Journal of Experimental Psychology: Learning, Memory, and Cognition

Finding Words in a Sea of Text: Word Search as a Measure of Sensitivity to Statistical Regularities in Reading

Erin S. Isbilen, Abigail Laver, Noam Siegelman, James S. Magnuson, and Richard N. Aslin

Online First Publication, February 13, 2025. https://dx.doi.org/10.1037/xlm0001412

AMERICAN PSYCHOLOGICAL ASSOCIATION

Journal of Experimental Psychology: Learning, Memory, and Cognition

ISSN: 0278-7393

https://doi.org/10.1037/xlm0001412

Finding Words in a Sea of Text: Word Search as a Measure of Sensitivity to Statistical Regularities in Reading

Erin S. Isbilen[1], Abigail Laver[2], Noam Siegelman[3, 4], James S. Magnuson[5, 6, 7], and Richard N. Aslin[1, 8]

[1] Yale Child Study Center, Yale University
[2] Department of Psychology, University of Pennsylvania
[3] Department of Psychology, The Hebrew University of Jerusalem
[4] Department of Cognitive and Brain Sciences, The Hebrew University of Jerusalem
[5] Department of Psychological Sciences, University of Connecticut
[6] Basque Center on Cognition, Brain and Language, Donostia-San Sebastían, Spain
[7] Ikerbasque: Basque Foundation for Science, Bilbao, Spain
[8] Department of Psychology, Yale University

Statistical learning (SL) is hypothesized to play a fundamental role in reading, yet the correlations between reading and SL are largely mixed. This inconsistency may result from the fact that most SL studies train participants to learn novel, nonlinguistic visual regularities, which overlooks two important factors: (a) SL performance varies across domains, and (b) most SL studies utilize tasks with short exposure phases with a limited set of novel structured stimuli. Rather than exposing participants to novel statistics, we explored how prior learning of the statistical regularities inherent in natural texts predicts individual differences in reading. We developed a novel measure of long-term orthographic SL by assessing participants' ability to chunk letter information based on its statistical properties. Adults were prompted to find high- and low-frequency English words (derived from written-language corpora) when a single target word was embedded in an array of background distractors comprising letters that do not form words. Performance on this task was compared against three established measures of component skills of reading: lexical decision, orthographic awareness, and spelling recognition. Participants were faster and more accurate at identifying high-frequency words, replicating classic psycholinguistic results. Performance was also impacted by semantic diversity—the variation of the semantic contexts a word appears in—independent of frequency. Critically, word search performance significantly predicted each reading subtest, suggesting that the task draws upon key reading-related skills. Sensitivity to orthographic statistical structure may serve as a crucial foundation that drives individual differences in reading, consistent with SL-based accounts of language.

Keywords: statistical learning, reading, individual differences, literacy, visual search

Supplemental materials: https://doi.org/10.1037/xlm0001412.supp

Statistical learning—the ability to detect statistical regularities in the environment—is foundational to many theories in cognitive science. This basic mechanism promotes the extraction of meaningful units from sensory input by enabling the brain to group recurring structures into larger chunks based on their statistical properties. One of the most formative demonstrations of this came from the pivotal Saffran et al. (1996) study, which found that infants as young as 8 months old could capitalize on syllable co-occurrences in an artificial language to discover words after only 2 min of passive exposure. Analogous effects have since been reported across a broad range of domains and modalities (e.g., Conway & Christiansen, 2005), including visual pattern learning of both spatial and temporal inputs (Fiser & Aslin, 2001, 2002), revealing the ubiquity of statistical learning across a wide range of materials.

The discovery of statistical learning's versatility marked a theoretical shift in many subfields of the cognitive sciences, especially

L. Robert Slevc served as action editor.

Erin S. Isbilen https://orcid.org/0000-0002-7233-0805

This research was in part supported by a National Institute of Child Health and Human Development, National Institutes of Health National Research Service Award postdoctoral research grant (F32HD104542) awarded to Erin S. Isbilen and a National Institutes of Health, National Institute of Child Health and Human Development research grant (HD-037082) awarded to Elissa Newport (subcontract to Richard N. Aslin).

Erin S. Isbilen played a lead role in conceptualization, data curation, formal analysis, funding acquisition, investigation, methodology, project administration, resources, software, validation, visualization, writing–original draft, and writing–review and editing. Abigail Laver played a supporting role in investigation, methodology, project administration, and software. Noam Siegelman played a supporting role in conceptualization, formal analysis, methodology, and writing–review and editing. James S. Magnuson played a supporting role in formal analysis, methodology, supervision, visualization, and writing–review and editing. Richard N. Aslin played a supporting role in conceptualization, funding acquisition, investigation, methodology, project administration, supervision, and writing–review and editing.

Correspondence concerning this article should be addressed to Erin S. Isbilen, Yale Child Study Center, Yale University, 300 George Street, No. 900, New Haven, CT 06511, United States. Email: erin.isbilen@yale.edu

in the study of language. It supplied a presumably domain-general process that could capture the immense richness of linguistic pattern learning, counter to long-standing hypotheses about innate, domain-specific mechanisms dedicated to language acquisition (Chomsky, 1959; Eimas, 1985). Since then, increased empirical attention has been devoted to deciphering whether the learning observed for artificial stimuli in the lab scales up to the complexity of learning in the real world and, specifically, whether statistical learning can predict individual differences in linguistic proficiency. This research is predicated on the ideas that statistical learning facility, regardless of domain, provides comparable advantages in language and that learning on the timescale of a single test session reflects relative individual differences in the learning that unfolds over years of linguistic exposure.

Ample literature now highlights how statistical learning scaffolds a diverse assortment of behaviors related to spoken language, including word learning (Swingley, 2005), and syntax acquisition (Kidd, 2012; Thompson & Newport, 2007). However, written language also contains a wealth of statistical regularities, with proficient reading necessitating the integration of statistical information from multiple sources (see Sawi & Rueckl, 2019, for a review). To date, statistical learning has been found to mediate three critical elements of reading—orthography, phonology, and semantics—as well as the connections between them (He & Tong, 2017). Specifically, it can support the discovery of letter combinations that form morphological units (Crepaldi et al., 2010; Vidal et al., 2021), and computational work that compares model results to human data provides a strong correlational case for the role of statistical learning in the mapping of phonology to orthography, with learners tracking the regularities between them (Harm & Seidenberg, 2004; Seidenberg & McClelland, 1989). Furthermore, sensitivity to orthographic regularities has been found to underlie spelling acquisition and ultimately reading outcomes in young children (Rothe et al., 2014; Treiman & Kessler, 2022). Preliterate children who exhibit advanced statistical learning on artificial language tasks display marked improvements in foundational reading precursors, including vocabulary, oral language, and phonological processing (Spencer et al., 2015). Enhanced statistical learning has also been associated with advanced reading in neurotypical populations (e.g., Arciuli & Simpson, 2012), while clinical populations with reading impairment display significant deficits in this ability (Gabay et al., 2015; Kahta & Schiff, 2019), which has sparked increased interest in using statistical learning to characterize reading disabilities. These findings, along with many others, illustrate the importance of statistical computations to literacy and have cemented statistical learning as a foundational process in many theories of reading development.

Despite the promise of statistical learning as a key component of reading, a growing body of research reveals that the correlations between performance on lab-based statistical learning tasks and literacy are not always consistent. In fact, an increasing number of studies raise questions about the robustness and reliability of statistical learning effects in predicting reading ability (e.g., Arnon, 2019; Schmalz et al., 2017), often uncovering weak or even null correlations between the two, which complicates our understanding of the role of statistical learning in literacy. These varied results may in part derive from the fact that statistical learning is often treated as a singular amodal mechanism that operates identically across domains, with the supposition that proficiency in one modality and domain denotes equivalent aptitude in any other (R. Frost et al., 2015). However, statistical learning is often nonuniform across inputs, which challenges the view of statistical learning as a unitary capacity. For example, statistical learning follows modality- and domain-specific developmental trajectories. The learning of visual and auditory-nonlinguistic statistical information improves steadily with age, while learning in the auditory-linguistic domain remains relatively consistent in children between the ages of 5 and 12 years old (Raviv & Arnon, 2018; Shufaniya & Arnon, 2018; see also Ren & Wang, 2023). Furthermore, learning in one modality does not always predict one's ability to learn the same kind of structure in another, and the ability to process different types of regularities—even in the same modality and domain—often fails to correlate within individuals (e.g., adjacent and nonadjacent auditory-linguistic regularities; Siegelman & Frost, 2015).

These findings challenge existing notions of how to characterize statistical learning proficiency: It varies not only across but also *within* individuals, subject to the constraints of modality, domain, and even structure (Bogaerts et al., 2022). This feeds into a major ongoing theoretical discussion about how statistical learning may encompass an assemblage of computations rather than a singular mechanism, drawing from other cognitive systems including attention and memory (Baker et al., 2004; Hall et al., 2015; Palmer & Mattys, 2016). This in turn suggests that statistical learning is more uniquely tailored to the input than previously assumed, with different processing weights being applied to multiple underlying mechanisms depending on the task at hand.

Inconsistencies, both within and between individuals, in learning across stimulus types raise an important consideration for how to approach the comparison between statistical learning and other cognitive skills, including literacy—the focus of the current work. Although a growing body of evidence attempts to link statistical learning to reading, the overwhelming majority utilize separate, artificial statistical learning tasks that present nonlinguistic visual stimuli—items that bear limited resemblance to the orthographic regularities of written language and which may draw upon slightly different suites of computations. To illustrate, a comprehensive meta-analysis by Ren et al. (2023) revealed that of the 137 studies in their sample that evaluated the connection between statistical learning and reading, 76 tested visual-nonlinguistic statistical learning while only 26 tested visual-linguistic statistical learning (30 of the remaining studies tested auditory-linguistic and five tested auditory-nonlinguistic). A similar breakdown was reported in an earlier meta-analysis by Lee et al. (2022) on the linkage between statistical learning and reading in individuals with and without developmental dyslexia, with 38 studies testing visual-nonlinguistic statistical learning, five testing visual-linguistic statistical learning (one tested both visual-linguistic and visual-nonlinguistic), three testing auditory-linguistic statistical learning (one tested both auditory-linguistic and nonlinguistic), and 11 testing audiovisual statistical learning (six linguistic, one nonlinguistic, four both linguistic and nonlinguistic). Furthermore, in both meta-analyses, many of these tasks test sequential visual statistical learning (e.g., individual items presented one after another), whereas reading requires the processing of both temporal and spatial information (e.g., grouping sequences of adjacent letters into words).

Despite the fact that most studies seeking to link statistical learning and reading utilize stimulus materials highly divergent from native-language orthographies, Lee et al. (2022) found a domain-general disadvantage of statistical learning in developmental dyslexia (mean $d = .47$, $p < .0001$), while Ren et al. (2023) reported, in their own words, a “significant, moderate correlation” between statistical learning and reading-related outcomes ($r = .24$, $p < .001$). Furthermore, although Ren et al. (2023) found no significant difference in the predictive power of statistical learning on reading as a function of learning modality or domain, it is possible that the two may interact. If statistical learning computations vary across domains, artificial stimuli and lab-based learning tasks may not fully capture the processing of the statistics required for reading because of stimulus-specific differences. Furthermore, it is possible that the initial stages of learning captured by a single statistical learning test session may yield less interindividual variation in performance than the learning that unfolds over long periods of time. Participants may also be more uniform in their acquisition of simple artificial statistics than they are in their ability to learn the complicated information embedded in natural text.

Part of the reliance on artificial tasks, in addition to offering greater experimental control, stems from a drive to establish the unique contribution of statistical learning without interference from prior knowledge. Participants do not enter the lab as blank slates; they arrive armed with a lifetime of linguistic experience. This point has been poignantly demonstrated by statistical learning studies that simulate second language acquisition, where the regularities of the first artificial language that participants were trained on impede acquisition of the second when the two share a syllabic inventory (Franco et al., 2011; Gebhart et al., 2009). Similarly, the biases that participants carry from their native language critically impact their ability to learn novel words in statistical learning tasks. Stimuli that feature high-frequency syllable combinations from participants’ native language yield significantly better learning than items with low syllable co-occurrence frequencies (Elazar et al., 2022; Siegelman et al., 2018), suggesting facilitation (or interference) from accrued linguistic knowledge.

While this observation has inspired some researchers to pivot away from linguistic stimuli in favor of artificial stimuli unknown by participants to circumvent the influence of prior learning (e.g., sequences of random sounds; Siegelman et al., 2018), we argue that prior knowledge may also supply a rich source of individual differences. As skilled reading necessitates assimilating a multitude of statistics over decades of exposure, naturalistic tasks that mine this wealth of distributional information may better inform how statistical learning shapes language functions by comparing like-to-like in the kinds of knowledge, and presumably computations, utilized across tasks. Indeed, aligning the structures targeted in statistical learning and natural language tasks yields more robust correlations specifically because their processing is supported by a common suite of mechanisms (Isbilen, McCauley, & Christiansen, 2022). This is in line with the multicomponential view of statistical learning that theorizes a suite of cognitive processes that vary according to the input characteristics, rather than a singular mechanism (R. Frost et al., 2015). Furthermore, tasks that test sensitivity to natural language statistics have been shown to afford clearer theoretical predictions about the precise statistical learning mechanisms that support skilled reading (e.g., print-speech and print-meaning regularities; Siegelman et al., 2020, 2022).

The present study seeks to establish how individual differences in prior statistical learning that participants bring to the task inform reading proficiency. Rather than evaluating how statistical learning on the timescale of minutes, as it unfolds over a single training session, is correlated with reading skill, we focus on sensitivity to written regularities resulting from learning over the course of years of reading experience. Instead of introducing participants to artificial statistics, we analyze their long-term statistical knowledge of real English text and how their experience as life-long statistical learners influences reading.

We created a novel visual search task based on classic word search games, where individuals hunt for words hidden among a host of random letters that do not form words. While most experiments that forge the link with reading focus on temporal visual statistical learning (Ren et al., 2023), our focus was on the spatial aspects of reading. This is because reading of extended text involves a series of eye movements and fixations, which naturally involve spatial aspects of learning and a heavy reliance on working memory (Rayner & McConkie, 1976). Furthermore, existing accounts suggest that the spatial and temporal components of statistical learning may be to some degree separable. For example, while both types of learning are highly correlated and appear to be governed by a common capacity for extracting regularities across time and space, individuals also differ in their ability to learn one kind of regularity over the other (Growns et al., 2020). This may be due to the involvement of separate brain regions, such as the greater involvement of the visual cortex in spatial statistical learning, which is more sensitive to spatial regularities over nonspatial (Chen & Vroomen, 2013). Testing sensitivity to spatial regularities may therefore better inform how statistical learning facilitates subcomponents of the reading process that exploit similar information. While prior studies of statistical learning and visual search do exist, they have primarily adopted a contextual cuing approach by training learners to associate particular stimuli with specific locations in arrays that feature multiple competing stimuli (Theeuwes et al., 2022), focusing on how spatial statistical learning can tune learners’ attention to relevant information while helping them suppress irrelevant information (Ferrante et al., 2018). By contrast, we were interested in testing how sensitivity to orthographic regularities in familiar text impacts visual search when words are embedded in horizontal lines of random letters, or how sensitivity to statistical structure guides the ability to detect visual words in noise.

In line with recent statistical learning studies that aim to reveal reliable individual differences by aligning the computations used in different tasks (Bogaerts et al., 2022; Isbilen, McCauley, & Christiansen, 2022), we created a task that draws on computations known to be central to statistical learning and language: statistically based chunking. Chunking describes the rapid grouping together of low-level features into higher levels of representation. This domain-general process enables the cognitive system to recode incoming sensory information into larger units of abstraction based on statistical regularities, such as syllables into words (Isbilen et al., 2020; Perruchet & Vinter, 1998) and words into multiword phrases (McCauley & Christiansen, 2019). Historically, segmentation in the context of statistical learning was implicitly assumed to be a unicomponential process, proceeding based on probability computation alone (R. Frost et al., 2015). However, subsequent work reveals that chunking is central to the segmentation behaviors observed in classic statistical learning tasks, with chunking working in tandem with statistical computations to facilitate the extraction of words from speech (Isbilen et al., 2020). The chunking hypothesis of statistical learning thereby provides more explicit predictions about the specific subcomponent skills involved in both the segmentation described by conventional statistical learning tasks and those involved in language and, in turn, better accounts for statistical learning performance (Isbilen, Frost, et al., 2022).

Segmenting words from speech requires recognizing candidate chunks while rejecting statistically lower probability items (e.g., rejecting "tyba" as a candidate word in the phrase "pretty baby"; Saffran, 2003). Performance on our word search task relies on similar statistical chunking mechanisms: Participants must rapidly group together letters (low-level linguistic features) among a host of "nonchunkable" competitors into a word (a higher level linguistic feature) based on the statistics of English. Importantly, statistically based chunking plays a foundational role in language (see Christiansen & Chater, 2016; Perruchet, 2019, for reviews), suggesting that it may serve as a bridge between classic statistical learning tasks and reading. Statistical learning as measured via chunking-based recall tasks significantly predicts cross-individual variation in the ability to process speech in noise (Conway et al., 2010). It is also highly predictive of participants' reading efficiency: For example, participants' ability to recall items derived from high-frequency phoneme statistics significantly predicts their reading speed on self-paced reading tasks (McCauley et al., 2017), as does their sensitivity to high-frequency consonant letter combinations (McCauley & Christiansen, 2015). Thus, chunking tasks likely tap into learners' sensitivity to orthographic statistics and better predict reading, since they recruit similar sets of computations and measure statistical knowledge that bears directly on literacy.

In our word search task, we embedded a single target word (four to eight letters long) in 12×12 visual arrays populated with random letter combinations. Participants were allotted 20 s to find the target word, which was written in a left-to-right orientation reflective of typical English text. We manipulated three statistical properties that critically impact visual word recognition: word frequency, letter bigram frequency, and semantic diversity. A long psycholinguistic legacy emphasizes the powerful influence of frequency on language learning and processing (see Brysbaert et al., 2018, for a review). High-frequency words are recognized faster than low-frequency words (Forster & Chambers, 1973), as are words containing frequent bigrams (two-letter combinations; Rice & Robinson, 1975). However, frequency is not the sole determinant of word recognition: The number of semantic contexts a word appears in also impacts processing. Semantic diversity (derived via latent semantic analysis; Landauer & Dumais, 1997) is an index of the mean similarity of the semantic contexts within which a word appears. The more unique contexts a word appears in, the greater the semantic diversity score. Adult (Hoffman & Woollams, 2015) and developmental studies (Hsiao & Nation, 2018) reveal faster reaction times to words that occur in more diverse semantic contexts independent of word frequency, although Cevoli et al. (2021) reported that while semantic diversity bolsters visual word recognition, it also significantly interacts with word frequency. Semantic diversity therefore encompasses a more abstract kind of statistical regularity that is relevant to reading: It tallies the number of unique semantic contexts a word appears in rather than the orthographic statistics that characterize the visual properties of a word.

In our preregistered hypotheses, we predicted that individuals would be more accurate at identifying high-frequency words than low-frequency words (Experiment 1). We further hypothesized that words with high bigram frequencies might be easier to find than those with low bigram frequencies, since participants appear to represent both bigram and whole-pattern information in statistical learning tasks (Siegelman et al., 2019). We also predicted that shorter words would be easier to find than longer words, in line with data showing that individuals take longer to read longer words (both high and low frequency) across development (e.g., Gerth & Festman, 2021; Kliegl et al., 2004). However, we were conservative in our estimates about the unique impact of semantic diversity, given the significant interaction with frequency reported by Cevoli et al. (2021). More importantly for our goals in designing the word search task, we hypothesized that the ability to successfully find the target word within a time-limited search would significantly correlate with key component skills of reading (written word identification, spelling recognition, and orthographic awareness; Experiment 2), since performance on both types of tasks likely recruits similar mechanisms and leverages comparable long-term statistical knowledge. Support for our hypotheses would significantly extend prior work on the formative role of statistical learning processes in language, illuminating how these foundational mechanisms support the domain of reading.

Experiment 1: Word Search as a Measure of Natural Language Statistical Learning

Method

Participants

A total of 80 English-speaking monolinguals who reported no experience with other languages were recruited from the Prolific participant platform (https://www.prolific.co/). This number was based on a power analysis from a pilot study with an estimated effect size of $d = .61$, power $= .9$, $p = .0001$, using the R package pwr (Champely et al., 2018). This effect size was estimated based on the results of a pilot version of the word search task, which was identical in its methods to the version reported here. Participants had no known auditory, visual, or language disorders and were compensated at the Prolific recommended rate at the time of posting the study ($8.00/hr). A total of seven participants were excluded from this data set for failing to meet criterion on attention checks and/or failure to follow instructions, yielding a final sample size of 73 participants (42 female, 25 male, six nonbinary; $M_{\text{age}} = 23$ years).

Materials

The target words were sourced from the English Lexicon Project (ELP; https://elexicon.wustl.edu/), an online corpus of words that contains their lexical characteristics and associated behavioral data from lexical decision and speeded naming tasks (Balota et al., 2007). Within this corpus, we utilized the Hyperspace Analogue to Language (Lund & Burgess, 1996) frequency norms, which are based on the data from 3,000 web-based newsgroups. These frequency norms are more representative of the statistics of written (rather than spoken) English, making them suitable for the purpose of this experiment.

Only nouns were selected as targets, to ensure approximately equal concreteness of the stimuli (i.e., the degree to which a word refers to something perceivable; Brysbaert et al., 2014), which was later verified in our analyses. None of the target words contained shorter words within them (e.g., the word "boxer" was excluded as a candidate stimulus, since it contains the words "box" and "ox"). All items had ELP lexical decision and naming accuracies of 100%, to ensure that they would be known by adult participants (see Mandera et al., 2020). Bigram frequency (i.e., the frequency of the two-letter combinations within each word) was controlled across stimuli, to guarantee that bigram and word frequency were not collinear (see the Results section for data on the variance inflation factors). The mean bigram frequencies (where a bigram is defined as a two-letter sequence) in the ELP corpus were calculated by the original authors as follows: First, the summed bigram frequency of each word was calculated. This was achieved by summing the successive frequencies of each word's bigrams (e.g., for DOG, the frequency of DO + OG in the entire corpus). The mean bigram frequency was then calculated by dividing the summed frequency by the number of successive bigrams in each word (Balota et al., 2007). Concreteness ratings in the corpus are based on the data from norming studies (Brysbaert et al., 2014), where the concreteness of words (defined as words that refer to things which can be experienced through one of the five senses) were rated on a 1–5 scale, where a higher score denoted greater concreteness. Estimates of the words' semantic diversity (the number of distinct semantic contexts a word appears in; Hoffman et al., 2013) and semantic neighborhood density (the mean proximity of semantic neighbors to the target; Shaoul & Westbury, 2010) were also extracted from the corpus and included as predictors of word finding ability in our analyses.

The final stimulus set consisted of 88 words: 44 high and 44 low frequency. Of these, 75 were monomorphemic. Raw frequencies were occurrences per 131 million words in the corpus. The natural log frequencies of the words were used. The target words were four to eight letters long and followed a Zipfian distribution that is reflective of English: Shorter words occurred as targets more often than longer words. This was done in part to tie the task to the statistics of reading more closely and to ensure that no single word type was more likely to occur in a certain starting position (e.g., an eight-letter word can only begin in a limited number of positions in a 12×12 array). Word length was equated across the different frequency types (high/low) to avoid conflating word length and frequency. All the words, and their associated statistics, are reported in Appendix A. In addition, summary statistics for the values of the high- and low-frequency words are reported in Table 1.

After the target words were selected, we constructed 88 boards, which comprised 12×12 square grids. An array of this size allowed us to vary word length more than would be possible with a smaller array and helped avoid repeated instances of target words occurring in the same position across trials. Each board contained a single target word, and the remaining spaces were populated by distractor letters (Figure 1). The distractor letters were sampled to reflect the distribution of letter frequency in English: High-frequency letters occurred in the background more often than low-frequency letters, once more based on the estimates of ELP. For example, since "E," a high-frequency letter, has a frequency of approximately 12% in ELP, it occurred as a background distractor 12% of the time, whereas "Z," a low-frequency letter in ELP, occurred as a distractor with its corresponding frequency (approximately 1%). These percentages were held constant across all boards. A table of these frequencies is reported in Appendix B. The mean bigram frequencies of the distractors were equated with those of the target words.

The background letters were generated such that they did not create English words when reading left-to-right (including two-letter words such as "it" or "on," etc.), or common English acronyms. The background distractors were also controlled such that they did not form words when combined with the target words. For example, for the target word "media," the word was not preceded by the letter "a," which would produce the word "am," or the letters "cla," which would produce the word "clam." Similarly, the word "media" was not followed by the letter "t," which would produce the word "at," or "dd," which would produce the word "add," or followed by the letter "l," which would produce the words "dial" and "medial." Efforts were also taken to reduce the presence of non-target words in the horizontal positions, although two-letter words (e.g., "at") do occasionally occur, but were never selected by participants. The positions of the words were controlled across frequency (high/low) and length (four to eight), such that no word type was more or less likely to occur in any given position in the grid. The target words only occurred in a horizontal, left-to-right configuration, reflecting the canonical structure of English text.

Procedure

Prior to beginning the word search task, participants were required to pass an attention check with 100% accuracy to be eligible for inclusion in the final sample. This attention check comprised a comprehension quiz about the task instructions: 1) what is the minimum number of letters a word can be in this task (four letters), 2) what orientation are words embedded in the board (left-to-right, the way English is read), 3) how many words are in each board (one), and 4) how should the letters be clicked when a word is found (starting with the first and ending with the last). Participants who passed the attention check but nonetheless exclusively selected random letters in the incorrect orientation (e.g., vertical and/or diagonal), which did not form words, were excluded ($N = 2$).

Participants were told that they would be playing a modified word search game. They were informed that each board only contained a single word and that their task was to find the word and click on each letter in the correct order, starting with the first and ending with the last. Unlike in traditional word search games, they were not told what the target words were prior to beginning: They were simply

Table 1
Summary Statistics by Item Type

Categorical frequency	Log frequency		Bigram frequency		Concreteness		Semantic diversity		Semantic neighborhood density	
	M (*SD*)	Range	*M* (*SD*)	Range	*M* (*SD*)	Range	*M* (*SD*)	Range	*M* (*SD*)	Range
High	10.10 (0.79)	8.74–11.70	1,420 (690)	100–2,961	3.83 (1.1)	1.33–5	1.63 (0.26)	.92–2.11	0.64 (0.04)	.51–.69
Low	6.64 (0.61)	5.30–8.00	1,510 (720)	151–2,778	4.06 (0.75)	1.88–5	1.37 (0.30)	.64–2.15	0.47 (0.07)	.31–.60

Figure 1
Example Boards Containing a Low-Frequency Word (Panel A) and a High-Frequency Word (Panel B)

(A)

T	H	N	R	M	S	D	L	O	A	A	I
U	E	E	A	H	T	C	T	Y	T	A	H
N	L	A	F	L	P	P	T	A	H	T	T
T	H	C	C	N	T	A	E	B	A	U	I
T	F	S	W	H	R	J	T	D	A	U	R
N	T	M	H	T	W	A	E	B	H	R	N
I	E	S	R	A	J	E	S	E	A	H	T
R	T	S	D	H	T	S	F	H	J	T	L
N	T	M	U	E	S	Q	H	N	A	E	E
G	M	O	L	A	R	U	A	H	E	R	L
I	P	A	E	B	T	N	L	T	S	E	P
E	O	E	E	I	R	R	M	D	N	I	N

(B)

M	N	I	D	A	T	I	I	X	S	M	I
N	S	L	N	T	H	L	A	A	O	O	I
T	I	H	H	F	X	U	H	O	L	C	I
T	E	S	Q	R	N	D	B	T	Y	N	A
F	A	F	I	M	J	L	H	P	C	C	T
N	I	R	P	O	L	M	T	N	T	R	P
P	N	L	N	H	S	U	T	E	L	P	E
A	E	J	S	M	N	L	U	H	R	N	I
T	T	D	S	J	A	O	O	I	E	U	R
E	U	I	S	U	A	H	O	A	E	U	O
T	L	L	T	Q	E	E	E	I	E	A	A
M	N	A	O	C	N	M	E	D	I	A	K

Note. Participants were tasked with clicking on each letter in the correct order within 20 s. The target words are highlighted here for illustration. See the online article for the color version of this figure.

instructed to search the boards for candidate words before the trial timed out.

Participants were informed that all of the words were four letters in length or longer and that words only appeared in a horizontal, left-to-right configuration, reflecting how they naturally read English. They were instructed to not look for words in the vertical or diagonal orientations. They were also informed that all of the targets were common English nouns and utilized American English spellings. Participants were given the chance to deactivate a square if they mis-selected.

A fixation cross located at the center of the board preceded every trial, and participants were required to click on it to initiate the trial. This was done to ensure that the starting position of the mouse was the same for every participant across every trial. Participants were given 20 s to find the word on each board, after which the trial timed out and the experiment automatically advanced to the next trial. This time-limited trial duration was selected from piloting to reduce ceiling effects and to better ensure that the task tapped into implicit statistical knowledge rather than explicit search strategies. All participants received the same boards to keep the background letter statistics and the word positions consistent across individuals, but the order of the boards was randomized across participants. The task of completing all 88 boards lasted approximately 40 min. Reaction times were calculated as the total amount of time that participants spent on each board (i.e., from board onset until the last letter of the target word was clicked).

Results

All analyses for this experiment were preregistered prior to data collection at AsPredicted.org (https://aspredicted.org/jf2yn.pdf). The analyses utilized the lme4 package (Bates et al., 2020) in the R Statistical Software, Version 4.0.2. (R Core Team, 2020). All data and code are available at https://osf.io/v8t4k/.

Prior to running the analyses, the data was assessed for potential collinearities among the predictors of interest, including the row and starting column of the target word in the 12 × 12 grid, starting column and word length, bigram and word frequency, and semantic diversity and word frequency. All predictors yielded low variance inflation factors (VIF = 1.25 or less), revealing no problematic collinearities in the data set.

To determine the optimal random effect structures of the logistic mixed effects models run on the word detection data, several versions of the models were attempted. The addition of row (how close to the top vs. the bottom of the board the target word occurs in) resulted in several failed model convergences and, hence, was not included as a random effect in subsequent analyses. This decision was validated by the fact that row was not a significant predictor of word finding (p = .83), beyond random effects associated with participants and items. By contrast, starting column (how close to the right or left side of the board the target word starts in) did significantly influence performance: Words that began closer to the left side of the board were easier for participants to find (Table 2). There was no significant interaction between row and starting column. Versions of the model that also included by-participant random slopes were attempted; however, they failed to converge. The final models thus included starting column, participant, and items as random intercepts, the maximal random effects structure for the models to converge (Barr et al., 2013). Each predictor was run as a single fixed effect in its own separate model to establish its unique contribution to word finding. Models where interactions or additional control variables were included are noted below.

Participants on average scored 36% correct in finding the target word across all trials (SD = 14%), including time-outs where the trial ended before the word was found (which accounted for 27% of all trials). As hypothesized, significant effects of both categorical frequency (high vs. low; Table 2) and log frequency were observed (Table 3), with participants exhibiting enhanced word finding for high-frequency words over low-frequency words (Figure 2). These

Table 2
Mean Accuracy and Reaction Times per Item Type

Word frequency	Accuracy (% correct)		Reaction time (s)	
	M (*SD*)	Range	*M* (*SD*)	Range
High	41% (16)	10%–80%	14.08 (2.85)	8.68–19.99
Low	31% (15)	7%–70%	15.01 (2.69)	8.44–19.99

effects remained unchanged by further controlling for semantic diversity ($\beta = .16$, $SE = .05$, $z = 2.91$, $p = .004$). Furthermore, average reaction times (excluding the trials where participants timed out) were significantly faster for the high-frequency words, $t(72) = -5.84$, $p < .0001$, $d = .33$ (Table 2).

Counter to our preregistered predictions, log bigram frequency was not a significant predictor of word finding. The interaction between log bigram frequency and word frequency was marginal (for both categorical and log word frequency: both $p = .06$). The number of syllables the target word contained did not influence word finding. However, the number of letters did, although in the opposite direction than anticipated: Longer words were easier to find.

To further explore the null effect of bigram frequency, we conducted an exploratory analysis on the interplay between bigram frequency and word position, which was not preregistered. This was inspired by data highlighting that the presence of high-frequency bigrams at morpheme boundaries can impede recognition of the individual constituents that compose compound words, reflected in longer reaction times in lexical decision tasks (Gagné et al., 2019). Although we did not use any compound words in our task, it is still possible that, rather than cumulative bigram frequency, the frequency of bigrams in specific positions in words across the corpus may influence word finding. To this end, we extracted the "bigram frequency by position" estimates for our items from ELP, which measures the summed bigram frequencies of each word by their position. For example, according to ELP, "the bigram frequency for DO in DOG counts DO bigrams only when they appear in the first two positions of a word" (Balota et al., 2007, p. 450). However, the bigram frequency by position estimates also had no significant effect on word finding accuracy ($p = .10$).

As expected, concreteness and semantic neighborhood density did not impact word finding. However, contrary to our predictions, there was a significant effect of semantic diversity when controlling for log frequency, indicating that words that occur in more disparate contexts facilitate word finding (Figure 3). There was no significant interaction between semantic diversity and either categorical or log word frequency ($p = .33$ or greater).

Table 3
Results From Individual Models: Predictors of Word Search Performance

Fixed effect	β	*SE*	*z*	*p*
Starting column	−.15	.04	−3.60	.0003**
Frequency (high vs. low)	−.66	.20	−3.28	.0012**
Log frequency	.16	.05	2.92	.004**
Log bigram frequency	−.04	.18	−.20	.84
Number of syllables	.08	.16	.50	.61
Number of letters	.27	.08	3.15	.002**
Concreteness	.10	.12	.93	.35
Semantic diversity	.72	.36	2.02	.04*
Semantic neighborhood density	1.71	1.04	1.63	.10

Note. *SE* = standard error.
* $p < .05$. ** $p < .01$.

Discussion

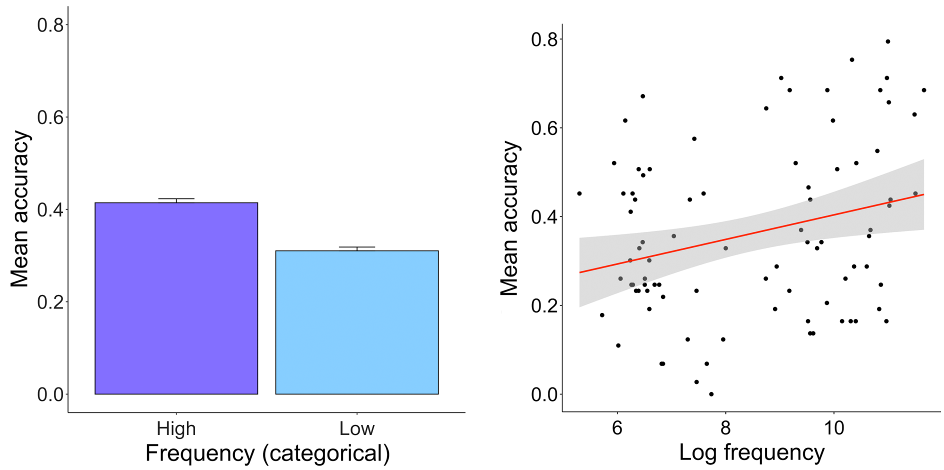
Experiment 1 assessed whether word search performance serves as a proxy for testing individual differences in sensitivity to statistical information in natural text. We tested the impact of several statistical properties on linguistic processing: word frequency, bigram frequency, and semantic diversity. As controls, we also included the number of letters, number of syllables, concreteness, and semantic neighborhood density.

In line with our preregistered predictions, the results reveal significant effects of both categorical and log frequencies on word finding, when accounting for the influence of participant, item, starting position, and semantic diversity. Reaction times were also faster for the high-frequency items, replicating classic findings in the psycholinguistic literature using a novel task (Forster & Chambers, 1973). Contrary to our predictions, there was no effect of bigram frequency. This was potentially due to the fact that there was relatively little variation in bigram frequency in our selected items to ensure that word and bigram frequency were not conflated. The bigram frequencies of the background distractors and the words were approximately equal to one another, meaning that bigram strength alone was not a reliable cue to word finding. However, the lack of bigram effects is consistent with published lexical decision data in large-scale databases, both those reported in ELP (Balota et al., 2007) and those reported in the British–English version of the corpus, the British Lexicon Project (Keuleers et al., 2012). However, higher bigram frequency does promote faster response times in reading aloud tasks in these databases, which suggests that bigram frequency may have stronger effects on phonological processing than orthographic processing (Schmalz & Mulatti, 2017).

The number of syllables did not affect word finding, although the length of the word as measured by the number of letters was a significant predictor, with longer words being easier to find. It is possible that these words stood out among the competitors more than shorter words. Although findings on the relationship between word length and lexical decision performance reveal a nonlinear, U-shaped curve rather than a linear relationship (New et al., 2006), the effect is monotonic for words between three and eight letters in length (the downturn is only for words between eight and 13 letters). Thus, the faster reaction times for longer words that we observed are consistent with prior work given the limited range of word lengths (New et al., 2006). There was no effect of concreteness, as this was relatively constant across the words in the stimulus set, or semantic neighborhood density, although there was a significant effect of semantic diversity. This effect remained significant when controlling for log word frequency, suggesting that semantic diversity imparted its own unique variance on word finding (in line with its effects in other tasks, e.g., Hoffman & Woollams, 2015; Hsiao & Nation, 2018).

Notably, overall accuracy was widely variable across individuals. Although mean performance across both high- and low-frequency items hovered around 36%, some participants scored as low as 9% while others scored as high as 80%. This wide range in performance suggests that the word search task may be highly sensitive to

Figure 2
Mean Accuracy by Categorical and Log Frequency



Note. Both higher categorical frequency and log frequency resulted in better word finding. Error bars in the bar chart reflect standard error. The gray band in the scatterplot reflects the 95% confidence interval. See the online article for the color version of this figure.

individual differences. Based on this observation, we investigated whether these differences were indicative of individual variation in component skills of reading proficiency that similarly rely on sensitivity to orthographic regularities.

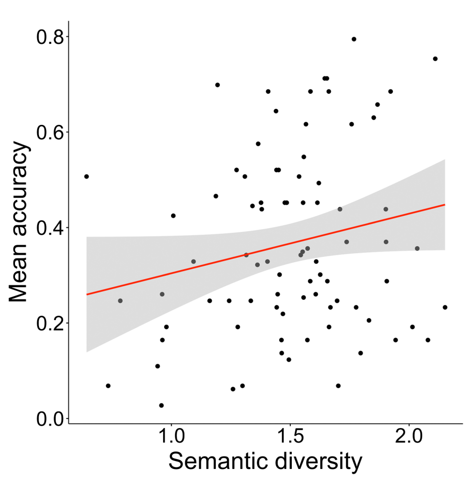
Experiment 2: Individual Differences in Reading Proficiency

Method

Participants

The 73 participants who completed the word search task were recontacted via Prolific to take part in the reading tasks, approximately 3 weeks after their completion of the original study. Of these, 11 did not return, yielding a final sample size of 62 participants. Participants were compensated monetarily at the same rate as Experiment 1 ($8.00/hr, as recommended by Prolific at the time of the study).

Figure 3
Mean Accuracy as a Function of Semantic Diversity



Note. Greater semantic diversity resulted in better word finding. The gray band reflects the 95% confidence interval. See the online article for the color version of this figure.

Materials

We selected three established measures of component skills of reading proficiency: lexical decision (using the Lexical Test for Advanced Learners of English; LexTALE; Lemhöfer & Broersma, 2012), Spelling Recognition, and Orthographic Awareness. These three tasks were included in the recent English Reading Online (ENRO) project (Siegelman et al., 2024; see details below). Major considerations behind the choice of these tests were that (a) they all can be administered online in a reliable and valid manner and (b) they are considered to capture variation in component skills that predict substantial variance in overall measures of reading skill. Indeed, data of over 7,000 adult participants from the ENRO project point to their high reliability and strong correlations with measures of reading comprehension and fluency in both monolingual and bilingual populations (Siegelman et al., 2024). Further, we selected these component skills because they can be thought of as drawing upon statistical knowledge about the writing system and knowledge of letter combinations in particular.

LexTALE Lexical Decision Task. This task measures written word identification skills. The stimuli for this task consisted of 60 items: 40 words and 20 nonwords, taken from Lemhöfer and Broersma (2012). The words were between four and 12 letters long (M = 7.3), with their frequencies ranging between 1 and 26 occurrences per million (M = 6.4) as estimated by the CELEX database (Baayen et al., 1995), a lexical database of English, Dutch and German. The nonwords were orthotactically legal and pronounceable and were created by the original authors by recombining existing morphemes (e.g., *plaudate*) or changing the number of letters in an existing English word (e.g., *skave*). The full list of items is available online from the original authors (https://www.lextale.com) and is reported again here in Appendix C.

ENRO Spelling Recognition. This task measures spelling abilities. A total of 44 words (22 correctly spelled, 22 misspelled) were used, taken from Siegelman et al. (2024; originally adapted from Andrews & Hersch, 2010). The full list of items is reported in Appendix D.

ENRO Orthographic Awareness. This task measures sensitivity to orthographic regularities. A total of 30 pairs of nonwords were used, taken from Siegelman et al. (2024), based on Siegel et al. (1995). One word of each pair contained a bigram that never occurs in English in a certain position (either word initial or word final). The other member of the pair contained an orthotactically legal bigram in the same position (e.g., *blaem* vs. *bleam*). The full list of items is reported in Appendix E.

Procedure

All tasks were administered in the same order across participants to minimize any potential order effects on performance (Mollon et al., 2017). However, the presentation of the items within each task was randomized across participants. In line with the original studies on which these tasks are based, the reading tasks were not timed. The time to complete all three tasks was 8 min on average.

Word Search Habits Questionnaire. Prior to beginning the experiment, participants were given a brief survey about their current and past word search game playing habits. Participants were asked to rate on a 1–7 scale how often they played word search games as a child (1 = *never*, 2 = *at least once a year*, 3 = *several times a year*, 4 = *at least once a month*, 5 = *several times a month*, 6 = *at least every week*, 7 = *every day*), as well as how often they currently play word search games as adults. This questionnaire took 1 min on average to complete.

LexTALE Lexical Decision Task. Participants were presented with words and nonwords (e.g., *ablaze*, *abergy*) presented on the screen one at a time, and asked to judge via a Yes/No button press whether a presented word was a real English word. Participants were told that all items reflected American English spelling.

ENRO Spelling Recognition. Participants were presented with words that were either correctly or incorrectly spelled (e.g., *accommodation*, *addmission*). Words were presented one at a time, and participants indicated via a Yes/No button press whether the presented word was spelled correctly. Participants were once more informed that all items reflected American English spelling.

ENRO Orthographic Awareness. Each trial presented two pairs of nonwords (e.g., *bolk* and *bolz*). Participants were instructed to select which of the two words they thought could spell a valid English word.

Results

The results of the word search habits questionnaire revealed that most participants played word search games at least once a month as children (M = 4.52, SD = 1.29, range = 2–7) and currently play several times a year as adults (M = 3.39, SD = 1.62, range = 1–7). Neither childhood nor current word search playing habits correlated with experimental word search performance (both p = .42 or greater).

Participants performed significantly above chance (50%) on each reading subtest—LexTALE Lexical Decision Task (LDT): $t(61)$ = 51.21, p < .0001, d = 6.50; Spelling Recognition: $t(61)$ = 31.48, p < .0001, d = 4.00; Orthographic Awareness: $t(61)$ = 39.65, p < .0001, d = 5.04. The means, ranges, and standard deviations for each measure are reported in Table 4.

Furthermore, word search score significantly predicted each of the reading measures, with better word finding correlating with higher LDT, $r(60)$ = .36, p = .004; Spelling Recognition, $r(60)$ = .37, p = .003; and Orthographic Awareness scores, $r(60)$ = .28, p = .03 (Figure 4). Lexical decision was significantly correlated with each of the other reading measures—LDT and Spelling Recognition: $r(60)$ = .60, p < .0001; LDT and Orthographic Awareness: $r(60)$ = .28, p = .03. However, Spelling Recognition and Orthographic Awareness performance were not significantly correlated, $r(60)$ = .21, p = .09. All between-task correlations are reported in Table 5.

In addition, word search accuracy was correlated with lexical decision reaction times: Participants who performed better on the word search task were also significantly faster to accept words and reject nonwords on the LDT, $r(60)$ = −.31, p = .01. However, there was no correlation between word search performance and reaction times on either the Spelling Recognition or Orthographic Awareness tasks (p = .34 or above).

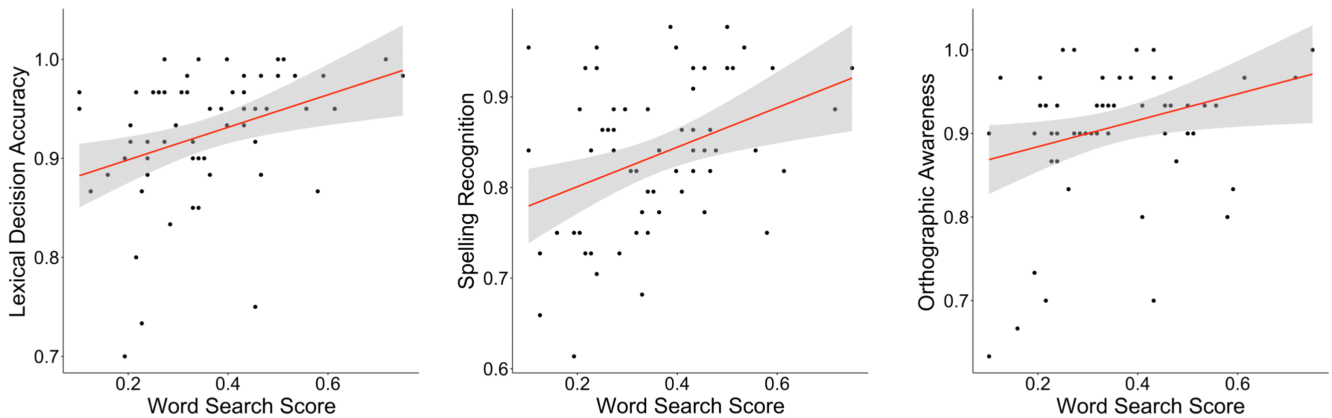
To generate a measure of overall reading proficiency, we created a composite score of the three reading measures (Siegelman et al., 2024). This was achieved by computing a z-score on each measure (LDT, Orthographic Awareness, and Spelling Recognition) for each participant and then calculating the average of these three scores for each participant. Word search performance was then used to predict this composite reading score. The results reveal that word search score was significantly correlated with reading proficiency, $r(60)$ = .44, p = .0003, with better readers finding significantly more words (Figure 5). A multiple regression analysis revealed that none of the subcomponent reading measures predicted word search more strongly than the other (all p = .12 or above), suggesting that word search performance is related to shared variance between these three measures rather than to one of these component skills specifically (see the Supplemental Materials for further analyses).

To investigate whether our word frequency measures differentially impacted word search performance for readers of varied skill levels, we performed several post hoc exploratory analyses, which were not preregistered. These analyses were motivated by prior research showing that differences in vocabulary size (large vs. small; Preston, 1935) can lead to different frequency effects across individuals in lexical decision tasks (Brysbaert et al., 2017; Keuleers et al., 2010; see Monaghan et al., 2017, for computational simulations of this behavioral data and Brysbaert et al., 2018, for a review). While our language measures did not include vocabulary size, it is possible that

Table 4
Summary Statistics for Each Reading Measure

Reading task	M (SD)	Range
Lexical decision task	92% (7)	70%–100%
Spelling recognition	84% (8)	61%–98%
Orthographic awareness	91% (8)	63%–100%

Figure 4
Correlations Between Word Search and Lexical Decision, Spelling Recognition, and Orthographic Awareness Performance



Note. Participants who scored higher on the word search task also displayed larger scores on each reading subtest. The gray bands reflect the 95% confidence intervals. See the online article for the color version of this figure.

differential frequency effects may also arise for readers of different skill levels in our word search task as well.

For these exploratory analyses, we performed a median split on the data to classify participants as high- or low-proficiency readers, similar to the median split performed in the studies cited above to classify individuals as having large or small vocabularies. Participants were classified as high-proficiency readers if their composite reading score was equal to or greater than the median (.16). Those who fell below the median were classified as low-proficiency readers. Individual models were then run for each variable of interest. The results revealed that both high- and low-proficiency readers were significantly impacted by categorical and log word frequency: High-frequency words were on average easier for participants to find regardless of reading skill, when controlling for participant, item, and starting column. However, semantic diversity only impacted word finding in high-proficiency readers, when additionally controlling for log word frequency. The results of these exploratory analyses are reported in Table 6.

Discussion

Experiment 2 sought to determine whether the individual differences observed in the word search task of Experiment 1 were reflective of individual differences in reading. We found that word search performance was a significant predictor of three component skills of reading, including lexical decision, orthographic awareness, and spelling recognition, suggesting that the word search task and these component skills similarly rely upon sensitivity to orthographic regularities.

Importantly, word search performance was not significantly predicted by word search habits (how frequently participants played word search games as children and how often they play now). Part of this may stem from differences in task demands: In traditional word search games, participants are given a list of words and must find the matching item in the array. They are also untimed, meaning that participants are free to deploy a variety of search strategies that may be less related to the implicit statistical computations that reading relies upon. Performance on the experimental word search task is thus not influenced by the frequency with which individuals play such games, but rather by general reading ability. This argument is augmented by the fact that the composite reading score based on the average *z*-scores of each standard reading measure significantly predicts word search ability, suggesting that these tasks may draw upon similar computations.

Furthermore, our exploratory analyses revealed that readers of all skill levels were significantly impacted by both categorical and log word frequencies on the search task, although only higher proficiency readers showed the semantic diversity effect reported in Experiment 1. However, since these analyses were post hoc, and other studies of this nature employ a much larger battery of items and range of frequencies to probe this question (e.g., 420 words in Brysbaert et al., 2017, vs. 88 words in our word search task), further research is required to examine this relationship.

Table 5
Correlation Matrix: Word Search and Reading Scores

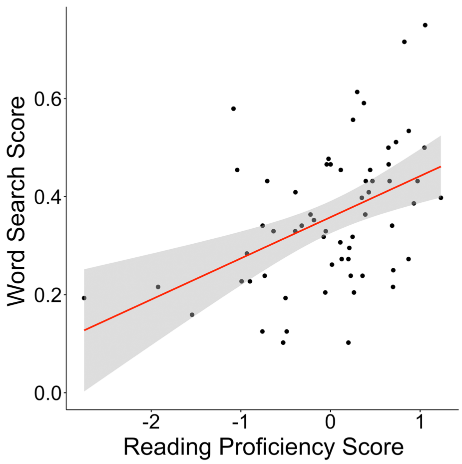
Reading task	Word search	Lexical decision task	Spelling recognition	Orthographic awareness
1. Lexical decision task	.36**	—		
2. Spelling recognition	.37**	.60***	—	
3. Orthographic awareness	.28**	.28**	.21	—

** $p < .01$. *** $p < .0001$.

General Discussion

Over the last 25 years, mounting research has sought to forge a link between statistical learning and reading. This work was motivated by the hypothesis that statistical learning is a fundamental gateway for enabling the acquisition of skilled reading and in turn the implementation of interventions that could potentially alleviate delays in achieving reading milestones in children with reading difficulties. While these studies of the linkage between statistical

Figure 5
Correlation Between Word Search Performance and the Composite Reading Proficiency Score



Note. Participants who exhibited greater reading proficiency as measured by the composite reading score also demonstrated enhanced word finding. The gray band reflects the 95% confidence interval. See the online article for the color version of this figure.

learning and reading have significantly enhanced our understanding of this connection by focusing on the acquisition of visual-nonlinguistic statistics and comparing performance to a battery of reading measures (see Lee et al., 2022; Ren et al., 2023, for meta-analyses), they have also generated mixed results. Here, we explored the idea that these mixed results may in part derive from two often-implicit assumptions of statistical learning research. The first is that statistical learning encompasses a single unitary mechanism that is activated regardless of task demands, the type of statistical regularities, and cognitive domains, implying that learners should demonstrate equal proficiency in any task involving statistical information (Bogaerts et al., 2022; R. Frost et al., 2015). The second is that learning assessed in a short exposure-test session provides a reliable measure of individual differences for tasks like reading that emerge over years of exposure to complex linguistic material.

Here, we took an initial step in addressing these assumptions through the development of an alternative to traditional visual statistical learning tasks. Unlike prior tasks that utilize artificial visual structures such as novel shapes or cartoon characters (Arciuli & Simpson, 2012; Fiser & Aslin, 2002), we examined how the assimilation of natural language orthographic statistics across a lifetime of experience predicts variation in reading. Our word search task provides a direct measure of individual differences in statistical learning by tapping into participants' sensitivity to the orthographic regularities present in natural text—information that forms the foundation of literacy (Chetail, 2015, 2017; Treiman & Kessler, 2022). This task can thus be conceptualized as a measure of long-term orthographic statistical learning that taps into the on-line processing of text and is correlated with established measures of reading proficiency.

Our results reveal that word search performance was critically impacted by word frequency, in line with a long history of psycholinguistic research that confirms the role of this first-order statistical property of linguistic input at the lexical level (e.g., Brysbaert et al., 2016; Forster & Chambers, 1973; Monsell et al., 1989; Preston, 1935). Participants were both faster and more accurate at identifying high-frequency words than low-frequency words embedded in visual noise comprising letters that do not form words, suggesting facilitation from long-term distributional learning.

In contrast to the effect of word frequency on search performance, there was no significant influence of word-internal bigram frequency, counter to our preregistered hypotheses. This finding may stem from the fact that the average bigram frequencies of the words were equated with the bigram frequencies of the background distractors, meaning that bigram frequency was not as strong of a cue to word locations as whole-word frequency. The bigram frequencies of the target words were also well-controlled to avoid conflating this measure with word frequency, meaning that only a very narrow range of bigram frequencies were built into the target words. However, it is worth noting that findings on bigram effects vary throughout the literature. For example, several large-scale analyses find little evidence of bigram effects in lexical decision (Balota et al., 2007; Keuleers et al., 2010), although higher bigram frequencies do have a facilitatory effect in reading aloud tasks (Schmalz & Mulatti, 2017). Importantly, bigram frequency effects vary along many dimensions beyond the scope of the current experiment's design. For example, where in the word frequent bigrams occur significantly affects lexical decision performance (e.g., the beginning vs. end, or at morpheme boundaries; Gagné et al., 2019), which in turn interacts with whole-word frequency (Chetail, 2015, 2017). Furthermore, the frequency of bigrams at word boundaries that arise from the combination of the letters that precede and follow the target word in the word search task may also influence processing: For example, in Figure 1, "molar" is preceded by "g" and followed by "u," and the frequency of the bigrams "gm" and "ru" may make it more or less difficult to segment the target word. These factors can be systematically manipulated in further experiments utilizing this paradigm aimed at exploring sensitivity to sublexical letter cluster frequency.

In addition to the effects of text-based sensitivity, the results reveal a unique contribution of semantic diversity on word finding independent of word frequency, and vice versa. Words that exhibit greater semantic diversity—that occur in a larger number of distinct contexts—were easier for participants to find. This result is noteworthy, as the explanatory power of these variables has been the topic of considerable debate in the word recognition literature. Prior research demonstrates that word frequency and semantic diversity often interact (Cevoli et al., 2021) and are highly correlated ($r > .95$;

Table 6
Individual Model Results of Word Search Performance by Reader Type

Reader type	Fixed effect	β	*SE*	*z*	*p*
High proficiency	Frequency (high vs. low)	−.55	.21	−2.61	.009**
	Log frequency	.13	.06	2.35	.02*
	Semantic diversity	.81	.36	2.24	.03*
Low proficiency	Frequency (high vs. low)	−.73	.21	−3.47	.0005**
	Log frequency	.17	.06	3.08	.002**
	Semantic diversity	.55	.38	1.46	.15

Note. *SE* = standard error.
* $p < .05$. ** $p < .01$.

Hoffman et al., 2013), making it difficult to disentangle the two (note that the correlation was $r = .43$ in our data set). In fact, some research shows that semantic diversity can sometimes wipe out frequency effects all together, suggesting that frequency may serve as a proxy for the contextual diversity of words (Adelman et al., 2006; Perea et al., 2013). On the other hand, others report that frequency, but not semantic diversity, impacts word learning in children and thus may be the more formative of the two variables (Joseph & Nation, 2018).

To address this paradox, a recent study based on five large-scale databases on word reading and lexical decision shows that, on the whole, semantic diversity and frequency effects are largely uncorrelated and therefore represent unique constructs (Chapman & Martin, 2022). The authors further detail that controlling for semantic diversity did not diminish the size of frequency effects, identical to our own data. This reinforces the hypothesis that semantic diversity and word frequency capture different facets of reading, which differentially impact lexical processing. Semantic diversity may better estimate the amount individuals read, evidenced by exposure to a greater number of semantic contexts, while frequency evaluates word-level statistical sensitivity. This idea is bolstered by our exploratory analyses showing that semantic diversity effects are stronger in skilled readers who presumably read more (although see below for caveats surrounding these post hoc analyses). Indeed, considerable debate exists about whether semantic diversity is an undiluted semantic measure or if it in part serves as a proxy for contextual diversity—the number of documents a word appears in (Adelman et al., 2006). Contextual diversity has been found to better correlate with LDT reaction times in some reports than word frequency, perhaps in part because it provides a better estimate of the probability of encountering words in a given context (Norris, 2009). Relatedly, Cevoli et al. (2021) stated that the term semantic diversity (as described by Hoffman et al., 2013) may be better branded as "textual diversity," since it measures the distribution of a word both across types of topics and types of contexts.

Our results also support our hypothesis that word search performance should be highly predictive of performance on three established component skills of reading: lexical decision, orthographic awareness, and spelling recognition. All of these tasks capture participants' sensitivity to orthographic regularities, which suggests that word search may similarly leverage long-term orthographic knowledge and computations relevant for processing this information. As introduced earlier, increasing research supports the notion that statistical learning may comprise an assemblage of mechanisms that vary according to the input, evidenced by uneven learning profiles across different types of stimuli (e.g., Arnon, 2020; Siegelman & Frost, 2015). Such diversity of effects also manifests in research that bridges statistical learning and language proficiency. Recent longitudinal work highlights the independent contributions of auditory and visual statistical learning to specific facets of language development in children (Kidd et al., 2023). Auditory-verbal statistical learning is more highly correlated with sentence production, since both rely on the sequencing of syllable transitions, whereas visual-nonverbal statistical learning is more closely associated with language comprehension tasks involving visually presented images (see also Kidd & Arciuli, 2016; Perfors & Kidd, 2022; and see Qi et al., 2019, for evidence and discussion in the context of predicting reading skills). Such observations have sparked a call for work that makes stronger predictions about the precise computations shared (or not) between tasks (Bogaerts et al., 2022). Aligning the information targeted by statistical learning and natural language tasks can lead to more robust correlations because it better differentiates between domain-specific and domain-general (i.e., shared) cognitive processes (Isbilen, McCauley, & Christiansen, 2022).

Returning to the multicomponential view of statistical learning, here we argue for the involvement of statistically based chunking as the relevant subcomponent of orthographic statistical learning that predicts general reading outcomes. Statistically based chunking is integral to numerous aspects of language acquisition and processing. It operates across linguistic levels, with recurring phoneme combinations leading to the discovery of syllables in the auditory modality (Perruchet & Pacton, 2006; Perruchet & Vinter, 1998), which leads to the formation of words and multiword units that scaffold grammatical development (Ambridge & Lieven, 2015; Christiansen & Chater, 2016; McCauley & Christiansen, 2019). Comparable computations are also at play in orthographic processing, with the chunking of commonly recurring letters enabling the discovery of complex letter clusters (e.g., *str: string*, *strange*, *street*, *etc.*; Grainger & Ziegler, 2011) and larger morphological units (Lelonkiewicz et al., 2020, 2023), which in turn support lexical processing when reading. Chunking-based computational models also better approximate human statistical learning performance than nonchunking models that rely on probability computation alone (Frank et al., 2010; French et al., 2011; see Orbán et al., 2008, for chunking models of visual-nonlinguistic pattern learning), including segmentation in the classic Saffran et al. (1996) statistical learning task (Perruchet & Vinter, 1998). This may in part explain the predictive power of statistical learning to other linguistic processes. Sensitivity to high-frequency sublexical phonological chunks guides individual performance on nonword repetition, a classic task of phonological awareness used to diagnose language and reading impairment (Jones, 2016). Participants' sensitivity to high-frequency chunks in written text also predicts reading speed and efficiency in both their native language (Bonhage et al., 2017; McCauley et al., 2017) and second language (Pulido, 2021), further underscoring the central role of chunking to higher order linguistic processes.

The word search task that we designed fundamentally relies on the chunking of orthographic statistics. Successful word search performance requires individuals to leverage their knowledge of accrued regularities to chunk whole words from continuous text, similar to how knowledge of phonological regularities helps individuals delineate word boundaries in continuous speech (Saffran, 2001). The ability of text-based chunking to predict reading proficiency is consistent with findings from the auditory processing literature showing that individuals who perform better on recall-based verbal statistical learning tasks exhibit more robust perception of auditory words in noise while controlling for working and short-term memory, attention, vocabulary, and general intelligence (Conway et al., 2010). This is because statistical learning is crucial to amassing the long-term sequential knowledge that underlies word predictability and, in turn, the capacity to use this knowledge during degraded listening conditions. Further support for the involvement of statistical chunking in our word search task comes from the null effects of semantics on word finding, including semantic neighborhood density and concreteness (with the exception of semantic diversity, which indexes the occurrence of items in distinct semantic contexts and therefore embodies a type of co-occurrence statistic). While these factors critically impact processing on lexical decision and word naming tasks (e.g., Hendrix & Sun, 2021; Reilly & Desai, 2017), the pronounced influence of the frequency measures suggests that the word search task successfully leveraged the cognitive processes it aimed to target. Furthermore, word search simulates the spatial aspects of reading, requiring the grouping of contiguous information in complex arrays of text, whereas most visual statistical learning studies focus on the acquisition of temporal regularities, where individuals are trained on sequences of shapes presented one after the other in the absence of background clutter. It is possible that our word search task better approximates the computations used for reading, as evidenced by the reaction time differences between high- and low-frequency words, an effect that is observed for word reading but not other visual search tasks (Rayner & Raney, 1996).

Our final set of exploratory analyses was inspired by classic findings by Preston (1935), who reported larger frequency effects for individuals with smaller vocabularies. Similar results have since been replicated in both first (Davies et al., 2017; Kuperman & Van Dyke, 2013) and second language learners (Cop et al., 2015). Regardless of vocabulary size, all individuals show faster and more accurate processing of high-frequency words because such items tend to be overlearned, but for individuals with smaller vocabularies, their processing of very low-frequency words suffers more due to minimal exposure (Brysbaert et al., 2018; Monaghan et al., 2017). Based on this effect, we examined whether our main variables of interest, frequency and semantic diversity, varied in how they impacted readers of different skill levels.

These exploratory analyses revealed that readers, regardless of skill level, were significantly impacted by the word frequency manipulations in the word search task. By contrast, semantic diversity only affected high-proficiency readers, suggesting that a certain threshold level of proficiency (or reading experience) may be required for the effects of semantic diversity to emerge. These findings diverge somewhat from prior observations in the literature, which show stronger frequency effects in individuals with smaller vocabularies, although we did not measure vocabulary size specifically. It is possible that the task demands of our word search task, which is more challenging and yields larger individual differences than our own lexical decision task, may show more pronounced frequency effects in general. Prior research also used a much larger set of items (e.g., 420 words in Brysbaert et al., 2017, vs. 88 in our word search task), which affords more datapoints per participant and a much broader range of frequencies. It also provides a better estimate of whether readers of different proficiency levels are impacted by frequency to the same degree. Although these post hoc analyses targeted questions that were not factored into our original study design, they may still serve as a useful springboard for future research, especially studies devoted to determining how vocabulary size relates to statistical learning and reading.

Tasks that tap into the statistics relevant to reading may provide valuable insights into language processing beyond those offered by artificial language paradigms. First, such tasks may be more ecological than artificial language stimuli, which differ dramatically from natural text and the strategies required for real-world reading (although note that one limitation of the present study is that segmenting words from continuous text is not a process required for reading English, where words are separated by spaces, although it is a process required for reading languages like Chinese, where readers must segment words in sentences). Second, they circumvent the domain-specificity issue introduced above, which may contribute to the inconsistent correlations observed between visual statistical learning and reading across the literature. However, future experimentation comparing the efficacy of word search and traditional visual statistical learning tasks in predicting reading is required to substantiate this possibility. Additional future directions include the use of the word search task with eye tracking to evaluate whether individuals of differing reading skill levels are uniquely impacted by the statistics of the background distractors or bigram information that spans word boundaries, as findings by Olivers et al. (2014) suggest that low-proficiency readers are slower to select relevant information in nonlinguistic visual search tasks. It may also prove to be an effective method for use with children, who may benefit from the gamified aspect of the task. Furthermore, the words embedded in the search-array boards can be manipulated for other psycholinguistic factors, such as imageability and age of acquisition, which may provide a wellspring of data for future individual differences research. Evaluating how word search correlates with measures of central component skills of reading such as rapid automatized naming and phonological awareness—tasks that are among the strongest predictors of literacy (Clayton et al., 2020; S. J. Frost et al., 2009; Powell & Atkinson, 2021)—can also help elucidate the unique contributions of and interactions between sensitivity to phonological and orthographic regularities. Finally, although the present study cannot speak to the directionality of the connection between statistical learning and reading—whether better statistical learners become better readers or whether individuals with greater amounts of print exposure become more sensitive to statistical regularities—a combination of measures and longitudinal studies may be better able to elucidate this question.

In conclusion, the present study sought to bridge individual differences in statistical learning to variation in reading proficiency. Rather than deploying an artificial statistical learning task, we focused on how sensitivity to the statistical information ingrained in natural text supports literacy. We replicated classic psycholinguistic frequency effects in a novel word search paradigm that relied upon the ability to chunk statistical information in text. Performance on this task was highly predictive of key component skills of reading proficiency, suggesting a role for statistically based chunking in reading development.

References

Adelman, J. S., Brown, G. D., & Quesada, J. F. (2006). Contextual diversity, not word frequency, determines word-naming and lexical decision times. *Psychological Science*, *17*(9), 814–823. https://doi.org/10.1111/j.1467-9280.2006.01787.x

Ambridge, B., & Lieven, E. (2015). A constructivist account of child language acquisition. In B. MacWhinney, & W. O'Grady (Eds.), *The handbook of language emergence* (pp. 478–510). Wiley.

Andrews, S., & Hersch, J. (2010). Lexical precision in skilled readers: Individual differences in masked neighbor priming. *Journal of Experimental Psychology: General*, *139*(2), 299–318. https://doi.org/10.1037/a0018366

Arciuli, J., & Simpson, I. C. (2012). Statistical learning is related to reading ability in children and adults. *Cognitive Science*, *36*(2), 286–304. https://doi.org/10.1111/j.1551-6709.2011.01200.x

Arnon, I. (2019). Statistical learning, implicit learning, and first language acquisition: A critical evaluation of two developmental predictions. *Topics in Cognitive Science*, *11*(3), 504–519. https://doi.org/10.1111/tops.12428

Arnon, I. (2020). Do current statistical learning tasks capture stable individual differences in children? An investigation of task reliability across modality. *Behavior Research Methods*, *52*(1), 68–81. https://doi.org/10.3758/s13428-019-01205-5

Baayen, R. H., Piepenbrock, R., & Gulikers, L. (1995). The CELEX lexical database (release 2). *Distributed by the linguistic data consortium*. University of Pennsylvania.

Baker, C. I., Olson, C. R., & Behrmann, M. (2004). Role of attention and perceptual grouping in visual statistical learning. *Psychological Science*, *15*(7), 460–466. https://doi.org/10.1111/j.0956-7976.2004.00702.x

Balota, D. A., Yap, M. J., Cortese, M. J., Hutchison, K. A., 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

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278. https://doi.org/10.1016/j.jml.2012.11.001

Bates, D., Mächler, M., Bolker, B., Walker, S., Christensen, R. H., Singmann, H., Dai, B., Scheipl, F., Grothendieck, G., Green, P., Fox, J., Bauer, A., & Krivitsky, P. N. (2020). *Package "lme4"* (Version 1.1-26). R Package. https://cran.r-project.org/web/packages/lme4/lme4.pdf

Bogaerts, L., Siegelman, N., Christiansen, M. H., & Frost, R. (2022). Is there such a thing as a 'good statistical learner'? *Trends in Cognitive Sciences*, *26*(1), 25–37. https://doi.org/10.1016/j.tics.2021.10.012

Bonhage, C. E., Meyer, L., Gruber, T., Friederici, A. D., & Mueller, J. L. (2017). Oscillatory EEG dynamics underlying automatic chunking during sentence processing. *NeuroImage*, *152*, 647–657. https://doi.org/10.1016/j.neuroimage.2017.03.018

Brysbaert, M., Lagrou, E., & Stevens, M. (2017). Visual word recognition in a second language: A test of the lexical entrenchment hypothesis with lexical decision times. *Bilingualism: Language and Cognition*, *20*(3), 530–548. https://doi.org/10.1017/S1366728916000353

Brysbaert, M., Mandera, P., & Keuleers, E. (2018). The word frequency effect in word processing: An updated review. *Current Directions in Psychological Science*, *27*(1), 45–50. https://doi.org/10.1177/0963721417727521

Brysbaert, M., Stevens, M., Mandera, P., & Keuleers, E. (2016). The impact of word prevalence on lexical decision times: Evidence from the Dutch Lexicon Project 2. *Journal of Experimental Psychology: Human Perception and Performance*, *42*(3), 441–458. https://doi.org/10.1037/xhp0000159

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

Cevoli, B., Watkins, C., & Rastle, K. (2021). What is semantic diversity and why does it facilitate visual word recognition? *Behavior Research Methods*, *53*(1), 247–263. https://doi.org/10.3758/s13428-020-01440-1

Champely, S., Ekstrom, C., Dalgaard, P., Gill, J., Weibelzahl, S., Anandkumar, A., Ford, C., Volcic, R., De Rosario, H., & De Rosario, M. H. (2018). *Package "pwr"* (R Package Version 1.2).

Chapman, C. A., & Martin, R. C. (2022). Effects of semantic diversity and word frequency on single word processing. *Journal of Experimental Psychology: General*, *151*(5), 1035–1068. https://doi.org/10.1037/xge0001123

Chen, L., & Vroomen, J. (2013). Intersensory binding across space and time: A tutorial review. *Attention, Perception, & Psychophysics*, *75*(5), 790–811. https://doi.org/10.3758/s13414-013-0475-4

Chetail, F. (2015). Reconsidering the role of orthographic redundancy in visual word recognition. *Frontiers in Psychology*, *6*, Article 645. https://doi.org/10.3389/fpsyg.2015.00645

Chetail, F. (2017). What do we do with what we learn? Statistical learning of orthographic regularities impacts written word processing. *Cognition*, *163*, 103–120. https://doi.org/10.1016/j.cognition.2017.02.015

Chomsky, N. (1959). On certain formal properties of grammars. *Information and Control*, *2*(2), 137–167. https://doi.org/10.1016/S0019-9958(59)90362-6

Christiansen, M. H., & Chater, N. (2016). The now-or-never bottleneck: A fundamental constraint on language. *Behavioral and Brain Sciences*, *39*, Article e62. https://doi.org/10.1017/S0140525X1500031X

Clayton, F. J., West, G., Sears, C., Hulme, C., & Lervåg, A. (2020). A longitudinal study of early reading development: Letter-sound knowledge, phoneme awareness and RAN, but not letter-sound integration, predict variations in reading development. *Scientific Studies of Reading*, *24*(2), 91–107. https://doi.org/10.1080/10888438.2019.1622546

Conway, C. M., Bauernschmidt, A., Huang, S. S., & Pisoni, D. B. (2010). Implicit statistical learning in language processing: Word predictability is the key. *Cognition*, *114*(3), 356–371. https://doi.org/10.1016/j.cognition.2009.10.009

Conway, C. M., & Christiansen, M. H. (2005). Modality-constrained statistical learning of tactile, visual, and auditory sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(1), 24–39. https://doi.org/10.1037/0278-7393.31.1.24

Cop, U., Keuleers, E., Drieghe, D., & Duyck, W. (2015). Frequency effects in monolingual and bilingual natural reading. *Psychonomic Bulletin & Review*, *22*(5), 1216–1234. https://doi.org/10.3758/s13423-015-0819-2

Crepaldi, D., Rastle, K., & Davis, C. J. (2010). Morphemes in their place: Evidence for position-specific identification of suffixes. *Memory & Cognition*, *38*(3), 312–321. https://doi.org/10.3758/MC.38.3.312

Davies, R. A. I., Arnell, R., Birchenough, J. M. H., Grimmond, D., & Houlson, S. (2017). Reading through the life span: Individual differences in psycholinguistic effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *43*(8), 1298–1338. https://doi.org/10.1037/xlm0000366

Eimas, P. D. (1985). The perception of speech in early infancy. *Scientific American*, *252*(1), 46–53. https://www.scientificamerican.com/article/the-perception-of-speech-in-early-i/

Elazar, A., Alhama, R. G., Bogaerts, L., Siegelman, N., Baus, C., & Frost, R. (2022). When the "Tabula" is anything but "Rasa:" What determines performance in the auditory statistical learning task? *Cognitive Science*, *46*(2), Article e13102. https://doi.org/10.1111/cogs.13102

Ferrante, O., Patacca, A., Di Caro, V., Della Libera, C., Santandrea, E., & Chelazzi, L. (2018). Altering spatial priority maps via statistical learning of target selection and distractor filtering. *Cortex*, *102*, 67–95. https://doi.org/10.1016/j.cortex.2017.09.027

Fiser, J., & Aslin, R. N. (2001). Unsupervised statistical learning of higher-order spatial structures from visual scenes. *Psychological Science*, *12*(6), 499–504. https://doi.org/10.1111/1467-9280.00392

Fiser, J., & Aslin, R. N. (2002). Statistical learning of higher-order temporal structure from visual shape sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*(3), 458–467. https://doi.org/10.1037/0278-7393.28.3.458

Forster, K. I., & Chambers, S. M. (1973). Lexical access and naming time. *Journal of Verbal Learning and Verbal Behavior*, *12*(6), 627–635. https://doi.org/10.1016/S0022-5371(73)80042-8

Franco, A., Cleeremans, A., & Destrebecqz, A. (2011). Statistical learning of two artificial languages presented successively: How conscious? *Frontiers in Psychology*, *2*, Article 229. https://doi.org/10.3389/fpsyg.2011.00229

Frank, M. C., Goldwater, S., Griffiths, T. L., & Tenenbaum, J. B. (2010). Modeling human performance in statistical word segmentation. *Cognition*, *117*(2), 107–125. https://doi.org/10.1016/j.cognition.2010.07.005

French, R. M., Addyman, C., & Mareschal, D. (2011). TRACX: A recognition-based connectionist framework for sequence segmentation and chunk extraction. *Psychological Review*, *118*(4), 614–636. https://doi.org/10.1037/a0025255

Frost, R., Armstrong, B. C., Siegelman, N., & Christiansen, M. H. (2015). Domain generality versus modality specificity: The paradox of statistical learning. *Trends in Cognitive Sciences*, *19*(3), 117–125. https://doi.org/10.1016/j.tics.2014.12.010

Frost, S. J., Landi, N., Mencl, W. E., Sandak, R., Fulbright, R. K., Tejada, E. T., Jacobsen, L., Grigorenko, E. L., Constable, R. T., & Pugh, K. R. (2009).

Phonological awareness predicts activation patterns for print and speech. *Annals of Dyslexia*, *59*(1), 78–97. https://doi.org/10.1007/s11881-009-0024-y

Gabay, Y., Thiessen, E. D., & Holt, L. L. (2015). Impaired statistical learning in developmental dyslexia. *Journal of Speech, Language, and Hearing Research*, *58*(3), 934–945. https://doi.org/10.1044/2015_JSLHR-L-14-0324

Gagné, C. L., Spalding, T. L., & Schmidtke, D. (2019). LADEC: The large database of English compounds. *Behavior Research Methods*, *51*, 2152–2179. https://doi.org/10.3758/s13428-019-01282-6

Gebhart, A. L., Aslin, R. N., & Newport, E. L. (2009). Changing structures in midstream: Learning along the statistical garden path. *Cognitive Science*, *33*(6), 1087–1116. https://doi.org/10.1111/j.1551-6709.2009.01041.x

Gerth, S., & Festman, J. (2021). Reading development, word length and frequency effects: An eye-tracking study with slow and fast readers. *Frontiers in Communication*, *6*, Article 743113. https://doi.org/10.3389/fcomm.2021.743113

Grainger, J., & Ziegler, J. C. (2011). A dual-route approach to orthographic processing. *Frontiers in Psychology*, *2*, Article 54. https://doi.org/10.3389/fpsyg.2011.00054

Growns, B., Siegelman, N., & Martire, K. A. (2020). The multi-faceted nature of visual statistical learning: Individual differences in learning conditional and distributional regularities across time and space. *Psychonomic Bulletin & Review*, *27*(6), 1291–1299. https://doi.org/10.3758/s13423-020-01781-0

Hall, M. G., Mattingley, J. B., & Dux, P. E. (2015). Distinct contributions of attention and working memory to visual statistical learning and ensemble processing. *Journal of Experimental Psychology: Human Perception and Performance*, *41*(4), 1112–1123. https://doi.org/10.1037/xhp0000069

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

He, X., & Tong, X. (2017). Statistical learning as a key to cracking Chinese orthographic codes. *Scientific Studies of Reading*, *21*(1), 60–75. https://doi.org/10.1080/10888438.2016.1243541

Hendrix, P., & Sun, C. C. (2021). A word or two about nonwords: Frequency, semantic neighborhood density, and orthography-to-semantics consistency effects for nonwords in the lexical decision task. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *47*(1), 157–183. https://doi.org/10.1037/xlm0000819

Hoffman, P., Lambon Ralph, M. A., & Rogers, T. T. (2013). Semantic diversity: A measure of semantic ambiguity based on variability in the contextual usage of words. *Behavior Research Methods*, *45*(3), 718–730. https://doi.org/10.3758/s13428-012-0278-x

Hoffman, P., & Woollams, A. M. (2015). Opposing effects of semantic diversity in lexical and semantic relatedness decisions. *Journal of Experimental Psychology: Human Perception and Performance*, *41*(2), 385–402. https://doi.org/10.1037/a0038995

Hsiao, Y., & Nation, K. (2018). Semantic diversity, frequency and the development of lexical quality in children's word reading. *Journal of Memory and Language*, *103*, 114–126. https://doi.org/10.1016/j.jml.2018.08.005

Isbilen, E. S., Frost, R. L. A., Monaghan, P., & Christiansen, M. H. (2022). Statistically based chunking of nonadjacent dependencies. *Journal of Experimental Psychology: General*, *151*(11), 2623–2640. https://doi.org/10.1037/xge0001207

Isbilen, E. S., McCauley, S. M., & Christiansen, M. H. (2022). Individual differences in artificial and natural language statistical learning. *Cognition*, *225*, Article 105123. https://doi.org/10.1016/j.cognition.2022.105123

Isbilen, E. S., McCauley, S. M., Kidd, E., & Christiansen, M. H. (2020). Statistically-induced chunking recall: A memory-based approach to statistical learning. *Cognitive Science*, *44*(7), Article e12848. https://doi.org/10.1111/cogs.12848

Jones, G. (2016). The influence of children's exposure to language from two to six years: The case of nonword repetition. *Cognition*, *153*, 79–88. https://doi.org/10.1016/j.cognition.2016.04.017

Joseph, H., & Nation, K. (2018). Examining incidental word learning during reading in children: The role of context. *Journal of Experimental Child Psychology*, *166*, 190–211. https://doi.org/10.1016/j.jecp.2017.08.010

Kahta, S., & Schiff, R. (2019). Deficits in statistical leaning of auditory sequences among adults with dyslexia. *Dyslexia*, *25*(2), 142–157. https://doi.org/10.1002/dys.1618

Keuleers, E., Diependaele, K., & Brysbaert, M. (2010). Practice effects in large-scale visual word recognition studies: A lexical decision study on 14,000 dutch mono- and disyllabic words and nonwords. *Frontiers in Psychology*, *1*, Article 174. https://doi.org/10.3389/fpsyg.2010.00174

Keuleers, E., Lacey, P., Rastle, K., & Brysbaert, M. (2012). The British Lexicon Project: Lexical decision data for 28,730 monosyllabic and disyllabic English words. *Behavior Research Methods*, *44*(1), 287–304. https://doi.org/10.3758/s13428-011-0118-4

Kidd, E. (2012). Implicit statistical learning is directly associated with the acquisition of syntax. *Developmental Psychology*, *48*(1), 171–184. https://doi.org/10.1037/a0025405

Kidd, E., & Arciuli, J. (2016). Individual differences in statistical learning predict children's comprehension of syntax. *Child Development*, *87*(1), 184–193. https://doi.org/10.1111/cdev.12461

Kidd, E., Arciuli, J., Christiansen, M. H., & Smithson, M. (2023). The sources and consequences of individual differences in statistical learning for language development. *Cognitive Development*, *66*, Article 101335. https://doi.org/10.1016/j.cogdev.2023.101335

Kliegl, R., Grabner, E., Rolfs, M., & Engbert, R. (2004). Length, frequency, and predictability effects of words on eye movements in reading. *The European Journal of Cognitive Psychology*, *16*(1–2), 262–284. https://doi.org/10.1080/09541440340000213

Kuperman, V., & Van Dyke, J. A. (2013). Reassessing word frequency as a determinant of word recognition for skilled and unskilled readers. *Journal of Experimental Psychology: Human Perception and Performance*, *39*(3), 802–823. https://doi.org/10.1037/a0030859

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review*, *104*(2), 211–240. https://doi.org/10.1037/0033-295X.104.2.211

Lee, S. M. K., Cui, Y., & Tong, S. X. (2022). Toward a model of statistical learning and reading: Evidence from a meta-analysis. *Review of Educational Research*, *92*(4), 651–691. https://doi.org/10.3102/00346543211073188

Lelonkiewicz, J. R., Ktori, M., & Crepaldi, D. (2020). Morphemes as letter chunks: Discovering affixes through visual regularities. *Journal of Memory and Language*, *115*, Article 104152. https://doi.org/10.1016/j.jml.2020.104152

Lelonkiewicz, J. R., Ktori, M., & Crepaldi, D. (2023). Morphemes as letter chunks: Linguistic information enhances the learning of visual regularities. *Journal of Memory and Language*, *130*, Article 104411. https://doi.org/10.1016/j.jml.2023.104411

Lemhöfer, K., & Broersma, M. (2012). Introducing LexTALE: A quick and valid lexical test for advanced learners of English. *Behavior Research Methods*, *44*(2), 325–343. https://doi.org/10.3758/s13428-011-0146-0

Lund, K., & Burgess, C. (1996). Producing high-dimensional semantic spaces from lexical co-occurrence. *Behavior Research Methods, Instruments, & Computers*, *28*(2), 203–208. https://doi.org/10.3758/BF03204766

Mandera, P., Keuleers, E., & Brysbaert, M. (2020). Recognition times for 62 thousand English words: Data from the English Crowdsourcing Project. *Behavior Research Methods*, *52*(2), 741–760. https://doi.org/10.3758/s13428-019-01272-8

McCauley, S. M., & Christiansen, M. H. (2015). Individual differences in chunking ability predict on-line sentence processing. In D. C. Noelle & R. Dale (Eds.), *Proceedings of the 37th annual conference of the cognitive science society* (pp. 1553–1558). Cognitive Science Society.

McCauley, S. M., & Christiansen, M. H. (2019). Language learning as language use: A cross-linguistic model of child language development.

Psychological Review, *126*(1), 1–51. https://doi.org/10.1037/rev0000126

McCauley, S. M., Isbilen, E. S., & Christiansen, M. H. (2017). Chunking ability shapes sentence processing at multiple levels of abstraction. In G. Gunzelmann, A. Howes, T. Tenbrink, & E. J. Davelaar (Eds.), *Proceedings of the 39th annual conference of the cognitive science society* (pp. 2681–2686). Cognitive Science Society.

Mollon, J. D., Bosten, J. M., Peterzell, D. H., & Webster, M. A. (2017). Individual differences in visual science: What can be learned and what is good experimental practice? *Vision Research*, *141*, 4–15. https://doi.org/10.1016/j.visres.2017.11.001

Monaghan, P., Chang, Y. N., Welbourne, S., & Brysbaert, M. (2017). Exploring the relations between word frequency, language exposure, and bilingualism in a computational model of reading. *Journal of Memory and Language*, *93*, 1–21. https://doi.org/10.1016/j.jml.2016.08.003

Monsell, S., Doyle, M. C., & Haggard, P. N. (1989). Effects of frequency on visual word recognition tasks: Where are they? *Journal of Experimental Psychology: General*, *118*(1), 43–71. https://doi.org/10.1037/0096-3445.118.1.43

New, B., Ferrand, L., Pallier, C., & Brysbaert, M. (2006). Reexamining the word length effect in visual word recognition: New evidence from the English Lexicon Project. *Psychonomic Bulletin & Review*, *13*, 45–52. https://doi.org/10.3758/BF03193811

Norris, D. (2009). Putting it all together: A unified account of word recognition and reaction-time distributions. *Psychological Review*, *116*(1), 207–219. https://doi.org/10.1037/a0014259

Olivers, C. N., Huettig, F., Singh, J. P., & Mishra, R. K. (2014). The influence of literacy on visual search. *Visual Cognition*, *22*(1), 74–101. https://doi.org/10.1080/13506285.2013.875498

Orbán, G., Fiser, J., Aslin, R. N., & Lengyel, M. (2008). Bayesian learning of visual chunks by human observers. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(7), 2745–2750. https://doi.org/10.1073/pnas.0708424105

Palmer, S. D., & Mattys, S. L. (2016). Speech segmentation by statistical learning is supported by domain-general processes within working memory. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, *69*(12), 2390–2401. https://doi.org/10.1080/17470218.2015.1112825

Perea, M., Soares, A. P., & Comesaña, M. (2013). Contextual diversity is a main determinant of word identification times in young readers. *Journal of Experimental Child Psychology*, *116*(1), 37–44. https://doi.org/10.1016/j.jecp.2012.10.014

Perfors, A., & Kidd, E. (2022). The role of stimulus-specific perceptual fluency in statistical learning. *Cognitive Science*, *46*(2), Article e13100. https://doi.org/10.1111/cogs.13100

Perruchet, P. (2019). What mechanisms underlie implicit statistical learning? Transitional probabilities versus chunks in language learning. *Topics in Cognitive Science*, *11*(3), 520–535. https://doi.org/10.1111/tops.12403

Perruchet, P., & Pacton, S. (2006). Implicit learning and statistical learning: One phenomenon, two approaches. *Trends in Cognitive Sciences*, *10*(5), 233–238. https://doi.org/10.1016/j.tics.2006.03.006

Perruchet, P., & Vinter, A. (1998). PARSER: A model for word segmentation. *Journal of Memory and Language*, *39*(2), 246–263. https://doi.org/10.1006/jmla.1998.2576

Powell, D., & Atkinson, L. (2021). Unraveling the links between rapid automatized naming (RAN), phonological awareness, and reading. *Journal of Educational Psychology*, *113*(4), 706–718. https://doi.org/10.1037/edu0000625

Preston, K. A. (1935). The Speed of Word Perception and Its Relation to Reading Ability. *The Journal of General Psychology*, *13*(1), 199–203. https://doi.org/10.1080/00221309.1935.9917878

Pulido, M. F. (2021). Individual chunking ability predicts efficient or shallow L2 processing: Eye-tracking evidence from multiword units in relative clauses. *Frontiers in Psychology*, *11*, Article 607621. https://doi.org/10.3389/fpsyg.2020.607621

Qi, Z., Sanchez Araujo, Y., Georgan, W., Gabrieli, J., & Arciuli, J. (2019). Hearing Matters More Than Seeing: A Cross- Modality Study of Statistical Learning and Reading Ability. *Scientific Studies of Reading*, *23*(1), 101–115. https://doi.org/10.1080/10888438.2018.1485680

R Core Team. (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Raviv, L., & Arnon, I. (2018). The developmental trajectory of children's auditory and visual statistical learning abilities: Modality-based differences in the effect of age. *Developmental Science*, *21*(4), Article e12593. https://doi.org/10.1111/desc.12593

Rayner, K., & McConkie, G. W. (1976). What guides a reader's eye movements? *Vision Research*, *16*(8), 829–837. https://doi.org/10.1016/0042-6989(76)90143-7

Rayner, K., & Raney, G. E. (1996). Eye movement control in reading and visual search: Effects of word frequency. *Psychonomic Bulletin & Review*, *3*(2), 245–248. https://doi.org/10.3758/BF03212426

Reilly, M., & Desai, R. H. (2017). Effects of semantic neighborhood density in abstract and concrete words. *Cognition*, *169*, 46–53. https://doi.org/10.1016/j.cognition.2017.08.004

Ren, J., & Wang, M. (2023). Development of statistical learning ability across modalities, domains, and languages. *Journal of Experimental Child Psychology*, *226*, Article 105570. https://doi.org/10.1016/j.jecp.2022.105570

Ren, J., Wang, M., & Arciuli, J. (2023). A meta-analysis on the correlations between statistical learning, language, and reading outcomes. *Developmental Psychology*, *59*(9), 1626–1644. https://doi.org/10.1037/dev0001577

Rice, G. A., & Robinson, D. O. (1975). The role of bigram frequency in the perception of words and nonwords. *Memory & Cognition*, *3*(5), 513–518. https://doi.org/10.3758/BF03197523

Rothe, J., Schulte-Körne, G., & Ise, E. (2014). Does sensitivity to orthographic regularities influence reading and spelling acquisition? A 1-year prospective study. *Reading and Writing*, *27*(7), 1141–1161. https://doi.org/10.1007/s11145-013-9479-7

Saffran, J. R. (2001). Words in a sea of sounds: The output of infant statistical learning. *Cognition*, *81*(2), 149–169. https://doi.org/10.1016/S0010-0277(01)00132-9

Saffran, J. R. (2003). Statistical language learning mechanisms and constraints. *Current Directions in Psychological Science*, *12*(4), 110–114. https://doi.org/10.1111/1467-8721.01243

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, *274*(5294), 1926–1928. https://doi.org/10.1126/science.274.5294.1926

Sawi, O. M., & Rueckl, J. G. (2019). Reading and the neurocognitive bases of statistical learning. *Scientific Studies of Reading*, *23*(1), 8–23. https://doi.org/10.1080/10888438.2018.1457681

Schmalz, X., Altoè, G., & Mulatti, C. (2017). Statistical learning and dyslexia: A systematic review. *Annals of Dyslexia*, *67*(2), 147–162. https://doi.org/10.1007/s11881-016-0136-0

Schmalz, X., & Mulatti, C. (2017). Busting a myth with the Bayes factor: Effects of letter bigram frequency in visual lexical decision do not reflect reading processes. *The Mental Lexicon*, *12*(2), 263–282. https://doi.org/10.1075/ml.17009.sch

Seidenberg, M. S., & McClelland, J. L. (1989). A distributed, developmental model of word recognition and naming. *Psychological Review*, *96*(4), 523–568. https://doi.org/10.1037/0033-295X.96.4.523

Shaoul, C., & Westbury, C. (2010). Exploring lexical co-occurrence space using HiDEx. *Behavior Research Methods*, *42*(2), 393–413. https://doi.org/10.3758/BRM.42.2.393

Shufaniya, A., & Arnon, I. (2018). Statistical learning is not age-invariant during childhood: Performance improves with age across modality. *Cognitive Science*, *42*(8), 3100–3115. https://doi.org/10.1111/cogs.12692

Siegel, L. S., Share, D., & Geva, E. (1995). Evidence for superior orthographic skills in dyslexics. *Psychological Science*, *6*(4), 250–254. https://doi.org/10.1111/j.1467-9280.1995.tb00601.x

Siegelman, N., Bogaerts, L., Armstrong, B. C., & Frost, R. (2019). What exactly is learned in visual statistical learning? Insights from Bayesian modeling. *Cognition*, *192*, Article 104002. https://doi.org/10.1016/j.cognition.2019.06.014

Siegelman, N., Bogaerts, L., Elazar, A., Arciuli, J., & Frost, R. (2018). Linguistic entrenchment: Prior knowledge impacts statistical learning performance. *Cognition*, *177*, 198–213. https://doi.org/10.1016/j.cognition.2018.04.011

Siegelman, N., Elgort, I., Brysbaert, M., Agrawal, N., Amenta, S., Arsenijević Mijalković, J., Chang, C. S., Chernova, D., Chetail, F., Clarke, A. J. B., Content, A., Crepaldi, D., Davaabold, N., Delgersuren, S., Deutsch, A., Dibrova, V., Drieghe, D., Filipović Đurđević, D., Finch, B., … Kuperman, V. (2024). Rethinking first language–second language similarities and differences in English proficiency: Insights from the English reading online (ENRO) project. *Language Learning*, *74*(1), 249–294. https://doi.org/10.1111/lang.12586

Siegelman, N., & Frost, R. (2015). Statistical learning as an individual ability: Theoretical perspectives and empirical evidence. *Journal of Memory and Language*, *81*, 105–120. https://doi.org/10.1016/j.jml.2015.02.001

Siegelman, N., Rueckl, J. G., Lo, J. C. M., Kearns, D. M., Morris, R. D., & Compton, D. L. (2022). Quantifying the regularities between orthography and semantics and their impact on group- and individual-level behavior. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *48*(6), 839–855. https://doi.org/10.1037/xlm0001109

Siegelman, N., Rueckl, J. G., Steacy, L. M., Frost, S. J., van den Bunt, M., Zevin, J. D., Seidenberg, M. S., Pugh, K. R., Compton, D. L., & Morris, R. D. (2020). Individual differences in learning the regularities between orthography, phonology and semantics predict early reading skills. *Journal of Memory and Language*, *114*, Article 104145. https://doi.org/10.1016/j.jml.2020.104145

Spencer, M., Kaschak, M. P., Jones, J. L., & Lonigan, C. J. (2015). Statistical learning is related to early literacy-related skills. *Reading and Writing*, *28*(4), 467–490. https://doi.org/10.1007/s11145-014-9533-0

Swingley, D. (2005). Statistical clustering and the contents of the infant vocabulary. *Cognitive Psychology*, *50*(1), 86–132. https://doi.org/10.1016/j.cogpsych.2004.06.001

Theeuwes, J., Bogaerts, L., & van Moorselaar, D. (2022). What to expect where and when: How statistical learning drives visual selection. *Trends in Cognitive Sciences*, *26*(10), 860–872. https://doi.org/10.1016/j.tics.2022.06.001

Thompson, S. P., & Newport, E. L. (2007). Statistical learning of syntax: The role of transitional probability. *Language Learning and Development*, *3*(1), 1–42. https://doi.org/10.1080/15475440709336999

Treiman, R., & Kessler, B. (2022). Statistical learning in word reading and spelling across languages and writing systems. *Scientific Studies of Reading*, *26*(2), 139–149. https://doi.org/10.1080/10888438.2021.1920951

Vidal, Y., Viviani, E., Zoccolan, D., & Crepaldi, D. (2021). A general-purpose mechanism of visual feature association in visual word identification and beyond. *Current Biology*, *31*(6), 1261–1267.e3. https://doi.org/10.1016/j.cub.2020.12.017

(*Appendices follow*)

Appendix A

Targets Used in the Word Search Task and Their Linguistic Properties

Word	Frequency (categorical)	Log frequency	Concreteness rating	Semantic neighborhood density	Semantic diversity	Mean bigram frequency	No. of letter	No. of syllable
Body	High	11.663	4.79	0.684	1.922	483	4	2
Bull	High	9.025	4.85	0.625	1.654	1,056.67	4	1
Cult	High	9.292	2.71	0.625	1.451	687.667	4	1
Debt	High	9.56	2.72	0.615	1.464	962.333	4	1
Desk	High	9.52	4.87	0.579	1.463	2,438.67	4	1
Dirt	High	8.941	4.86	0.582	1.656	1,229.67	4	1
Duty	High	9.616	2.19	0.644	1.796	667.667	4	2
Folk	High	9.392	3.57	0.636	1.737	664.667	4	1
Food	High	10.997	4.8	0.672	1.768	683.667	4	1
Gene	High	10.149	3.59	0.621	0.962	2,434.00	4	1
Girl	High	10.605	4.85	0.659	1.584	556.667	4	1
Jazz	High	9.529	3.9	0.636	1.187	100	4	1
Lake	High	9.981	4.88	0.653	1.566	1,082.67	4	1
Loss	High	10.309	2.19	0.667	1.943	1,295.00	4	1
Luck	High	11.013	1.33	0.601	1.867	655.667	4	1
Lung	High	9.51	4.82	0.571	1.315	2,118.67	4	1
Mile	High	9.175	3.63	0.613	1.669	2,073.33	4	1
Role	High	10.402	2.19	0.686	2.079	2,285.67	4	1
Tree	High	10.212	5	0.658	1.552	2,366.67	4	1
Yard	High	8.912	4.82	0.61	1.663	1,235.67	4	1
Acne	Low	6.284	4.89	0.427	1.334	1,248.00	4	2
Aunt	Low	7.953	4.17	0.6	1.259	1,780.67	4	1
Axle	Low	7.462	4.5	0.57	0.958	1,171.67	4	2
Calf	Low	7.298	4.48	0.48	1.494	1,656.00	4	1
Deed	Low	7.591	3.86	0.554	1.477	2,652.00	4	1
Diva	Low	6.771	3.64	0.515	1.161	1,072.33	4	2
Fern	Low	6.344	5	0.506	1.443	2,402.00	4	1
Fuzz	Low	6.588	4.37	0.392	1.455	171	4	1
Germ	Low	6.469	3.89	0.434	1.44	2,652.00	4	1
Halo	Low	6.841	4.03	0.553	1.702	1,892.67	4	2
Haze	Low	6.843	4.04	0.432	1.469	584.333	4	1
Hymn	Low	7.735	4.04	0.559	1.259	150.667	4	1
Mole	Low	7.337	4.41	0.552	1.552	1,892.67	4	1
Moth	Low	6.244	4.69	0.508	1.362	1,064.00	4	1
Puma	Low	6.389	4.86	0.425	1.362	902	4	2
Reef	Low	8	4.7	0.568	1.093	1,953.00	4	1
Smog	Low	6.589	4.14	0.395	0.979	665.333	4	1
Verb	Low	7.65	2.85	0.577	0.734	2,778.33	4	1
Wick	Low	6.326	4.69	0.444	1.709	1,234.00	4	1
Zeal	Low	6.551	2.33	0.499	1.777	1,794.33	4	1
Birth	High	10.059	4.46	0.661	1.537	937.75	5	1
Blood	High	10.855	4.86	0.66	1.662	1,027.50	5	1
Chaos	High	9.869	2.79	0.613	1.831	881.5	5	2
Laser	High	9.688	4.5	0.606	1.404	2,916.00	5	2
Media	High	10.968	3.57	0.681	1.572	2,305.75	5	2
Movie	High	11.023	4.59	0.668	1.008	881.25	5	2
Night	High	11.488	4.52	0.681	1.851	786.75	5	1
Queen	High	9.768	4.45	0.668	1.544	1,407.75	5	1
Scene	High	10.371	3.93	0.674	1.906	2,150.00	5	1
Style	High	10.674	2.67	0.677	1.903	1,944.25	5	1
Aroma	Low	6.686	3.67	0.482	1.243	2,001.25	5	3
Cider	Low	6.504	4.97	0.465	1.341	2,648.75	5	2
Envoy	Low	6.816	3.5	0.577	1.299	1,110.50	5	2
Gauze	Low	5.717	4.62	0.31	1.556	404	5	1
Lymph	Low	6.016	3.48	0.501	0.942	779	5	1
Molar	Low	6.258	4.83	0.446	0.785	1,811.50	5	2
Niece	Low	6.506	4.28	0.583	1.447	1,426.25	5	1
Snack	Low	7.42	4.36	0.533	1.365	1,000.00	5	1
Spree	Low	6.279	2.92	0.451	1.614	1,896.25	5	1
Tunic	Low	5.717	3.75	0.421	1.279	1,601.25	5	2
Church	High	10.801	4.9	0.663	1.557	1,022.00	6	1
Degree	High	10.648	3	0.654	2.034	1,839.00	6	2

(*Appendices continue*)

Appendix A (*continued*)

Word	Frequency (categorical)	Log frequency	Concreteness rating	Semantic neighborhood density	Semantic diversity	Mean bigram frequency	No. of letter	No. of syllable
Energy	High	10.978	3.11	0.665	1.645	2,538.80	6	3
Length	High	10.835	4.07	0.654	2.014	2,500.40	6	1
Object	High	10.866	3.66	0.649	1.697	595	6	2
Script	High	10.412	4.72	0.634	1.274	1,087.20	6	1
Celery	Low	6.596	4.8	0.418	0.643	2,715.60	6	3
Clique	Low	6.466	2.68	0.426	1.556	861	6	1
Grocer	Low	5.298	4.3	0.385	1.554	2,266.80	6	2
Pebble	Low	6.404	4.86	0.414	1.609	1,334.60	6	2
Recess	Low	6.236	3.07	0.431	1.626	2,707.20	6	2
Sludge	Low	6.469	4.23	0.439	1.194	543.4	6	1
Alcohol	High	9.562	4.76	0.621	1.38	1,359.50	7	3
Receipt	High	8.741	4.86	0.51	1.607	1,445.83	7	2
Science	High	11.505	2.79	0.664	1.485	1,660.33	7	1
Variety	High	10.333	2.13	0.676	2.11	1,604.17	7	3
Bouquet	Low	6.392	4.74	0.367	1.309	790.333	7	2
Cheddar	Low	6.057	4.46	0.398	0.961	1,909.33	7	2
Giraffe	Low	5.938	4.73	0.392	1.442	950.5	7	2
Replica	Low	7.041	3.86	0.565	1.573	2,139.33	7	3
Currency	High	9.182	4.25	0.604	1.407	1,788.71	8	3
Daughter	High	9.877	4.79	0.65	1.585	1,812.43	8	2
Judgment	High	8.748	1.68	0.604	1.341	1,330.00	8	2
Security	High	11.044	2.82	0.668	1.902	1,585.29	8	4
Chivalry	Low	6.109	1.88	0.412	1.377	1,095.14	8	3
Juncture	Low	6.144	3	0.354	1.758	1,601.14	8	2
Obstacle	Low	7.46	3.48	0.556	2.151	1,640.57	8	3
Souvenir	Low	6.475	4.45	0.464	1.62	1,504.29	8	3

Appendix B

Word Search Background Letter Frequencies

Letter	% occurrence
e	12
t	10
a	8
o	8
i	8
n	7
s	6
r	5
h	5
l	4
d	3
c	3
u	2
m	2
f	2
p	2
w	2
g	2
y	2
b	1
v	1
k	1
x	1
j	1
q	1
z	1

(*Appendices continue*)

Appendix C

LexTALE Lexical Decision Items

Item	Answer (is a word)
Abergy	No
Ablaze	Yes
Alberation	No
Allied	Yes
Bewitch	Yes
Breeding	Yes
Carbohydrate	Yes
Celestial	Yes
Censorship	Yes
Cleanliness	Yes
Crumper	No
Cylinder	Yes
Destription	No
Dispatch	Yes
Eloquence	Yes
Exprate	No
Fellick	No
Festivity	Yes
Flaw	Yes
Fluid	Yes
Fray	Yes
Hasty	Yes
Hurricane	Yes
Ingenious	Yes
Interfate	No
Kermshaw	No
Kilp	No
Lengthy	Yes
Listless	Yes
Lofty	Yes
Magrity	No
Majestic	Yes
Mensible	No
Moonlit	Yes
Muddy	Yes
Nourishment	Yes
Plaintively	Yes
Plaudate	No
Proom	No
Pudour	No
Pulsh	No
Purrage	No
Quirty	No
Rascal	Yes
Rebondicate	No
Recipient	Yes
Savory	Yes
Scholar	Yes
Scornful	Yes
Screech	Yes
Shin	Yes
Skave	No
Slain	Yes
Spaunch	No
Stoutly	Yes
Turmoil	Yes
Turtle	Yes
Unkempt	Yes
Upkeep	Yes

Note. LexTALE = Lexical Test for Advanced Learners of English.

(*Appendices continue*)

Appendix D

ENRO Spelling Recognition Items

Word	Answer (correctly spelled)
Accommodation	Yes
Addmission	No
Announcment	No
Aplause	No
Appreciate	Yes
Attitude	Yes
Bachelor	Yes
Basicly	No
Behavior	Yes
Benafit	No
Classafied	No
Commitment	Yes
Conscientious	Yes
Consequence	Yes
Conveinient	No
Distinguish	Yes
Elementary	Yes
Exhibition	Yes
Fulcrum	Yes
Gaurantee	No
Honerable	No
Imminant	No
Implie	No
Importent	No
Independent	Yes
Interpretation	Yes
Interrogate	Yes
Jeapordise	No
Misary	No
Missellaneous	No
Mortgage	Yes
Necesscarily	No
Occurence	No
Partitionining	No
Plagarism	No
Political	Yes
Pollution	Yes
Proliferate	Yes
Reciept	No
Rendezvous	Yes
Seperate	No
Sincirely	No
Sufficient	Yes
Suspicious	Yes

Note. ENRO = English Reading Online.

(*Appendices continue*)

Appendix E

ENRO Orthographic Awareness Items

Nonword 1	Nonword 2	Correct answer
Bleam	Blaem	Bleam
Bolz	Bolk	Bolk
Clid	Cklid	Clid
Dayke	Dake	Dake
Dlun	Lund	Lund
Fague	Fageu	Fague
Fant	Tanf	Fant
Fenth	Femth	Fenth
Filk	Filv	Filk
Frength	Frenkth	Frength
Fyeth	Fieth	Fieth
Gmup	Gnup	Gnup
Kmort	Knort	Knort
Miln	Milg	Miln
Moke	Moje	Moke
Netch	Neetch	Netch
Nitl	Nilt	Nilt
Pihgt	Pight	Pight
Poat	Paot	Poat
Powl	Lowp	Powl
Qoust	Quost	Quost
Sckap	Skap	Skap
Tamb	Tabm	Tamb
Tign	Tagn	Tign
Tirtt	Tirth	Tirth
Trenths	Trenthz	Trenths
Tridth	Tribth	Tridth
Visn	Vism	Vism
Vrine	Wrine	Wrine
Wolg	Wolt	Wolt

Note. ENRO = English Reading Online.

Received January 11, 2024
Revision received July 25, 2024
Accepted August 3, 2024 ■