level of concreteness	Neighborhood density
3	Neighborhood density	Neighborhood density	Word frequency
4	Word frequency	Word frequency	Level of concreteness
5	Phonotactic probability	Phonotactic probability	

Downloaded from: https://pubs.asha.org ERIC on 03/03/2023, Terms of Use: https://pubs.asha.org/pubs/rights_and_permissions

Table 5. Goodness of fit results for first, second, and third grade multivariate adaptive regression splines models.

Error metric	First grade	Second grade	Third grade
R^2	0.84	0.72	0.42
MAE	0.09	0.12	0.08
MSE	0.01	0.02	0.01
RMSE	0.12	0.15	0.11
GCV	0.01	0.03	0.01

Note. MAE = mean absolute error; MSE = mean standard error; RMSE = root-mean-square error; GCV = general cross validation.

with the third-grade model, R^2(1st) = .84 and R^2(2nd) = .72, compared with R^2(3rd) = .42.

Discussion

A secondary data analysis of an investigation examining the effects of a supplemental vocabulary intervention were conducted using MARS to identify the relations between lexical characteristics and the word learning outcomes (i.e., the ability to recall the meanings of target vocabulary) from first-, second-, and third-grade students. The lexical characteristics examined were age of acquisition, level of concreteness, word frequency, phonotactic probability, and neighborhood density. Model results indicated that age of acquisition was the most important characteristic related to word learning across grade levels. The second most important characteristic related to word learning was level of concreteness in first and second grades, and neighborhood density in third grade.

Lexical Characteristics

Age of Acquisition

The MARS analysis revealed significant relations between age of acquisition and students' vocabulary learning in the first-, second-, and third-grade models. We found that words with a younger age of acquisition rating were easier for children to learn their meanings than words with older age of acquisition ratings, which is what we predicted. Our findings support the results of a lexical access studies by Meschyan and Hernandez (2002) and Newman and German (2002) who found it was easier to name words with lower age of acquisition. Although this seems rather intuitive, and somewhat circular, this is an interesting factor to discuss. The level of importance attributed to age of acquisition is surprising considering the nature in which these ratings were obtained. Adults were asked to recall the age at which they learned a word. *Learned* was defined as understanding the word if others used it, but that they did not necessarily use it themselves. This can be a difficult task, especially when trying to recall learning at a very young age. Yet researchers have examined the validity of this and found that adult ratings of age of acquisition are valid (Gilhooly & Gilhooly, 1980; Gilhooly & Logie, 1980). Findings from this study reinforce age of acquisition as a reliable metric.

In second grade, there was an increase in learning at the hinge at AoA rating 11.44. This increase is an interesting artifact. It could be that the word(s) had other contributing factors, like a higher level of concreteness, that lead to increased learning. The definition and/or the contexts used for instruction may have also contributed to the increase in learning.

Now that we know age of acquisition strongly relates to children's sophisticated vocabulary learning from a range of grade levels, additional analyses and studies are warranted to discover the range of AoA ratings that lead to optimized learning for each grade level. Because of the way MARS models data, we have detailed information about how age of acquisition impacts learning using hinges. When designing future studies, the hinge data could help when selecting words for instruction by pinpointing the exact age range most appropriate for each grade level. This selection of words would be more precise than using general linear trends. It may be that teachers should focus instruction on words and their meanings acquired later (within reason given the age of students) because they are more difficult for children to learn than words that are acquired at an earlier age.

Level of Concreteness

MARS modeled level of concreteness as the second most important lexical characteristic related to learning words and their meanings in the first- and second grade–level models. Our results indicate, that for children in first and second grades, words that were more concrete, or high in imageability, were easier to define than words that were more abstract, meaning they were more difficult to explain and picture. The third-grade model selected concreteness as the least important variable; it did not seem to significantly impact word learning. We hypothesized that it would be easier for all children to define words that were more concrete, so this finding was an interesting deviation. Descriptive statistics were examined to determine if the average concreteness level of the words taught in first, second, and third grades differed significantly. If there were differences, it could explain the differences in the model's selection of important variables. The average concreteness levels did not differ greatly across grade levels; however, words taught in third grade had a lower maximum value (4.15 compared with 5 and 4.97) and varied less (SD = 0.62 compared with 0.97 and 1). While these small differences could impact children's ability to acquire words and their meanings, there may be something innately different about the age of children, how they

Downloaded from: https://pubs.asha.org ERIC on 03/03/2023, Terms of Use: https://pubs.asha.org/pubs/rights_and_permissions

acquire the meanings of new vocabulary terms, and what lexical characteristics influence learning the most.

The hinge data provided insightful information about the underlying process of word learning regarding abstract and concrete concepts taught in first and second grades. Our findings are supported by prior research that found imageability predicted preschoolers' word learning (Hadley et al., 2021). Also, more imageable words were learned earlier and more easily than words that were less imageable (McDonough et al., 2011). Again, this finding is rather intuitive. Words that are more concrete have specific meanings, whereas words that are more abstract often have nuanced meanings that depend on context. Children can acquire more abstract terms, but if they have no referent to associate the word with, it can be difficult to retain the word's meaning. It could be that as children age, their life experiences make them well suited to understand and describe more abstract concepts. This is supported by Ponari et al. (2018) who found that young children rely heavily on emotional valence to learn words, but that as they approach 8 or 9 years of age (and enter third grade), they rely more heavily on linguistic information to learn abstract words. This shift, coupled with the restricted range of ratings in our sample, could explain why concreteness did not impact word learning in third grade. Further research is warranted to explore this phenomenon to better understand word learning across a larger group of children to determine if there are underlying processes that facilitate the acquisition of abstract terms. This characteristic coupled with age of acquisition could facilitate the creation of a developmental sequence for vocabulary instruction.

Word Frequency

Word frequency was included as an important variable in each of the grade models and was 8%–12% as important for first grade word learning as age of acquisition, was 17% –28% as important for second grade word learning as age of acquisition, and was 56%–62% as important for third grade word learning as age of acquisition. For first and second grades, it appears that as word frequency values increased, learning word meanings did not change dramatically. For third grade, meanings of words that were more frequent were easier to learn, corroborating our original hypothesis and the findings of other researchers who found that words that occur more frequently were easier for children to learn (Dollaghan, 1994; Hoover et al., 2010; Maekawa & Storkel, 2006; Storkel, 2004).

The words in this study did not include words with very high measures of word frequency, so our findings must be interpreted carefully due to the restricted range of frequencies. Word frequency values ranged from 0.02 to 801 and were heavily skewed for all grade levels. For third grade in particular, word frequencies varied less and had a more restricted range (SD = 6.39, range: 0.08–35.65) compared with first (SD = 57.94, range: 0.27–509.37) and second (SD = 105.91, range: 0.02–801.82) grades. While the words chosen may not seem to have a lower frequency among adults, they may have infrequent use by younger children. Further analyses should investigate word frequency norms for children by examining childhood literature or television shows and movies made for children. Either of these methods would mirror popular adult word frequency norms derived from print or television and movies (e.g., Brysbaert & New, 2009; Francis & Kučera, 1982). If differences existed between the frequency norms of children and adults, it would allow for a more robust measure used to examine the relation between frequency and children's vocabulary learning.

Phonotactic Probability and Neighborhood Density

MARS included neighborhood density and phonotactic probability in the first- and second-grade models, and only neighborhood density was included in the third-grade model. We hypothesized that it would be easier for children to recall meanings of words from sparser neighborhoods. For the first- and second-grade models, this was true. Learning slowly declined for words as neighborhood density increased (words were from denser neighborhoods). However, this was not the case for third grade; in fact, it was the opposite.

In third grade, neighborhood density was the second most important variable related to word learning; it was 62%–69% as important to word learning as age of acquisition. There was a sharp increase in learning until the hinge at 82.79 and then a steady increase in learning as neighborhood density measures increased. Word learning increased as neighborhood density also increased. It was easier for children to learn the meanings of words from denser neighborhoods than those from sparse neighborhoods. Our findings support that of Storkel and Lee (2011) who found that initially young children did better learning sparse sound sequences right after training, but that learning of denser sound sequences improved over time. It could be that as children get older, lexical information plays a more important role in learning the meanings of words than it does for younger children who rely on other factors to integrate new concepts (Ponari et al., 2018).

Most studies investigating these two variables relied on word recall and recognition tasks rather than on definition tasks. During the ILIAD study, researchers also measured children's ability to label target vocabulary words when presented with a stimulus picture. The percentage of children demonstrating word recall was highly correlated with the percentage of children who could correctly define

Downloaded from: https://pubs.asha.org ERIC on 03/03/2023, Terms of Use: https://pubs.asha.org/pubs/rights_and_permissions

target words across grade levels (first grade ρ = 0.92, second grade ρ = 0.83, third grade ρ = 0.85). Storkel (2013) found that phonotactic probability influenced children's abilities to recall and recognize words, whereas neighborhood density facilitated the integration of new concepts. Similarly, Storkel et al. (2006) found that phonotactic probability contributed to adult word learning while neighborhood density facilitated the integration of new and existing lexical representations. By third grade, neighborhood density remained in the model while phonotactic probability was removed. Many studies investigating the influence of neighborhood density and phonotactic probability measured lexical access and not necessarily the ability to also recall meanings for words (Hoover et al., 2010; Storkel et al., 2006; Storkel & Lee, 2011, Storkel et al., 2013). However, the process of word retrieval and the ability to recall definitions warrants further investigation, especially if we can disentangle the differential effects each characteristic has on learning.

Phonotactic probability is directly affected by word length. It is calculated using the sum of log values, which is equivalent to the log of the values multiplied. When multiplied together, values in this range will always decrease. This leads to a decrease in phonotactic probability as word length increases. Word length was not a factor controlled for in this study. Phonotactic probabilities for the words in our analysis ranged from 0.01 to 0.08. These small probabilities were not significantly related to word learning nor to neighborhood density. This restricted range of probabilities could explain why phonotactic probability was removed from the third-grade model.

Neighborhood density is negatively correlated to word length. The density increases for shorter words that have more similar neighbors, and decreases in density as word length increases, where longer words have fewer similar neighbors. Because our words varied in length, we had a large range of density measures, from 0 to 69,210.62. About half of the words (53%) had a neighborhood density of 0–5, and only 17% of the words had density measures over 100.

Previous research has found a relation between phonotactic probability and neighborhood density (Hoover et al., 2010; Storkel et al., 2006; Storkel & Lee, 2011). However, results of this analysis did not reveal a relationship between the two. Additionally, there was little-to-no relation among these factors and word learning in first- and second-grade models. On average, the correlation between neighborhood density and phonotactic probability was −0.23 (range of −0.18 to -0.26). These findings are similar to that of Storkel et al. (2006) who were unable to demonstrate an interaction between phonotactic probability and neighborhood density in a study examining adult word learning. Our findings could be attributed to the correlation between word length and these lexical characteristics since words varied in length and were typically multisyllabic. When words vary in length, problems in analysis and interpretation can occur (Storkel, 2004). This could explain why we did not find significant relations between word learning and phonotactic probability and neighborhood density.

There are several limitations of this investigation worth noting. First, this vocabulary program was implemented in schools that primarily served children from low-income families. Researchers like Hart and Risley (1995) have demonstrated a significant difference in language acquisition across children from families with varying socioeconomic levels. The results could differ if implemented in schools serving children from families with a wide range of income levels. Second, we were unable to control for child-level characteristics in our analyses. If included, analyses may reveal differential effects of lexical characteristics on the acquisition of words and their meanings based on a child's individual abilities. Third, some of our lexical characteristic measures differed greatly in their variability and range of values, which may skew our results. If this was controlled for, analyses may reveal different effects.

Clinical Implications

The cornerstone of vocabulary instruction is to select words children will need to know to comprehend academic texts. These words are often sophisticated synonyms for known words (i.e., gorgeous instead of pretty), but are difficult to infer meaning from context alone. Trying to organize vocabulary targets for instruction can be difficult. Among various attempts to order vocabulary targets (Beck et al., 2002; Biemiller, 2006; Biemiller & Boote, 2006; Marzano & Simms, 2013), there is still much variability in the words chosen for instruction. This variability must be minimized to make word selection a more structured process. Further investigations are needed to determine the optimal developmental groupings for words using relevant lexical characteristics. Grouping words and their meanings based on developmental appropriateness will provide teachers a more focused list of words, reducing the variability and increasing the uniformity in the word selection process.

Based on the promising results of applying MARS to the ILIAD data set, other analytic problems may be pursued using this method. MARS provided nuanced information about the influence of lexical characteristics on word learning using information based on hinges. Hinge placement created local regressions that combined to form a complete regression model for each of the relevant predictor variables, and these localized trends varied for each lexical characteristic. These localized trends allowed us to see what was happening in small intervals with more precision whereas a linear regression line would

Downloaded from: https://pubs.asha.org ERIC on 03/03/2023, Terms of Use: https://pubs.asha.org/pubs/rights_and_permissions

Table 6. Multivariate adaptive regression splines model prediction of second grade word learning.

Word	Age of acquisition	Level of concreteness	Word frequency	Neighborhood density	Phonotactic probability	Predicted learning
advise	8.89	2.03	12.20	69.83	.09	22%
illegal	9.21	2.37	23.51	38.42	.26	34%
space	5.67	3.54	66.06	32.03	.19	76%

only provide an overall trend. Hinges provided information about change within the models that could help to better explain how lexical characteristics affect word learning. The next step is to apply MARS to a different word learning data set. Outcomes that corroborate our findings may have important educational implications for vocabulary instruction. If we can model similar results with word learning outcomes from studies with new participants and different words, it would strengthen our argument for using lexical characteristics to create a developmentally appropriate sequence of vocabulary targets for instruction that would facilitate improved word learning in children.

One way we could use this information is to predict the percent of children who would learn words and their meanings using the model. To illustrate this example, three words were randomly selected. The words, their respective lexical characteristics, and the predicted learning for second-grade students are listed in Table 6. The model predicted that only 22% of second graders would learn the word advise, compared with the word space, where the model predicted 76% of second graders would learn it. This information could be used to select vocabulary words that are more appropriate for children given their grade level. The results of these future studies could expand our understanding of the way lexical characteristics, and not simply a word's tier, affects children's vocabulary mastery across various developmental stages. This may not completely resolve the issue of "which words to teach when," but is an attempt to fill one of the gaps in vocabulary instruction.

Data Availability Statement

The data sets generated during and/or analyzed during this study are available in the figshare repository, https://doi.org/10.6084/m9.figshare.16699594.v1. The program code used for this study is available in the figshare repository, https://doi.org/10.6084/m9.figshare.16699816.v1.

Acknowledgments

This research was supported by the Institute of Education Sciences, U.S. Department of Education, through Grants R324L060023 and R324A200179. The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education. This research was not preregistered in an independent, institutional registry.

References

Balota, D. A., Yap, M. J., Hutchison, K. A., Cortese, M. J., Kessler, B., Loftis, B., Neely, J. H., Nelson, D. L., Simpson, G. B., & Treiman, R. (2007). The English Lexicon Project. *Behavior Research Methods, 39*(3), 445–459. https://doi.org/10.3758/BF03193014

Beck, I. L., McKeown, M. G., & Kucan, L. (2002). *Bringing words to life: Robust vocabulary instruction*. Guilford Press.

Biemiller, A. (2005). Size and sequence in vocabulary development: Implications for choosing words for primary grade vocabulary instruction. In E. H. Hiebert & M. L. Kamil (Eds.), *Teaching and learning vocabulary: Bringing research to practice* (pp. 223–243). Routledge.

Biemiller, A. (2006). Vocabulary development and instruction: A prerequisite for school learning. In D. K. Dickinson & S. B. Neuman (Eds.), *Handbook of early literacy research* (Vol. 2, pp. 41–51). Guilford Press.

Biemiller, A. (2010). *Words worth teaching: Closing the vocabulary gap*. McGraw-Hill SRA.

Biemiller, A., & Boote, C. (2006). An effective method for building meaning vocabulary in primary grades. *Journal of Educational Psychology, 98*(1), 44–62. https://doi.org/10.1037/0022-0663.98.1.44

Biemiller, A., & Slonim, N. (2001). Estimating root word vocabulary growth in normative and advantaged populations: Evidence for a common sequence of vocabulary acquisition. *Journal of Educational Psychology, 93*(3), 498–520. https://doi.org/10.1037/0022-0663.93.3.498

Bloom, P., & Markson, L. (1998). Capacities underlying word learning. *Trends in Cognitive Sciences, 2*(2), 67–73. https://doi.org/10.1016/S1364-6613(98)01121-8

Brysbaert, M., & New, B. (2009). Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods, 41*(4), 977–990. https://doi.org/10.3758/BRM.41.4.977

Brysbaert, M., Warriner, A., & 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

Coxhead, A. (2000). A new academic word list. *TESOL Quarterly, 34*(2), 213–238. https://doi.org/10.2307/3587951

Crow, S., & O'Leary, A. (2015). *Word health: Addressing the word gap as a public health crisis*. Next Generation.

Dale, E., & O'Rourke, J. (1976). *The living word vocabulary, the words we know: A national vocabulary inventory*. Dome, Inc.

Downloaded from: https://pubs.asha.org ERIC on 03/03/2023, Terms of Use: https://pubs.asha.org/pubs/rights_and_permissions

Dollaghan, C. A. (1994). Children's phonological neighbourhoods: Half empty or half full? *Journal of Child Language, 21*(2), 257–271. https://doi.org/10.1017/S0305000900009260

Francis, W., & Kučera, H. (1982). *Frequency analysis of English usage.*

Friedman, J., Hastie, T., & Tibshirani, R. (2001). *The elements of statistical learning*. Springer Series in Statistics.

Frisch, S., Large, N., & Pisoni, D. (2000). Perception of wordlikeness: Effects of segment probability and length on the processing of nonwords. *Journal of Memory and Language, 42*(4), 481–496. https://doi.org/10.1006/jmla.1999.2692

Gierut, J., & Dale, R. (2007). Comparability of lexical corpora: Word frequency in phonological generalization. *Clinical Linguistics & Phonetics, 21*(6), 423–433. https://doi.org/10.1080/02699200701299891

Gilhooly, K. J., & Gilhooly, M. L. (1980). The validity of age-of-acquisition ratings. *British Journal of Psychology, 71*(1), 105–110. https://doi.org/10.1111/j.2044-8295.1980.tb02736.x

Gilhooly, K. J., & Logie, R. H. (1980). Age-of-acquisition, imagery, concreteness, familiarity, and ambiguity measures for 1,944 words. *Behavior Research Methods & Instrumentation, 12*(4), 395–427. https://doi.org/10.3758/BF03201693

Goldstein, H., Ziolkowski, R. A., Bojczyk, K. E., Marty, A., Schneider, N., Harpring, J., & Haring, C. D. (2017). Academic vocabulary learning in first through third grade in low-income schools: Effects of automated supplemental instruction. *Journal of Speech, Language, and Hearing Research, 60*(11), 3237–3258. https://doi.org/10.1044/2017_JSLHR-L-17-0100

Gray, S., & Yang, H.-C. (2015). Selecting vocabulary words to teach. *SIG 1 Perspectives on Language Learning and Education, 22*(4), 123–130. https://doi.org/10.1044/lle22.4.123

Hadley, E., Dedrick, R., Dickinson, D., Kim, E., Hirsh-Pasek, K., & Golinkoff, R. (2021). Exploring the relations between child and word characteristics and preschoolers' word-learning. *Journal of Applied Developmental Psychology, 77,* 101332. https://doi.org/10.1016/j.appdev.2021.101332

Hadley, E. B., & Mendez, K. Z. (2021). A systematic review of word selection in early childhood vocabulary instruction. *Early Childhood Research Quarterly, 54,* 44–59. https://doi.org/10.1016/j.ecresq.2020.07.010

Hart, B., & Risley, T. (1995). *Meaningful differences in the everyday experiences of young American children*. Brookes.

Hoover, J. R., Storkel, H. L., & Hogan, T. P. (2010). A cross-sectional comparison of the effects of phonotactic probability and neighborhood density on word learning by preschool children. *Journal of Memory and Language, 63*(1), 100–116. https://doi.org/10.1016/j.jml.2010.02.003

Kilic Depren, S. (2018). Prediction of students' science achievement: An application of multivariate adaptive regression splines and regression trees. *Journal of Baltic Science Education, 17*(5), 887–903. http://doi.org/10.33225/jbse/18.17.887

Kolyshkina, I., Brownlow, M., & Taylor, J. (2013). Improving every child's chance in life. In *2013 IEEE 13th International Conference on Data mining Workshops* (pp. 180–184). IEEE. https://doi.org/10.1109/ICDMW.2013.61

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

Luce, P. A., & Pisoni, D. B. (1998). Recognizing spoken words: The neighborhood activation model. *Ear and Hearing, 19*(1), 1–36. https://doi.org/10.1097/00003446-199802000-00001

Maekawa, J., & Storkel, H. L. (2006). Individual differences in the influence of phonological characteristics on expressive vocabulary development by young children. *Journal of Child Language, 33*(3), 439–459. https://doi.org/10.1017/S0305000906007458

Martís, R., Alonso, J., Catalán, C., Fuentes, R., & Suárez, A. (2015). Prediction of the student success rate by means of quality teaching survey variables applying a multivariate adaptive regression splines (Mars) models. In *Toulon-Verona Conference "Excellence in Services"* (pp. 1–14).

McDonough, C., Song, L., Hirsh-Pasek, K., Golinkoff, R. M., & Lannon, R. (2011). An image is worth a thousand words: Why nouns tend to dominate verbs in early word learning. *Developmental Science, 14*(2), 181–189. https://doi.org/10.1111/j.1467-7687.2010.00968.x

Marzano, R., & Simms, J. (2013). *Vocabulary for the common core*. Marzano Research Laboratory.

McLoyd, V. C. (1998). Socioeconomic disadvantage and child development. *American Psychologist, 53*(2), 185–204. https://doi.org/10.1037/0003-066X.53.2.185

Meschyan, G., & Hernandez, A. (2002). Age of acquisition and word frequency: Determinants of object-naming speed and accuracy. *Memory & Cognition, 30*(2), 262–269. https://doi.org/10.3758/BF03195287

Milborrow, S., & Milborrow, M. (2007). *The earth package*. The R Project for Statistical Computing.

National Reading Panel. (2000). *Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction*. National Institute of Child Health and Human Development.

Newman, R. S., & German, D. I. (2002). Effects of lexical factors on lexical access among typical language-learning children and children with word-finding difficulties. *Language and Speech, 45*(3), 285–317. https://doi.org/10.1177/00238309020450030401

Paivio, A., Yuille, J. C., & Madigan, S. A. (1968). Concreteness, imagery, and meaningfulness values for 925 nouns. *Journal of Experimental Psychology, 76*(1, Pt. 2), 1–25. https://doi.org/10.1037/h0025327

Ponari, M., Norbury, C. F., & Vigliocco, G. (2018). Acquisition of abstract concepts is influenced by emotional valence. *Developmental Science, 21*(2), e12549. https://doi.org/10.1111/desc.12549

RStudio Team. (2020). *RStudio: Integrated development for R.* RStudio. http://www.rstudio.com/

Sanders, H. (2021). *"Multivariate adaptive regression splines (MARS) method for word learning analysis" code for analysis.* Figshare. https://doi.org/10.6084/m9.figshare.16699816.v1

Sanders, H., Peters-Sanders, L., & Goldstein, H. (2021). *ILIAD data—Cleaned* [data set]. Figshare. https://doi.org/10.6084/m9.figshare.16699594.v1

Spreen, O., & Schulz, R. W. (1966). Parameters of abstraction, meaningfulness, and pronunciability for 329 nouns. *Journal of Verbal Learning and Verbal Behavior, 5*(5), 459–468. https://doi.org/10.1016/S0022-5371(66)80061-0

Stahl, S. A., & Nagy, W. E. (2007). *Teaching word meanings*. Routledge. https://doi.org/10.4324/9781410615381

Stokes, S. F. (2010). Neighborhood density and word frequency predict vocabulary size in toddlers. *Journal of Speech, Language, and Hearing Research, 53*(3), 670–683. https://doi.org/10.1044/1092-4388(2009/08-0254)

Storkel, H. L. (2004). Methods for minimizing the confounding effects of word length in the analysis of phonotactic probability and neighborhood density. *Journal of Speech, Language, and Hearing Research, 47*(6), 1454–1468. https://doi.org/10.1044/1092-4388(2004/108)

Storkel, H. L. (2009). Developmental differences in the effects of phonological, lexical and semantic variables on word learning

Downloaded from: https://pubs.asha.org ERIC on 03/03/2023, Terms of Use: https://pubs.asha.org/pubs/rights_and_permissions

by infants. *Journal of Child Language, 36*(2), 291–321. https://doi.org/10.1017/S030500090800891X

Storkel, H. L. (2013). A corpus of consonant–vowel–consonant real words and nonwords: Comparison of phonotactic probability, neighborhood density, and consonant age of acquisition. *Behavior Research Methods, 45*(4), 1159–1167. https://doi.org/10.3758/s13428-012-0309-7

Storkel, H. L., Armbrüster, J., & Hogan, T. P. (2006). Differentiating phonotactic probability and neighborhood density in adult word learning. *Journal of Speech, Language, and Hearing Research, 49*(6), 1175–1192. https://doi.org/10.1044/1092-4388(2006/085)

Storkel, H. L., Bontempo, D. E., Aschenbrenner, A. J., Maekawa, J., & Lee, S. Y. (2013). The effect of incremental changes in phonotactic probability and neighborhood density on word learning by preschool children. *Journal of Speech, Language, and Hearing Research, 56*(5), 1689–1700. https://doi.org/10.1044/1092-4388(2013/12-0245)

Storkel, H. L., & Lee, S. Y. (2011). The independent effects of phonotactic probability and neighbourhood density on lexical acquisition by preschool children. *Language and Cognitive Processes, 26*(2), 191–211. https://doi.org/10.1080/01690961003787609

Vaden, K. I., Halpin, H. R., & Hickok, G. S. (2009). *Irvine Phonotactic online dictionary* (Version 2.0). http://www.iphod.com

Vitevitch, M. S., & Luce, P. A. (2004). A web-based interface to calculate phonotactic probability for words and nonwords in English. *Behavior Research Methods, Instruments, & Computers, 36*(3), 481–487. https://doi.org/10.3758/BF03195594

Vitevitch, M. S., Luce, P. A., Charles-Luce, J., & Kemmerer, D. (1997). Phonotactics and syllable stress: Implications for the processing of spoken nonsense words. *Language and Speech, 40*(1), 47–62. https://doi.org/10.1177/002383099704000103

Walker, D., Greenwood, C., Hart, B., & Carta, J. (1994). Prediction of school outcomes based on early language production and socioeconomic factors. *Child Development, 65*(2), 606–621. https://doi.org/10.1111/j.1467-8624.1994.tb00771.x

Willett, J. B., & Singer, J. D. (1988). Another cautionary note about R^2: Its use in weighted least-squares regression analysis. *The American Statistician, 42*(3), 236–238. https://doi.org/10.1080/00031305.1988.10475573

Downloaded from: https://pubs.asha.org ERIC on 03/03/2023, Terms of Use: https://pubs.asha.org/pubs/rights_and_permissions