Reading Research Quarterly

INTERNATIONAL LITERACY ASSOCIATION

ARTICLE OPEN ACCESS

Eye Tracking and Simulating the Spacing Effect During Orthographic Learning

Signy Wegener[1] | Anne Castles[1] | Elisabeth Beyersmann[2] | Kate Nation[3] | Hua-Chen Wang[4] | Erik D. Reichle[2]

[1]Australian Centre for the Advancement of Literacy, Australian Catholic University, North Sydney, New South Wales, Australia | [2]School of Psychological Sciences, Macquarie University, Sydney, New South Wales, Australia | [3]Department of Experimental Psychology, University of Oxford, Oxford, UK | [4]School of Education, Macquarie University, Sydney, New South Wales, Australia

Correspondence: Signy Wegener (signy.wegener@acu.edu.au)

Received: 10 June 2024 | **Revised:** 23 December 2024 | **Accepted:** 23 February 2025

Funding: This work was supported by a grant from the Australian Research Council (grant number DP200100311) to A.C., E.B., K.N., and E.D.R.

Keywords: eye tracking | E-Z Reader | orthographic learning | spacing effect | word reading

ABSTRACT

Spreading out study opportunities over time improves the retention of verbal material compared to consecutive study, yet little is known about the influence of temporal spacing on orthographic learning specifically. The current study addressed four questions: (1) do readers' eye movements during orthographic learning differ under spaced and massed conditions? (2) is the spacing effect observed in offline post-tests? (3) can readers' eye movements during learning be linked with learning success in offline post-tests? (4) can E-Z Reader simulate the spacing effect during orthographic learning? Eighty adults silently read sentences containing novel words while their eye movements were monitored. Sentences were read four times; half of the items were spaced while half were massed. Participants completed a post-test assessing their written word form learning (orthographic choice or spelling). Simulations with E-Z Reader were used to interpret the human data. During orthographic learning, massed items had shorter total reading times than spaced items. A spacing advantage was noted in the offline post-tests. Longer fixations during learning were associated with higher response accuracy at post-test. Implementing a processing deadline enabled E-Z Reader to simulate participants' eye movements; simulations suggested that massed items may have received less attentional processing than spaced items during learning. Temporal spacing results in longer fixations during learning and better learning outcomes using offline tests. The combination of human eye movements and computational modeling provides useful insights into how reading and memory intersect and points to new directions for future research.

Reading is a powerful driver of written word learning, or *orthographic learning* (Castles and Nation 2006; Nation and Castles 2017). This specific form of learning is generally viewed as an important contributor to the development of fluent and accurate word reading, and hence the transition from beginning to skilled reading. The *self-teaching hypothesis* (Share 1995) theorizes that the process of phonological decoding provides an opportunity for the reader to learn the spellings of decoded words and from this, orthographic knowledge is acquired. Orthographic learning is thought to operate throughout the lifetime, at the level of individual written words (Share 2008), and as such, memory representations of newly encountered written words can be acquired by both children and adults via their independent reading whenever a novel word is encountered in print. While phonological decoding exerts a strong influence on orthographic learning (de Jong and Share 2007; Kyte and Johnson 2006), the distributional properties of readers' experiences with words, such as the frequency with which they have been seen in print, also make a contribution to written word learning (for review see Brysbaert et al. 2018). Other

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2025 The Author(s). *Reading Research Quarterly* published by Wiley Periodicals LLC on behalf of International Literacy Association.

distributional properties of words might also influence learning; if this is so, manipulations of exposures to new written words in educational contexts might hold potential as a method of improving learning efficiency during independent reading. An obvious candidate factor is the timing with which new words are encountered in reading. For instance, a new word might be encountered multiple times within the same paragraph, meaning that exposures to this new word are very close together in time. Alternatively, exposures to a new written word could be more spread out in time, such that a reader might encounter a new word for the first time during sentence reading, but not encounter that word again until several minutes, hours, or days later.

The *spacing effect* refers to the finding that spreading out exposures to new information over time gives rise to better performance on post-tests than when new information is presented consecutively during learning (Cepeda et al. 2006). Importantly, in these studies, the number of exposures to the new information is kept the same; it is only their distribution in time that differs. A wealth of evidence supports the utility of the spacing principle for promoting learning in contexts that require phonological and semantic learning. Most notably, spacing improves performance on verbal learning tasks, while recent work points to the utility of the principle when it is applied to learning academic content within both laboratory and classroom settings (Cepeda et al. 2006; Wiseheart et al. 2019).

A range of theories of the spacing effect point to the potential role of several factors, either independently or in combination. According to the *encoding variability* account (Estes 1955; Glenberg 1979), the spacing effect is driven by the availability of a wide range of cues at the point of encoding when exposures are spread out in time, which in turn provide more routes to retrieval at post-test than massed encounters. According to *study-phase retrieval* accounts (Thios and D'Agostino 1976), the spacing effect is driven by attempts to retrieve prior learning episodes such that massed exposures are less likely to require retrieval of past encounters than spaced exposures. According to *deficient processing* accounts (Challis 1993; Cuddy and Jacoby 1982), the spacing effect is driven by attentional factors such that close repetitions attract fewer attentional resources than spaced repetitions. Turning to reading specifically, theories of visual word recognition (Coltheart et al. 2001; McClelland and Rumelhart 1981; Morton 1969; Rumelhart and McClelland 1982) suggest a fourth account. These theories imply that massed encounters with words prompt only a single change in activation levels—that is, at the first visual exposure. Due to the close proximity in time of repeated encounters, activation levels remain high, and thus do not change markedly with each subsequent exposure. In contrast, spaced encounters with words might prompt multiple changes in activation levels—one at each exposure—because the delay between exposures allows activation to rise and fall before the next encounter with the word. Because changes in activation levels are conceptualized as a key driver of learning in activation theories of word reading, spaced exposures should therefore result in better learning than massed exposures.

Recently, Wegener, Wang, Beyersmann, Nation, et al. (2022) observed that relatively little work has to date investigated the role of temporal spacing within orthographic learning specifically. Experimental orthographic learning paradigms embed novel written words into sentences that are otherwise composed of known written words, thereby creating an analogue for the experience of a reader encountering a written word in print for the first time. Orthographic learning trials require participants to read sentences either aloud (most commonly) or silently. The outcome of these learning opportunities on orthographic form learning is subsequently assessed using tasks tapping the *recognition* and/or *recall* of the novel word spellings. Because the spacing effect is defined in terms of the outcome of learning, the presence of a spacing effect in orthographic learning is supported if, at post-test, participants recognize and/or recall the spellings of novel words more accurately when they had been spread out in time during learning than when they had occurred consecutively.

The influence of temporal spacing on post-test measures of orthographic form learning following independent reading has been investigated in three prior studies. In an intentional second language vocabulary learning experiment with English-speaking adults, Koval (2019) showed participants an English word (e.g., *bridge*) before presenting a novel word with the same meaning in Finnish (e.g., *silta*) embedded in an English sentence. Half of the items were massed (experienced consecutively) during exposure, while half were spaced (experienced once per experimental block). At post-test, participants' recognition of the novel orthographic forms was assessed using a version of the orthographic choice task. A significant recognition advantage was found for items in the spaced condition compared to the massed condition 15 min following the learning phase (immediate post-test) and 44–78 h following the learning phase (delayed post-test). Importantly, the distractors in this task were formed by scrambling the letters in the target word or by transposing its syllables (e.g., *tasil*). This approach meant that the target and distractors had different pronunciations, leaving open the possibility that the task could have been performed by relying on memories of the decoded phonology of the novel words when responding, rather than memories of the words' spellings.

Wegener, Wang, Beyersmann, Nation, et al. (2022) compared spaced with massed exposures to novel written words in an orthographic learning study with children in Grades 3 and 4, and evaluated the influence of this manipulation on two within-participants outcome measures. The first was a traditional orthographic choice task in which participants were presented with a learned spelling (e.g., *sloak*) and a homophonic spelling (e.g., *sloke*), along with two one-letter-different distractor spellings (e.g., *sloat* and *slote*). This approach has the advantage that, in order to distinguish the target from its homophonic foil, the participant must have learned something about the word's spelling (Share 2004). The second task was a spelling-to-dictation task in which participants were provided with the phonology of the novel words and were asked to reproduce the spellings they saw during the learning phase. A clear spacing advantage was found on the orthographic choice task, which was administered first, but the effect was not observed on the spelling-to-dictation task, which was administered second.

To clarify these discrepant findings, a follow-up study with skilled readers was reported by Wegener, Wang, Beyersmann, Reichle, et al. (2023). Following an orthographic learning phase

in which spacing was manipulated, participants completed one of three post-test measures. The first two outcome measures were orthographic choice (recognition) and spelling-to-dictation (cued-recall). The third outcome measure was a letter-cue spelling task (cued-recall) in which participants were provided with the first and last letters of the trained words and were asked to complete the word's spelling (e.g., if the participant had read the word *blawm* they were presented with the stimulus *b_ _ _ m*). The authors administered each of the three post-test measures between participants on the basis that whatever post-test measure was administered first offered the most direct test of the effect of temporal spacing on orthographic learning, precluding any potential influences arising from test order. A significant spacing effect was found on each of the three outcome measures, leading the authors to conclude that temporal spacing boosts learning outcomes relative to massing regardless of the nature of the post-test measure of orthographic form learning.

If spacing out exposures to new written words during reading promotes the learning of orthographic forms, as these three studies suggest (Koval 2019; Wegener, Wang, Beyersmann, Nation, et al. 2022; Wegener, Wang, Beyersmann, Reichle, et al. 2023), this raises questions about whether there might also be differences in the nature of the processing that occurs for spaced and massed items during the orthographic learning phase itself. Evidence from verbal (semantic) learning tasks is consistent with this possibility. For instance, when participants are allowed to control their own study time, less time tends to be devoted to massed than to spaced items (e.g., Krug et al. 1990; Shaughnessy et al. 1972).

An informative method of gaining insight into the processing that occurs moment-to-moment during orthographic learning is eye movement monitoring (Nation and Castles 2017). A growing number of studies have used eye movements over repeated exposures to novel words to explore the evolution of eye movement behavior, with perhaps the key finding being that fixation durations (Ginestet et al. 2020, 2021; Joseph et al. 2014; Joseph and Nation 2018; Pellicer-Sánchez 2016; Wegener et al. 2020) and number of fixations (Ginestet et al. 2020, 2021; Pellicer-Sánchez 2016) reduce with building visual experience. Only two prior studies, both with skilled readers, have recorded participants' eye movements during orthographic learning of spaced and massed novel words. Pagán and Nation (2019) embedded rare English words into sentences which participants read silently as their eye movements were monitored. Participants were not instructed to intentionally learn the novel words; rather, learning of meaning was incidental and accrued with experience. Each word appeared four times under spaced or massed conditions, but at the point of analysis, eye movement data relating to fixation durations on the target word during learning were collapsed across the four exposure trials. Analyses at this global level revealed that massed exposures attracted shorter fixation durations than spaced exposures in both gaze duration and go-past time.

In addition to the offline measures described earlier, Koval (2019) also used eye movements to gain insight into online processing during orthographic learning. Unlike Pagán and Nation (2019), Koval (2019) instructed participants to intentionally try to learn the novel words; and, at the point of analysis, information about exposure number was retained. A key finding was that total reading times reduced across exposures in both conditions, but the reduction was particularly pronounced in the massed compared to the spaced condition, with massed items attracting shorter fixation durations than spaced items at the second, third, and fourth orthographic exposures. A similar but less pronounced pattern was observed in both first fixation and gaze duration. In a separate mediation analysis, spaced exposures attracted longer fixations than massed exposures, which in turn were associated with improved performance on a post-test measure of learning success.[1] Notably, the post-test measure of learning success was a composite of two measures of orthographic form recognition (immediate and delayed) and two measures of semantic learning (immediate and delayed). Recalling that the orthographic form recognition task in this experiment could have been completed based on participants' memory of the decoded phonology, the measure of learning success used by Koval (2019) could be viewed as a composite measure of phonological, semantic, and perhaps orthographic learning. Whether a similar pattern might be observed using a purer measure of orthographic form learning remains to be established.

A complementary method that holds great potential for understanding the processing that occurs moment-to-moment during reading is computational modeling of eye movements. Computational models are formal instantiations of theories using mathematical equations or computer programs, which can be used to describe, explain, and make predictions (Reichle 2021). Applied to eye movements during reading, they can be used to both interrogate models of word reading and to generate and test hypotheses about the processes involved in reading. Here, we will describe and report simulations using *E-Z Reader* (Reichle et al. 1998; Reichle et al. 2003; Reichle et al. 2012; Veldre et al. 2023), which is a computational model of how vision and cognition interact with the oculomotor system to guide eye movements during reading.

Perhaps due to its inherent complexity, the type of reading that is characterized by being fully engaged with the text and understanding whatever is being read can rapidly shift from this "default" mode into lapses of attention that are indicative of "mindless reading" (Reichle et al. 2010), on one hand, and a variety of sophisticated "strategies" (e.g., skimming rapidly displayed subtitles; see Reichle et al. 2021), on the other. An example of the latter that will be explored in the simulations reported below involves the fairly common situation in which a reader encounters an unknown word and then attempts to infer something about its meaning from the sentence context. At least three aspects of this situation make it theoretically interesting. The first is the question of how readers will treat these new items when they are initially encountered. The second is the question of how readers use information in the sentences to infer the likely meanings of these new items. The final is the question of how repeated encounters with these new items might influence their learning. Although it is in principle possible to address these questions by examining the patterns of eye movements that readers make when they encounter novel words in sentences, in practice the patterns of eye movements that are observed during reading are extremely complex and thus not readily open to interpretation. For that reason, our strategy is to interpret the patterns of eye movements observed in our experiment using the E-Z Reader

model (Veldre et al. 2023). Our approach thus entails using the model to simulate and understand the mental processes that might be engaged when readers encounter and make sense of novel words.

1.1 | The Current Experiment

We posed four questions with the aim of investigating the influence of temporal spacing on adults' orthographic form learning during independent reading. In the orthographic learning phase, participants silently read sixteen sentences a total of four times, with each sentence containing an embedded novel word, and their eye movements were monitored as they read. Half of the items were temporally distributed (sentences appeared once in each of four blocks) while the other half were massed (presented consecutively within a single block). After a brief retention interval, the acquisition of word-specific orthographic knowledge was assessed between subjects using either an orthographic choice task (recognition memory) or a spelling-to-dictation task (cued-recall).

1.1.1 | Research Question 1

During the orthographic learning phase of the experiment, do participants' fixation durations (first fixation; gaze duration; total reading time) during reading differ under spaced and massed conditions? Based on data reported by Koval (2019), when comparing the first and second trials, we expected to observe an interaction between condition and trial number, driven by a larger reduction of fixation durations in the massed than spaced condition. For the subsequent trials, we expected to observe a spacing effect on fixation durations, such that massed items attract shorter fixation durations than spaced items. We also asked whether the probability of skipping the target word differs under spaced and massed conditions.

1.1.2 | Research Question 2

Are novel orthographic forms experienced with temporal spacing during the learning phase better recognized and recalled than novel orthographic forms experienced consecutively during learning? We expected to observe a spacing effect on post-test accuracy data, with spaced items being learned better than massed items. If found, this would replicate the findings of Wegener, Wang, Beyersmann, Reichle, et al. (2023) by demonstrating evidence of a spacing effect on orthographic form learning among skilled readers. We did not expect to observe a significant difference in the magnitude of the spacing effect for participants completing the orthographic choice compared to the spelling-to-dictation test.

1.1.3 | Research Question 3

We sought to link participants' eye movements during the learning phase with their performance on the orthographic learning outcome measure at test by asking whether participants' fixation durations during the learning phase are a significant predictor of their orthographic form learning at test? We expected to find that longer fixations during the orthographic learning phase would be associated with better learning of novel orthographic forms at test.

1.1.4 | Research Question 4

We used E-Z Reader (Veldre et al. 2023), which is a computational model of eye movements during reading, to simulate the pattern of eye movements observed in our experiment. Our goal in doing this was to both determine what additional theoretical assumptions—if any—might be necessary to explain our results, and by doing this, gain insight into the mental processes that might be engaged when readers encounter novel words in written sentences and attempt to learn something about them.

The experimental design, hypotheses, and analysis plan for Research Questions 1–3 were preregistered: https://aspredicted.org/LCW_7Z1. The simulations used to address Research Question 4 were exploratory in nature and thus not preregistered.

2 | Methods

There were two distinct experimental designs within the orthographic learning experiment. The first related to the learning phase and took the form of a 2 (within-subject manipulation of spacing: massed vs. spaced) × 4 (within-subject manipulation of trial number: 1–4) design. The second related to the test phase in which participants completed an outcome task measuring their recognition or recall of the novel orthography. This took the form of a 2 (within-subject manipulation of spacing: massed vs. spaced) × 2 (between-subject manipulation of test type: orthographic choice vs. spelling-to-dictation) + 1 (a continuous variable reflecting fixation durations on the target word made during the learning phase).

2.1 | Participants

Eighty undergraduate students participated in the experiment for course credit. All were native speakers of English; 26% of the sample reported that they fluently spoke at least one additional language. The mean age of participants was 20.5 years (SD = 0.61) and 60 were female (75%). Participants were randomly allocated by a computer program to one of two test type conditions: 39 participants received the orthographic choice task; 41 participants received the spelling-to-dictation task.

2.2 | Materials

2.2.1 | Novel Words

Stimuli were taken from Wegener, Wang, Beyersmann, Reichle, et al. (2023). There were 16 homophonic pairs (e.g., *smaup/smawp*) of 4- to 5-letter, monosyllabic pseudowords with regular

pronunciations according to grapheme-phoneme rules (Rastle and Coltheart 1999). One item from each pair was allocated to experimental list 1 and the other was allocated to experimental list 2 (see Appendix A). Each manipulated grapheme occurred with different spellings in two pseudowords within each list. For example, participants who learned list 1 encountered *smaup* and *blawm*; participants who learned list 2 encountered *smawp* and *blaum*. This was done to control for any preference participants might have for a given spelling. The manipulated grapheme was in the vowel position for 62.5% of items (e.g., *smaup*/*smawp*) and in a consonant position for 37.5% of items (e.g., *sleff*/*sleph*). Additionally, each list was divided into two subsets so that the allocation of items to conditions could be counterbalanced (e.g., some participants learned *smaup* in the massed condition and *blawm* in the spaced condition, whereas others learned *smaup* in the spaced condition and *blawm* in the massed condition). These procedures yielded four versions of the learning phase of the experiment, and participants were randomly allocated to one version.

2.2.2 | Experimental Sentences

Single line sentences from Wegener, Wang, Beyersmann, Reichle, et al. (2023) were presented with the novel words embedded as nouns in a mid-sentence position (see Appendix B). Sentences were a mean of 12.13 words in length (SD = 1.86); each referred to an invention created by a fictional character (Professor Parsnip; Wang et al. 2011) and contained a single presentation of one novel word.

2.3 | Procedure

There were two phases in the behavioral experiment: a learning phase and a test phase.

2.3.1 | Orthographic Learning Task

E-Z Reader (Veldre et al. 2023) is a computational model of eye movements during silent reading. Therefore, participants were instructed to silently read a series of sentences referring to fictional inventions while their eye movements were monitored. They were told that they would read each sentence several times and that they should try to remember what they read. Each trial began with a drift correct target, which the participant fixated before the trial was initiated by the experimenter. Participants then read the sentence silently and used a button box to terminate each trial. Recalibration was performed as required. Eye movements were recorded with a desktop SR Research Ltd. Eyelink 1000 eye tracker in head stabilized mode, with a sampling rate of 1000 Hz. Participants read binocularly, but only movements of the right eye were recorded. Single line sentences were displayed on a white background in Courier New 14-point font at a viewing distance of 104 cm. Each character subtended approximately 0.24° of visual angle.

As in Wegener et al. (2023), half of the sentences in the orthographic learning task contained items that were presented in a massed fashion (the same sentence was read four times on consecutive trials) while half appeared in the spaced condition (the same sentence was read four times, once in each of the four experimental blocks). There were 16 trials in each of four randomized experimental blocks. Items from both the spaced and massed conditions appeared in each block; trial order was fixed within blocks to maintain the lag between sentences in the spaced condition. Blocks one to four were separated by two-and-a-half minute breaks during which participants watched nonverbal cartoons.

2.3.2 | Testing Phase

A seven-minute retention interval occurred between the end of the four orthographic learning blocks and the test phase. Participants watched a nonverbal cartoon during this delay. Participants were randomly allocated to one of two test types designed to assess their learning of the novel orthographic forms they had read during the learning phase. Test data were collected using the Gorilla online experiment platform (Anwyl-Irvine et al. 2020).

2.3.2.1 | Orthographic Choice Condition. Four written words were arranged horizontally on a computer screen, and participants were required to select, via mouse click, the spelling they read during the orthographic learning phase. The four written words were the target spelling, its homophonic spelling from the untrained list, and two visual distractors that were homophones of each other. The target appeared in each location an equal number of times. Trial order was randomized for each participant. The dependent variable was target recognition accuracy. A value of 1 was assigned by a computer program if the participant correctly selected the spelling they had seen during the learning phase, and a value of 0 was assigned if any of the other three spellings were selected.

2.3.2.2 | Spelling-to-Dictation Condition. Spoken target words were audio recorded. Each participant listened to the items in a randomized order and attempted to spell them by typing into a text box before pressing "Enter" to submit their responses. A value of 1 was assigned by a computer program if the participants' spelling matched the orthographic form they experienced during learning, and a value of 0 was assigned if the exact spelling was not produced. The dependent variable was spelling accuracy.

3 | Results

Data and scripts are available at: https://osf.io/mwhr2/. Data were analyzed in the R (R Core Team 2021) computing environment.

3.1 | Learning Phase

Models of the learning phase data had the following fixed effects structure: a within-subject manipulation of condition (massed; spaced), a within-subject manipulation of exposure number (1–4), and their interaction. The area of interest was the target

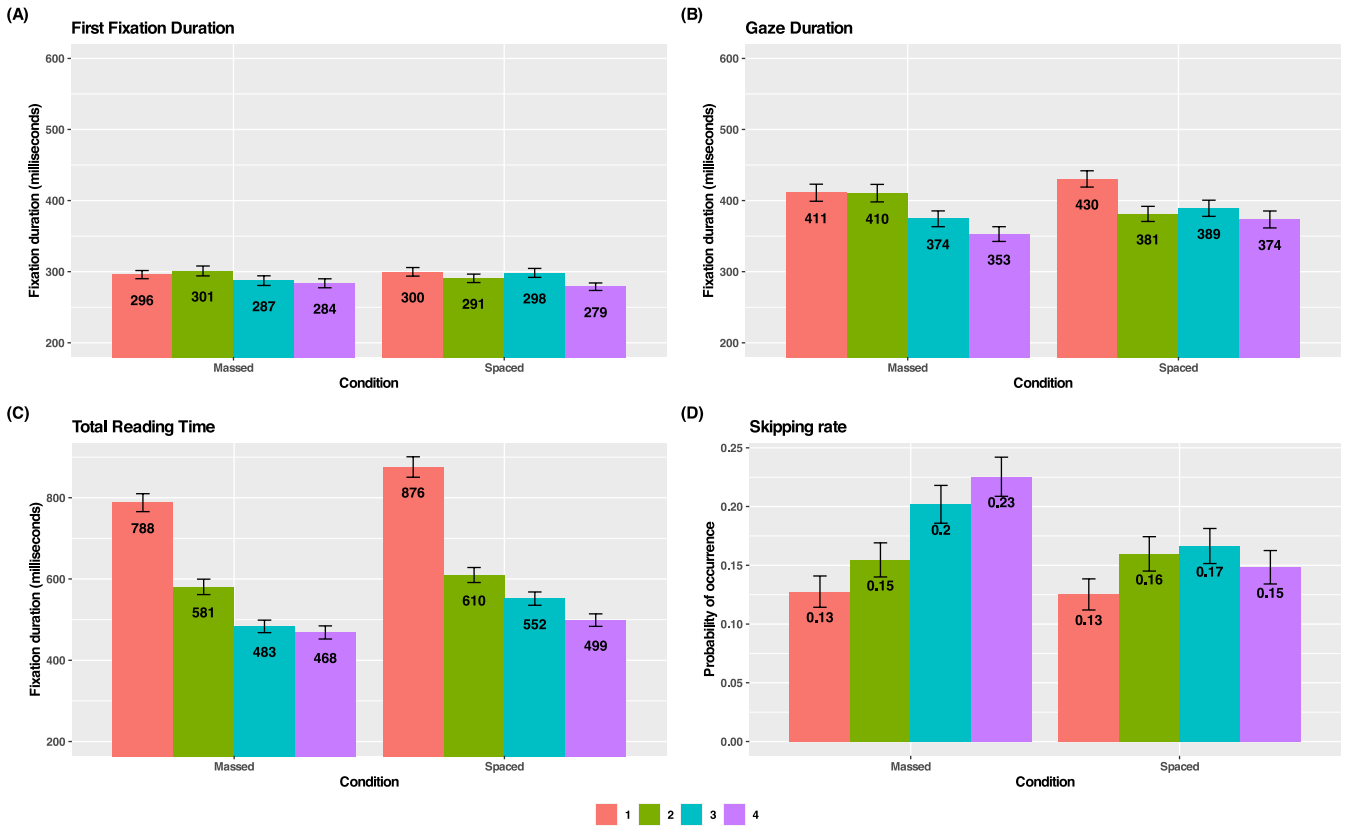


FIGURE 1 | Bar charts reflecting means and standard errors of eye movement-dependent variables. Time data are measured in milliseconds while skipping rate data reflect probability of occurrence. The bar colors indicate exposures 1–4.

word itself. Data cleaning was performed using DataViewer software (SR Research Ltd.). Fixations <80 ms or >1200 ms were merged with fixations within one character space, or they were deleted if there was no nearby fixation. Trials in which blinks occurred on the target words were removed prior to analysis (97.71% of the experimental data remained). The condition factor was deviation coded (massed: −0.5; spaced: +0.5), while the *MASS* package (Ripley et al. 2019) was used to code the exposure number factor using successive differences contrasts (i.e., 2nd vs. 1st exposure; 3rd vs. 2nd exposure; 4th vs. 3rd exposure). Models were implemented using the *lme4* package (Bates et al. 2020), and *p*-values were obtained using the lmerTest package (Kuznetsova et al. 2017). Models included random intercepts by participant and item. Maximal models included random slopes by condition and exposure; however, these models were singular and/or failed to converge. Therefore, a data-driven approach to model selection was adopted, beginning with the random intercepts model and incrementally adding random slopes. Following Matuschek et al. (2017), random slopes were retained if their addition significantly improved model fit ($p<0.20$). As per the preregistration, models were run for first fixation duration (duration of the initial fixation on the target word), gaze duration (sum of all first-pass fixations on the target word), total reading time (sum of all fixations on the target word, including any returns to the target), and skipping rate. Significant interactions were unpacked using the *emmeans* package (Lenth et al. 2019). Arithmetic means and standard errors for each of the dependent variables appear in Figure 1. Reading time data were log-transformed to improve the distributions of model residuals.

Results of the model reflecting first fixation duration are shown in Table 1. There is no significant fixed effect of spacing condition. Indeed, the only significant difference occurs between exposures 3 and 4, with fixation durations shortening between these two exposures. No two-way interactions between condition and exposure number were significant.

Results of the model reflecting gaze duration are shown in Table 2. There was no significant fixed effect of spacing condition. However, there were significant fixed effects when exposure 2 was compared to exposure 1; when exposure 3 was compared to exposure 2; and when exposure 4 was compared to exposure 3. Together, these findings show that participants' gaze durations reduced with each exposure to the target words. The only significant two-way interaction was between spacing condition and exposure 2 compared to exposure 1. When unpacked, the only significant simple effect was of exposure 2 vs. exposure 1 for spaced items such that gaze durations on items in the spaced condition reduced more from the first to the second exposure than items in the massed condition ($z=4.46$, $p<0.001$).

Results of the model reflecting total reading time are presented in Table 3. There was a significant effect of spacing condition in this model, indicating that items that were spaced during orthographic learning received longer total times than items in the massed

TABLE 1 | Results of the linear mixed effects model for first fixation duration.

	Condition	Exposure 2 vs. 1	Exposure 3 vs. 2	Exposure 4 vs. 3
Fixed effects				
b	0.01	−0.01	−0.02	−0.05
SE	0.02	0.02	0.02	0.02
t	0.46	−0.70	−1.24	−3.00
p	0.645	0.485	0.214	< 0.003[a]
	Condition × Exposure 2 vs. 1	**Condition × Exposure 3 vs. 2**		**Condition × Exposure 4 vs. 3**
Two-way interactions				
b	−0.03	0.03		−0.02
SE	0.30	0.03		0.03
t	−1.10	0.98		−0.51
p	0.271	0.326		0.609

[a]Denotes statistical significance.

TABLE 2 | Results of the linear mixed effects model for gaze duration.

	Condition	Exposure 2 vs. 1	Exposure 3 vs. 2	Exposure 4 vs. 3
Fixed effects				
b	0.05	−0.07	−0.10	−0.15
SE	0.03	0.02	0.02	0.02
t	1.88	−3.53	−5.06	−7.37
p	0.061	< 0.001[a]	< 0.001[a]	< 0.001[a]
	Condition × Exposure 2 vs. 1	**Condition × Exposure 3 vs. 2**		**Condition × Exposure 4 vs. 3**
Two-way interactions				
b	−0.11	0.00		−0.00
SE	0.04	0.04		0.04
t	−2.77	0.03		−0.11
p	0.006[a]	0.978		0.913

[a]Denotes statistical significance.

condition. Fixed effects comparing fixation durations at each of the orthographic exposures (2 vs. 1; 3 vs. 2; and 4 vs. 3) were all significant, indicating that total times decreased with each subsequent reading of the target words. None of the two-way interactions between condition and exposure number were significant.

Results of the logistic mixed effects model reflecting first-pass skipping are presented in Table 4. The fixed effect of spacing condition was not significant, suggesting that there was no overall difference in skipping rate for items experienced under massed versus spaced conditions. Fixed effects comparing skipping at each of the orthographic exposures (2 vs. 1; 3 vs. 2; and 4 vs. 3) were all significant, indicating that the skipping rate increased across orthographic exposures to the novel words. The only significant two-way interaction between condition and exposure number was when exposure 4 was compared with exposure 3. When unpacked, the only significant simple effect was of condition at exposure 4 ($z = 4.00$, $p < 0.001$), indicating that spaced items had a lower probability of being skipped than massed items at the fourth orthographic exposure. The simple effect of condition at exposure 3 was not significant ($z = 2.18$, $p = 0.088$), nor were the simple effects of exposure 4 to 3 for massed items ($z = -1.10$, $p = 0.546$) or exposure 4 to 3 for spaced items ($z = 0.94$, $p = 0.546$).

3.2 | Test Phase

Two logistic linear mixed effects models were constructed. The first model was not preregistered, but we elected to run and report it because it used a complete dataset (it is referred to as the full data set in the OSF repository); that is, it contained each participant's responses to all items (8 items that were massed during learning and 8 items that were spaced). This model contained the following fixed effects structure: a within-subject manipulation of

TABLE 3 | Results of the linear mixed effects model for total reading time.

	Condition	Exposure 2 vs. 1	Exposure 3 vs. 2	Exposure 4 vs. 3
Fixed effects				
b	0.10	−0.35	−0.50	−0.55
SE	0.03	0.02	0.48	0.02
t	3.29	−15.99	−21.47	−24.67
p	0.001[a]	< 0.001[a]	< 0.001[a]	< 0.001[a]
	Condition × Exposure 2 vs. 1	**Condition × Exposure 3 vs. 2**	**Condition × Exposure 4 vs. 3**	
Two-way interactions				
b	−0.04	0.06	0.00	
SE	0.04	0.04	0.04	
t	−0.95	1.37	−0.02	
p	0.343	0.170	0.987	

[a]Denotes statistical significance.

TABLE 4 | Results of the logistic linear mixed effects model for skipping.

	Condition	Exposure 2 vs. 1	Exposure 3 vs. 2	Exposure 4 vs. 3
Fixed effects				
b	−0.12	0.29	0.50	0.50
SE	0.20	0.12	0.12	0.12
z	−0.64	2.35	4.21	4.22
p	0.525	0.019[a]	< 0.001[a]	< 0.001[a]
	Condition × Exposure 2 vs. 1	**Condition × Exposure 3 vs. 2**	**Condition × Exposure 4 vs. 3**	
Two-way interactions				
b	0.06	−0.25	0.57	
SE	0.24	0.24	0.24	
z	0.23	−1.05	−2.38	
p	0.815	0.293	0.018[a]	

[a]Denotes statistical significance.

condition (massed; spaced) and a between-subjects manipulation of test type (orthographic choice; spelling-to-dictation). The model was implemented using the *lme4* package (Bates et al. 2020), and *p*-values were obtained using the lmerTest package (Kuznetsova et al. 2017). The model included random intercepts by participant and item. When random slopes by condition were included, it produced a singular fit, so the random intercepts model is reported here. This deviation to the preregistration is specified in Data S1, per Willroth and Atherton's (2024) recommendation.

Descriptive statistics are provided in Table 5, while Figure 2 visualizes both the data distributions and by-participant total scores in each condition. The model revealed a significant fixed effect of condition such that items learned with temporal spacing were acquired better than items experienced in consecutive sentences ($b = 0.95$; SE = 0.13; $z = 7.33$; $p < 0.001$). There was also a significant effect of test type such that response accuracy was higher for the orthographic choice than the spelling-to-dictation task ($b = -1.14$; SE = 0.19; $z = -6.03$; $p < 0.001$). In contrast, the two-way interaction between condition and test type was not

TABLE 5 | Accuracy means and standard errors for each experimental condition using the full dataset.

	Orthographic choice	Spelling-to-dictation
Massed	0.48 (0.03)[a]	0.28 (0.02)
Spaced	0.70 (0.03)[b]	0.44 (0.03)

[a]An error analysis revealed that, in the massed condition of the orthographic choice task, 28% of errors were homophone errors, while 72% of the errors were distractor errors.
[b]An error analysis revealed that, in the spaced condition of the orthographic choice task, 32% of errors were homophone errors, while 68% of the errors were distractor errors.

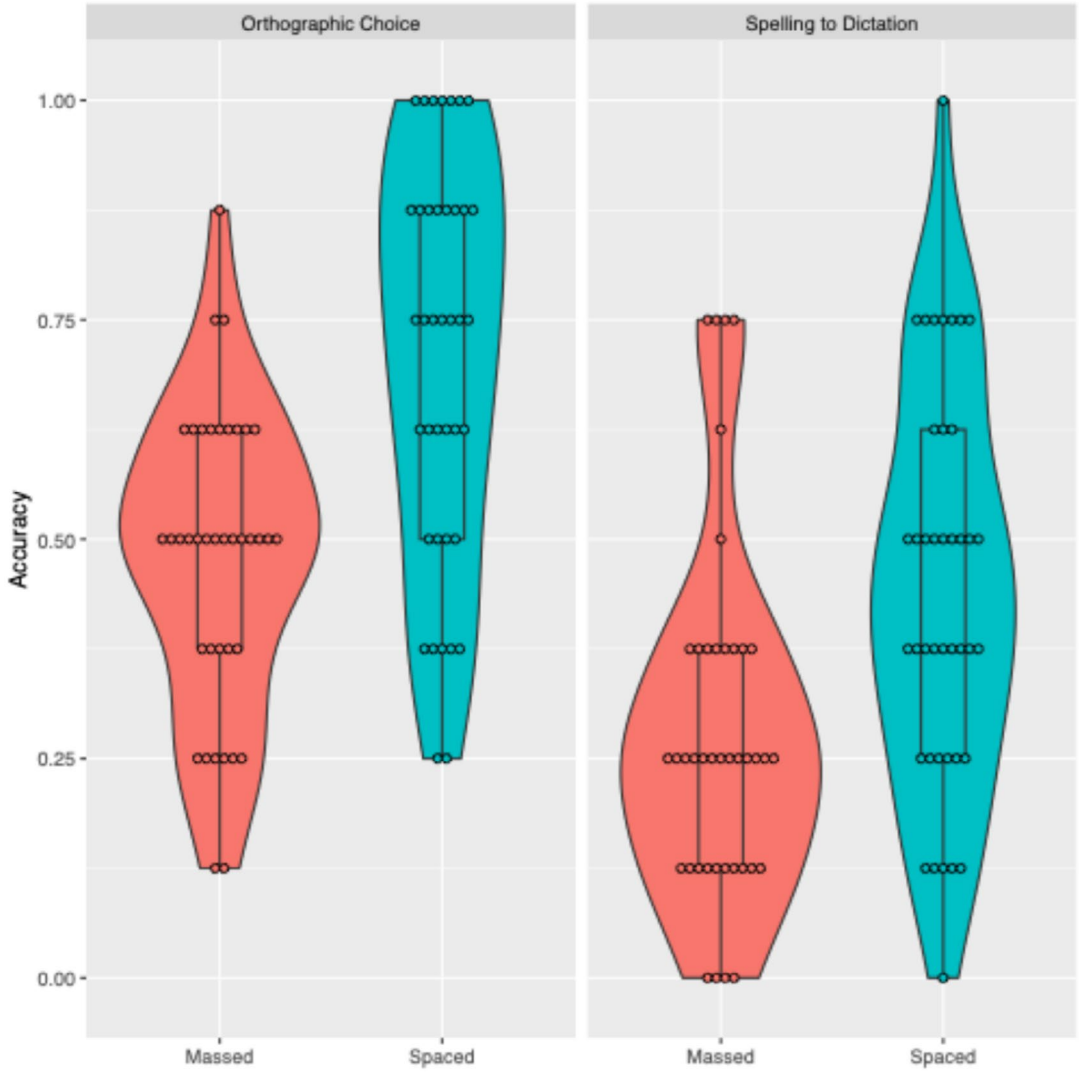


FIGURE 2 | Violin plots of response accuracy by condition and test type using the full dataset. Dots represent by-participant total scores by condition.

significant, indicating that the magnitude of the spacing effect did not significantly differ as a function of whether the participant completed the orthographic choice or spelling-to-dictation task at test ($b = -0.27$; SE = 0.25; $z = -1.06$; $p = 0.290$).

The second model relating to the test phase data was preregistered but included an additional data preparation step (see Data S1 for further details). The model contained the same fixed effects structure as before, namely a within-subject manipulation of condition (massed; spaced) and a between-subjects manipulation of test type (orthographic choice; spelling-to-dictation) with the addition of an item-level continuous measure of fixation durations during the learning phase. This fixation duration measure reflects the sum of the total reading time fixation durations made on the target word at each of the four exposures during the learning phase. If any one of an item's four orthographic learning trials contained a blink on first-pass reading, the item was removed from the analysis (7.89% of the data were removed), meaning that the model was calculated on a slightly smaller dataset than was used for the previous analysis (it is referred to as the 'reduced dataset' in the OSF repository). If a target word was simply not fixated at one or more of the four exposures, it contributed a value of zero to its sum, and the item was retained for the purpose of the analysis. This approach was taken because it is the most accurate reflection of the total time participants spent fixating each of the target words across the four visual exposures. Fixation durations were centered. The model was implemented using the *lme4* package (Bates et al. 2020) and *p*values were obtained using the lmerTest package (Kuznetsova et al. 2017). The dependent variable was accuracy.

The model included random intercepts by participant and item. When random slopes by condition were included, it produced a singular fit, so the random intercepts model is reported here. Consistent with the model produced using the full data set, there was an effect of condition such that learning was better for items that were spaced during learning than items that were experienced in consecutive sentences ($b = 0.84$; SE = 0.14; $z = 6.22$; $p < 0.001$); response accuracy was higher for the orthographic choice than the spelling-to-dictation task ($b = -1.14$; SE = 0.19; $z = -6.01$; $p < 0.001$); and the magnitude of the spacing effect did not significantly differ as a function of whether the participant completed the orthographic choice or spelling-to-dictation task

at test ($b = -0.17$; SE = 0.27; $z = -0.62$; $p = 0.533$). This model included an additional test in which response accuracy was predicted by fixation durations during the learning phase; this test showed that response accuracy increased as fixation durations increased ($b = 0.25$; SE = 0.08; $z = 3.22$; $p = 0.001$).

4 | Simulations

Five simulations were completed using the E-Z Reader model of eye-movement control in reading (Veldre et al. 2023). The objectives of this exercise were to first interpret our key findings, particularly the patterns of fixation durations and probabilities observed on the novel words used in our experiment, and to then provide a tentative account of our spacing manipulation, using the model to explain how the manipulation may have influenced the patterns of eye movements that were observed in our experiment. Before we do this, however, it is first necessary to describe the E-Z Reader model. This description will only be sufficient to understand the logic of the simulations and the implications of their results; more complete descriptions of the model and how the simulations were completed are presented in the Data S1 and S2, and for in-depth discussion of the model's assumptions and theoretical scope, see Reichle (2011, 2021).

4.1 | E-Z Reader

Figure 3 is a schematic diagram of the model. The model consists of a handful of equations that specify how long these processes require to complete, as well as factors that influence the accuracy of eye movements. As such, the model only provides a high-level description of the processes that guide the eyes during reading and does not provide detailed accounts of word identification or higher-level language processing.

As Figure 3 shows, the model has two core assumptions: (1) attention is allocated serially to support the lexical processing of only one word at a time, and (2) the completion of a preliminary stage of lexical processing, called the *familiarity check*, on a word triggers the programming of a saccade to move the eyes to the next word. By this account, the familiarity check indicates that a word is in the reader's lexicon (see Reichle and Perfetti 2003) and that a saccade can be programmed to move the eyes to the next word, whereas lexical access both allows post-lexical integration of the identified word and causes attention to shift to the next word.

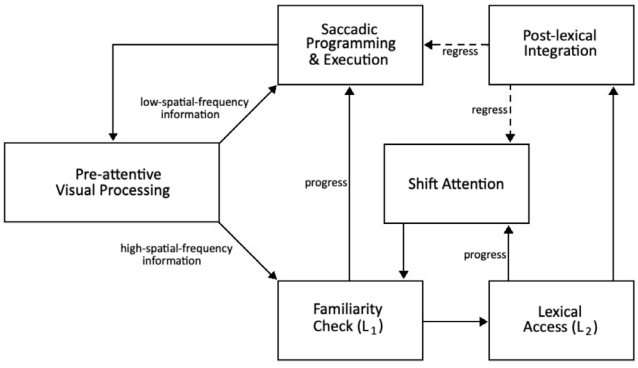


FIGURE 3 | Schematic diagram of the E-Z Reader model. The model's component processes are indicated by the boxes and the transfer of information and control is indicated by the arrows.

In the formalism of the model, the time (in ms) required to complete the familiarity check, $t(L_1)$, is given by Equation (1), where the parameters α_1 (=124), α_2 (=11.1), and α_3 (=76) respectively determine the maximum time to complete $t(L_1)$ and how that time is reduced as a function of the log of word n's frequency of occurrence in printed text and within-sentence cloze predictability. Thus, with all else being equal, common and/or predictable words require less time to process than uncommon or unpredictable words, consistent with what has been observed (e.g., Rayner et al. 2004). Similarly, the time required to complete lexical access, $t(L_2)$, is a fixed multiple of $t(L_1)$, as specified by the parameter Δ (=1.68) in Equation (2).

$$t(L_1) = \begin{cases} 0, \text{with } p = \text{predictability}_n \\ \alpha_1 - \alpha_2 \log_{10}\text{frequency}_n - \alpha_3 \text{predictability}_n, \text{with } p = 1 - \text{predictability}_n \end{cases} \quad (1)$$

$$t(L_2) = \Delta(\alpha_1 - \alpha_2 \log_{10}\text{frequency}_n - \alpha_3 \text{predictability}_n) \quad (2)$$

The values of the aforementioned parameters (and others that will be introduced shortly) and their interpretations are shown in Table 6 for each of the five simulations that will be reported below. Simulations 1 and 2 are our attempts to explain readers' eye movements when they first encounter novel words in sentences. Simulations 3, 4, and 5 then provide a tentative account of the key differences between the spaced and massed conditions. The logic of these simulations was to first fit the model to our experimental data in order to estimate the values of four or more of the model's parameters and to then use those parameter values to make informed inferences about what might be happening in the minds of our participants. The logic of this approach is thus similar to how one might use the time required for a train to travel between two towns separated by a known distance to estimate the train's average speed.

All of our simulations were completed using the actual materials from our experiment (see Data S1 and S2) and below we focus on explaining six dependent measures: (1) *first-fixation duration* (*FFD*), or the duration of the first fixation on a word during first-pass reading: (2) *single-fixation duration* (*SFD*), or the duration of a fixation on a word that is the recipient of only one fixation during the first pass; (3) *gaze duration* (*GD*), or the sum of all first-pass fixations on a word; and the probabilities of (4) fixating a word once (*Pr1*); (5) making two or more fixations on a word (*Pr2*); or (6) skipping a word (*PrS*). These measures collectively provide a fairly complete record of readers' eye movements during first-pass reading. We did not simulate second-pass measures (e.g., total reading time) because E-Z Reader does not provide a detailed account of higher-level language processing or how readers might engage in the re-reading of text. The model does not provide accurate predictions about second-pass measures. The goodness-of-fit

TABLE 6 | E-Z Reader parameters, interpretations, values, and goodness-of-fit (RMSD).

Parameter	Interpretation	Parameter values				
		Sim. 1	**Sim. 2**	**Sim. 3**	**Sim. 4**	**Sim. 5**
α_1	Mean maximal L_1 time (ms)	92	98	84	83	83
α_2	Frequency effect on L_1 time (ms)	13	7.4	6.4	1.7	1.6
α_3	Predictability effect on L_1 time (ms)	32	32	32	32	32
Δ	Prop. difference between L_1 and L_2	1.02	1.66	2.06	2.07	2.25
θ	Lexical-processing deadline (ms)	—	205	225	214	214
$p(\theta)$	Processing deadline probability	—	—	0.63	0.49	0.55
$\text{frequency}_{\text{novel}}$	Mean frequency of non-words	—	—	1	3	1
$\text{predictability}_{\text{novel}}$	Mean predictability of non-words	—	—	0	0	0.85
RMSD	Goodness-of-fit metric	0.590	0.458	0.364	0.560	0.505

Note: Dashes indicate parameters that were not used in a given simulation.

between the observed and simulated means of the six measures was quantified using the root mean squared deviation (*RMSD*; as described in appendix A of Reichle et al. 2012) between the observed and simulated means, with smaller RMSD values indicating better model fits (e.g., a value of 0 indicates a perfect fit with identical observed and simulated means). Finally, the simulations reported below were completed using 10,000 statistical subjects per condition to ensure stability of results, making it unnecessary to use inferential statistics to determine if two RMSD values are different.

4.2 | Simulation 1

This simulation used the standard version of E-Z Reader to simulate the reading of our sentences during the first block of trials. Figure 4 shows the observed mean values of our six dependent measures for all of the words in our corpus (i.e., known and novel words), organized into six classes defined by the $\log_{10}$(frequency)[2]. These observed means were calculated using the eye-movement data from the first block of trials. The words are grouped into frequency classes because the capacity to accurately simulate the effects of word frequency is an important "benchmark" that has often been used for evaluating the performance of eye-movement models (e.g., see Engbert et al. 2005, figure 9; Liu et al. 2024, figure 7; McDonald et al. 2005, figure 3; Reichle et al. 2012, figure A2).

Because the novel words are by definition completely new and thus also unpredictable, our starting assumption in Simulation 1 was that the mean time to complete the familiarity check on these items, $t(L_1)$, would equal α_1 (see Equation 1), the maximum time for completing L_1. The simulation results using the model's best-fitting parameters (see Table 6) are also shown in Figure 4. As can be seen, the model simulated the dependent measures for the middle- and high-frequency words reasonably well but fared less well with the lowest-frequency words and novel words (RMSD = 0.590). (To provide some basis for evaluating this fit, consider that good fits of the model typically produce RMSDs of ~0.3; see Veldre et al. 2023).

For the known words, the model systematically underpredicted the proportion of single fixations and overpredicted the proportion of skips. Although the discrepancy between the observed and simulated values for words of $\log_{10}(\text{frequency}) = 3$ is puzzling, we suspect that the simulated values are in closer agreement with what one would expect because, in our materials, the number of words in this class was fairly small ($N = 3$ vs. $N \geq 14$ in each of the other frequency classes), and because of the ubiquity of the *word-frequency effect*, or the finding that lower-frequency words are typically the recipients of more, longer fixations than higher-frequency words (see Rayner 1998, 2009).

For the novel words, the model underpredicted the fixation-duration measures (e.g., simulated first-fixation durations were 18 ms shorter than what was observed) and the proportion of single fixations (by 0.12) but overpredicted the proportion of refixations (by 0.14). The poor fit to the novel words was anticipated given that the purpose of this modeling exercise was to make inferences about their processing, but with the presupposition that readers would probably treat known and novel words differently.

The poor fits of the model in Simulation 1 are due to the fact that a single set of parameter values must simultaneously accommodate both known and novel words. The nature of this problem can be understood by considering α_1, the parameter that controls the maximum time required to complete the initial stage of lexical processing, $t(L_1)$, on a given word. Because the value of α_1 largely determines the duration of $t(L_1)$ and thus how long a word will be fixated (see Equation 1), the value of α_1 must be small enough to ensure that novel words are not the recipients of excessively long fixations but large enough to ensure that known words are not the recipients of excessively short fixations. Although the best-fitting value of this parameter ($\alpha_1 = 92$ ms; see Table 6) thus attempts to accommodate these two contrary objectives, Figure 4 shows that the best-fitting value of α_1 was not sufficient to accurately simulate the eye-movement measures on the two types of words. This suggests a fundamental limitation of the standard E-Z Reader model—that it cannot simultaneously explain why fixations on known words are as long as they

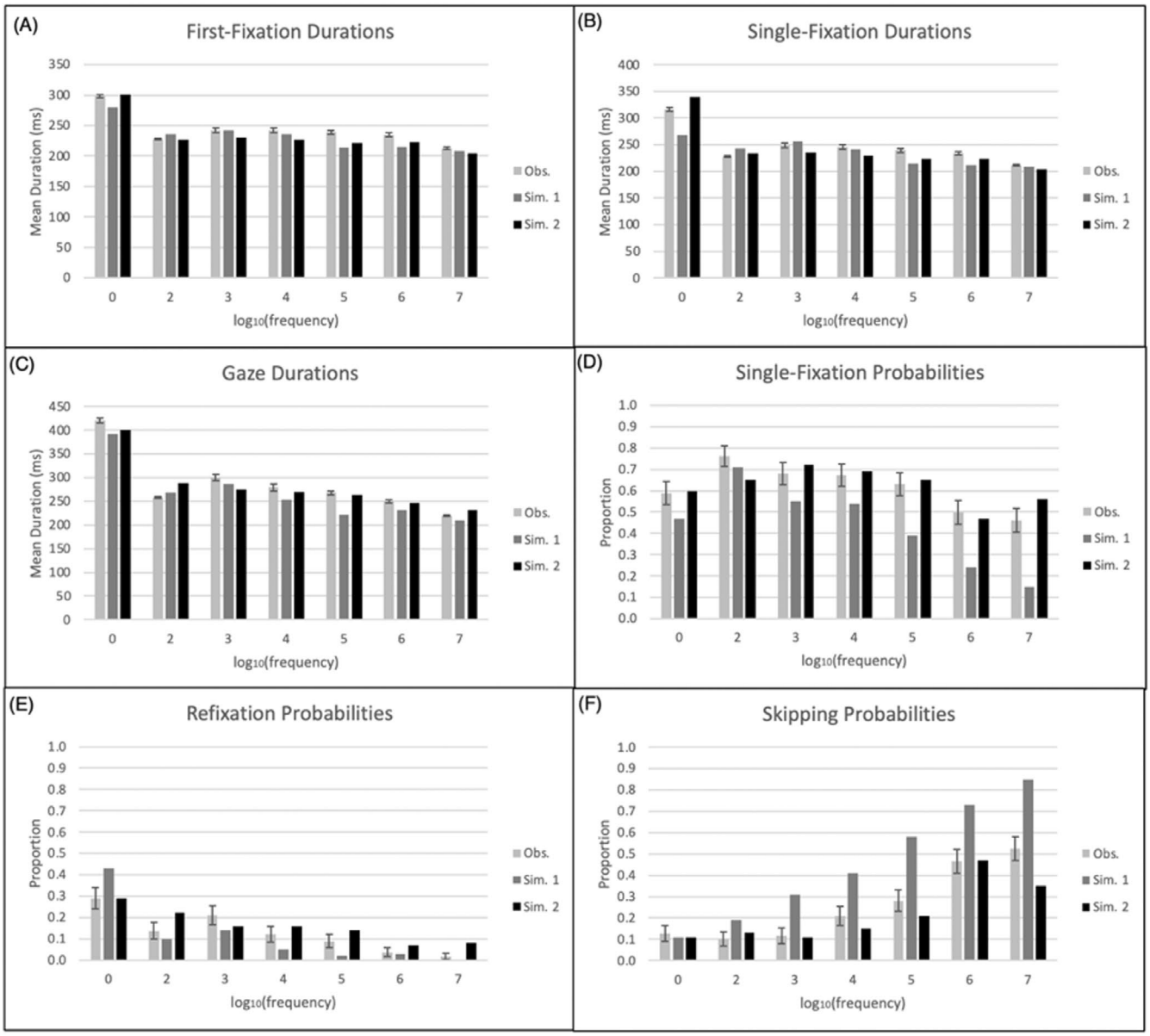


FIGURE 4 | Mean observed and simulated (A) FFD; (B) SFD; (C) GD; (D) Pr1; (E) Pr2; and (F) PrS for words (frequency classes 2–7) and novel words (frequency Class 0) during first exposure to novel words (simulations 1 and 2). Error bars are the observed standard errors.

are while fixations on novel words are not ridiculously long. The goal of Simulation 2 was to address this limitation.

4.3 | Simulation 2

To resolve the limitation identified in Simulation 1, we completed a second simulation to more explicitly address what might happen when readers encounter completely novel words. This second simulation introduced the assumption of a lexical-processing deadline. By this account, whenever lexical processing fails to complete (as it would when a novel word is encountered) after some amount of time, represented by the parameter θ, attention shifts to the next word, and a saccadic program is also initiated to move the eyes to the next word. Any value of $t(L_1)$ that exceeds θ is thus set equal to θ ms. Although the precise nature of the processing that occurs up until the processing deadline will not be specified in detail here, it would presumably include the encoding of the novel word's form and whatever information about its meaning has been garnered from the preceding sentence context, allowing a preliminary representation of the item to be formed in the mental lexicon. Normal processing would then resume, allowing attention and the eyes to shift to the next word and possibly some amount of post-lexical integration of the novel word.

To examine the feasibility of this simple account of novel words, we replicated the previous simulation using values of the lexical-processing parameters and θ (see Table 6) that maximized the goodness-of-fit between the observed and simulated means, the frequency classes, and the dependent measures (RMSD = 0.458). A comparison of Simulations 1 and 2 indicates that the introduction of a lexical-processing deadline improved the model's fit: As Figure 4 shows, the model now

provides reasonable fits of both the known and novel words. For example, Simulation 2 now predicts single-fixation and refixation rates that are much closer to those observed with the known words. And for the novel words, Simulation 2 now predicts only a 3-ms deviation between the observed and simulated first-fixation durations, a 0.01 deviation in the proportion of single fixations, and no significant deviation in the proportion of refixations.

4.4 | Interim Discussion

The introduction of a processing deadline allows E-Z Reader to simulate the finding that neither known words nor novel words are the recipients of fixations that are unreasonably long or short. Moreover, the notion of a processing deadline for initiating saccades is not new; it has been proposed in other models of eye-movement control (e.g., Engbert et al. 2005; Henderson and Ferreira 1990; Li and Pollatsek 2020; Reilly 1993; Reilly and Radach 2006) and has also previously been used within the framework of E-Z Reader (Reichle et al. 2012) to explain the long fixations that are observed in *z*-string "reading," an experimental paradigm wherein participants are instructed to "pretend" to read text in which all of the letters have been replaced with *z*s (e.g., see Rayner and Fischer 1996). Although it remains unclear whether this assumption is *necessary* to explain the eye movements observed with our novel words, our simulations show that they are *sufficient*. And arguably, the assumption is quite plausible because it involves a cognitive operation that might be under some degree of strategic control (e.g., thereby allowing adaptive reading behaviors such as the skimming of rapidly displayed subtitles in film; see Reichle et al. 2021), in contrast to the assumptions and/or parameters that control saccadic programming and execution, which are likely to be fixed across tasks and thus outside of strategic control.

Given our tentative account of how readers deal with novel words when they are first encountered, our second set of simulations represents an attempt to explain how readers respond to novel words when they are encountered during their fourth exposure (i.e., during the final block of trials). Additionally, this second set of simulations attempts to better understand how our spacing manipulation might have affected the learning of the novel words as indexed by our participants' patterns of eye movements.

As mentioned previously, there are four non-mutually exclusive accounts of why the spacing of learning trials is beneficial. The first is that the spacing effect reflects a cost associated with massed learning trials that comes from the to-be-learned materials being more familiar when they occur in rapid succession, causing the repeated materials to be less well attended and/or encoded. The second is that the spacing effect reflects an inherent advantage that comes from encoding the to-be-learned materials across varied contexts, which is more likely to vary with spaced versus massed presentation. The third is that any representation of the to-be-learned material has a resting activation threshold that must be exceeded to become active in working memory, and that spaced repetitions are more likely to provide opportunities for lowering the resting threshold by allowing them to reach their minimal value between repetitions. Finally, spacing effects might reflect the additional advantage that accrues from being forced to recall the to-be-remembered material from long-term memory rather than, for example, simply maintaining the material in working memory across the short retention intervals in the massed conditions. Although our simulations are *not* completely diagnostic about which of the preceding accounts best explains our experiment, as will be demonstrated below, they do offer some insights into how novel words are learned during reading across spaced versus massed presentations.

The simulations reported below were completed using the model as described previously, but fitting the model to the eye-movement data collected from the fourth block of trials. This was done to focus our efforts on explaining the maximal effects of learning by contrasting the first encounter with the novel words to their fourth and final encounter. Also, the simulations were completed twice, using eye-movement data from the spaced trials in Simulation 3 and the massed trials in Simulation 4. This was necessary because the effects of the spacing manipulation were not localized to the novel words, but were instead evident across all of the words in our sentences. Thus, our objective was to provide a global account of how spaced versus massed presentation of the novel words may have affected both the reading behavior of our participants and the nature of what was learned about the novel words as a consequence of these experiences. Because it was not possible to ascertain how much participants learned about the novel words across the course of the experiment, we used the model to quantify the effects of learning by introducing two final model assumptions. Both of these assumptions were motivated by the intuition that our participants might have varied in terms of how much they learned about the individual novel words from their experiences.

Therefore, the first of these assumptions is that some of the novel words may not have been sufficiently learned even after three encounters, effectively resulting in the lexical-processing deadline being evoked with some non-zero probability, represented by the parameter $p(\theta)$. The second assumption is that, for those novel words that were learned, the quality or strength of their representations might vary across participants and/or items. This second assumption was instantiated by allowing the mean frequency of the learned novel words, $\text{frequency}_{\text{novel}}$, to take on any value between 1 and 3. (The value of this parameter thus signifies how well the novel words are represented in memory after their third encounter, with the value then being used in Equation (1)). Thus, in Simulations 3 and 4, the values of these two parameters, as well as those that control the rate of lexical processing and the lexical-processing deadline, were again selected to maximize the goodness-of-fit between the mean observed and simulated eye-movement measures (see Data S1 and S2). Those best-fitting parameter values are shown in Table 6.

4.5 | Simulation 3

As indicated, this simulation was of our participants' final encounters with the novel words in the spaced-presentation condition. As shown in Figure 5, the model provided an adequate account of the eye-movement data (RMSD = 0.364). As Table 6

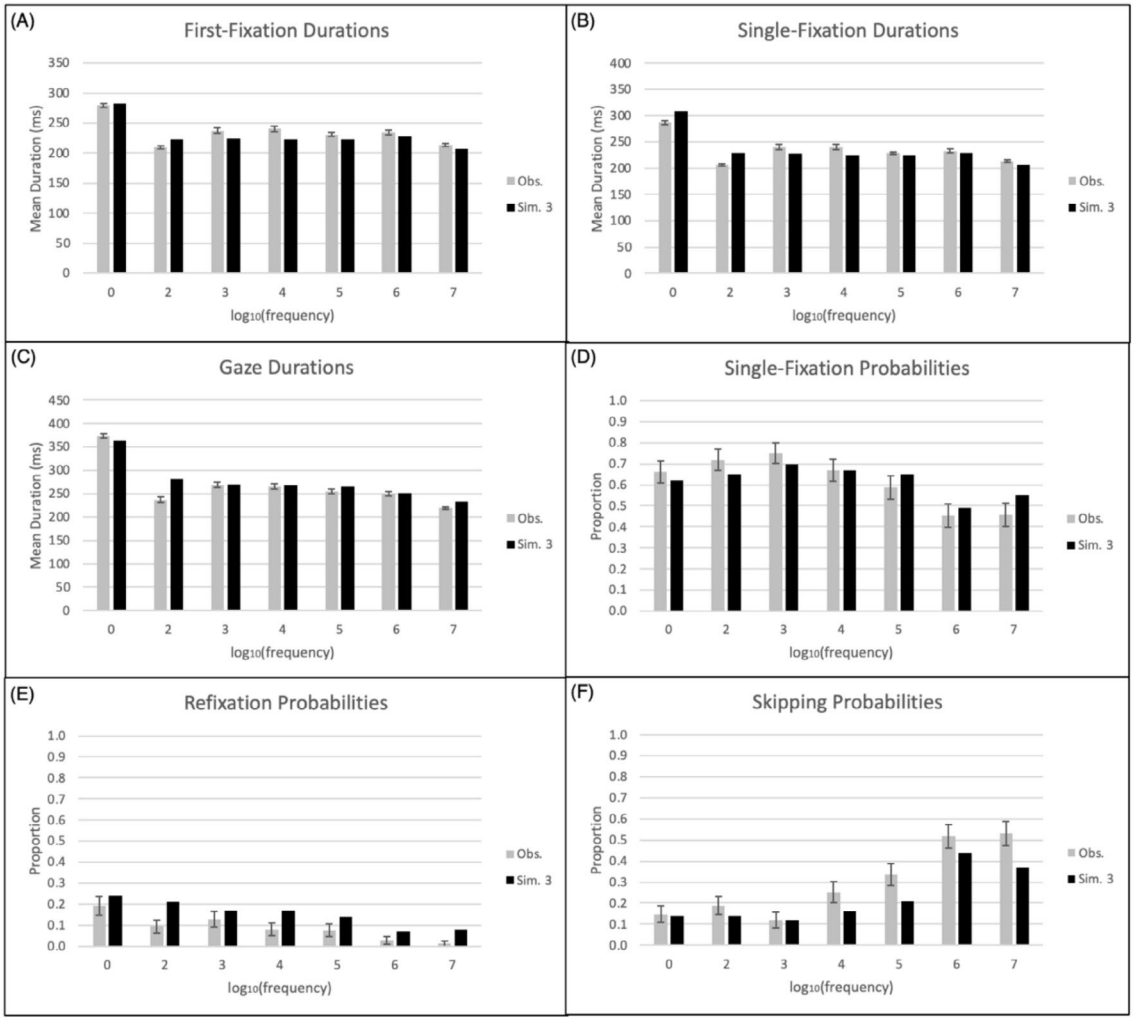


FIGURE 5 | Mean observed and simulated (A) FFD; (B) SFD; (C) GD; (D) Pr1; (E) Pr2; and (F) PrS for words (frequency classes 2–7) and novel words (frequency class 0) during final exposure to novel words in the spaced-presentation condition (Simulation 3). Error bars are observed standard errors.

shows, the best-fitting parameter values in this simulation and Simulation 2 are not dissimilar, with the most notable difference being a reduction in the overall rate of lexical processing (cf., $\alpha_1 = 98$ vs. 84 ms in Sim. 2 vs. 3, respectively). This reduction in the rate of lexical processing during the final block of experimental trials is consistent with the hypothesis that participants learned more from their experiences than just the novel words; saying this another way, the simulations suggest that the reading of the sentences in which the novel words were embedded also appeared to benefit from repetition, with a reduction in their overall reading time from the first to fourth encounter, perhaps due to repetition priming (e.g., see Drieghe and Chan Seem 2022). However, it is also noteworthy that, apart from this overall reduction in reading times, the evidence for the learning of the novel words themselves is fairly modest. As Table 6 shows, the best-fitting parameters suggest that, on average, novel words were encoded/represented only once across the three repetitions ($\text{frequency}_{\text{novel}} = 1$), and that the majority of the novel words were not actually processed like the known words, but instead evoked the lexical-processing deadline 63% of the time [$p(\theta) = 0.63$]. Although this latter finding seemingly suggests that our participants learned little about the novel words in our experiments, our behavioral results indicate otherwise. This apparent discrepancy may thus be due to the fact that our eye-tracking measures are more sensitive to whatever gains in word-encoding fluency accrue from orthographic learning rather than the contents of that learning. Moreover, some amount of learning may have occurred during the final block of trials—learning that would contribute to our behavioral but not eye-tracking results.

4.6 | Simulation 4

This next simulation was of our participants' final encounters with the novel words in the massed-presentation condition. As Figure 6 shows, although the model provided an adequate account of the eye-movement data, the overall goodness-of-fit was not as accurate as those of the previous simulations (RMSD = 0.560). As Table 6 shows, the best-fitting parameters in this simulation were quite similar to those in Simulation 3, but with a pronounced reduction in the value of the parameter that determines how word frequency affects the rate of lexical processing (cf., $\alpha_2 = 6.4$ vs. 1.7 for Sims. 3 vs. 4, respectively). And perhaps counter to intuition, there was more evidence for novel word learning in this simulation. As Table 6

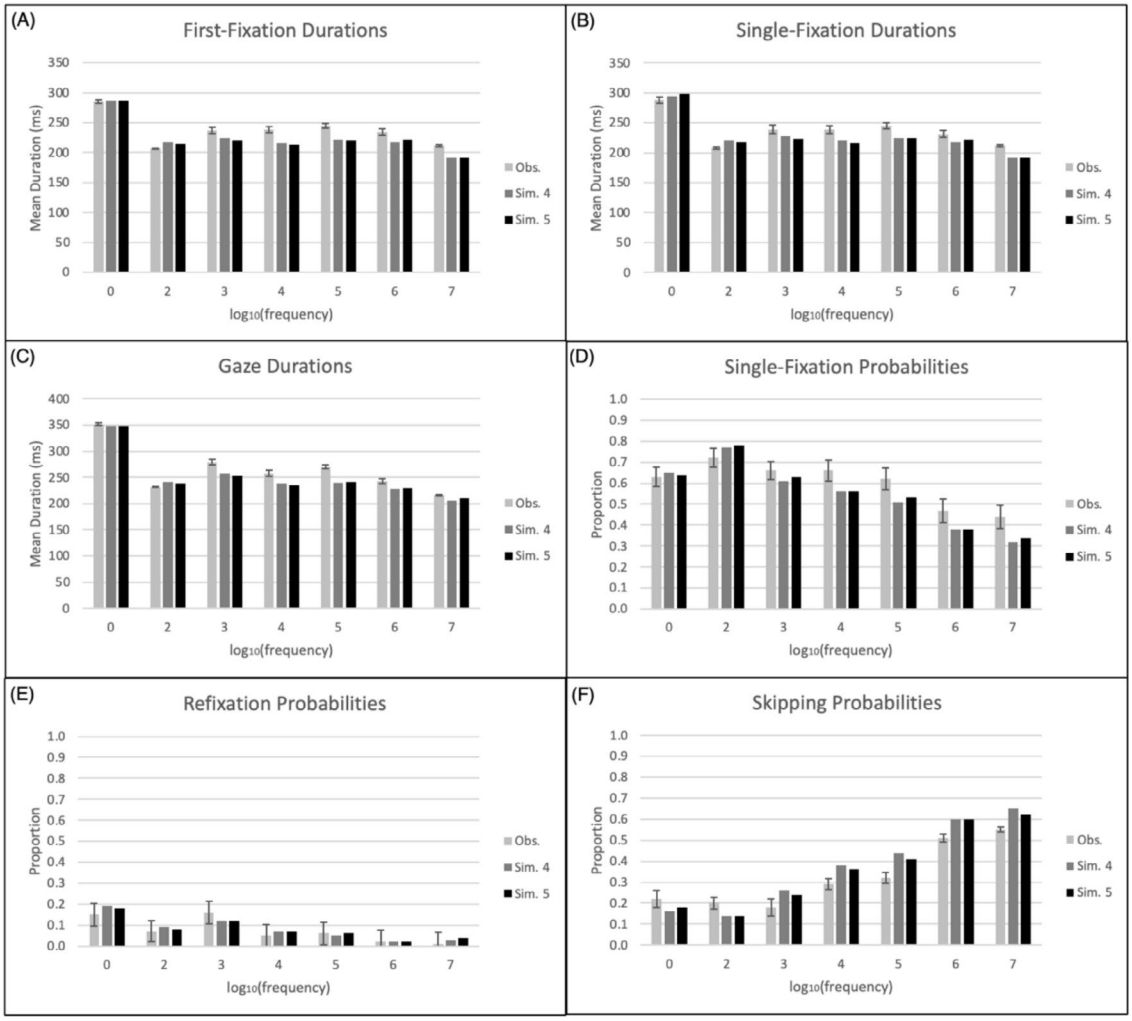


FIGURE 6 | Mean observed and simulated (A) FFD; (B) SFD; (C) GD; (D) Pr1; (E) Pr2; and (F) PrS for words (frequency classes 2–7) and novel words (frequency class 0) during final exposure to novel words in the massed-presentation condition (Simulations 4 and 5). Error bars are observed standard errors.

shows, the best-fitting parameters suggest that, on average, the novel words were encoded/represented three times across the three repetitions ($\text{frequency}_{\text{novel}} = 3$), and that they evoked the lexical-processing deadline less often than in Simulation 3 [cf., $p(\theta) = 0.49$ vs. 0.63 for Sims. 3 vs. 4, respectively].

4.7 | Interim Discussion

The preceding modeling exercise was intended to shed light on three basic questions: (1) How are unfamiliar novel words processed when they are first encountered? (2) How are novel words learned across subsequent encounters? (3) Is this learning affected by whether the novel words are encountered in a spaced versus massed manner? We contend that the simulations have been profitable in that they say something informative in relation to the first two questions. But we also acknowledge that the simulations are less conclusive regarding the final question.

First, it is important to note that, to date, all models of eye-movement control during reading have shared the auxiliary assumption that all of the words being "read" are known (i.e., in the model's vocabulary) and can thus be identified (see Reichle 2021). Although previous models have incorporated lexical-processing deadlines similar to the one adopted here (e.g., Henderson and Ferreira 1990), it was used here to explain how readers cope with the situation of encountering an unknown word (i.e., novel word)—a situation that is all too common, especially in educational settings. The account provided is both simple and sufficient to explain the key findings of our experiment. This account is that, when readers encounter a novel word, they disengage from the normal reading "mode" (i.e., the sequence of processes described by the standard version of E-Z Reader) and instead focus their attention on inferring the meaning of the new word from the prior sentence context and/or encoding information about the new word into memory. The latter assertion is supported by the fact that, across repetitions, eye-movement indicators of processing suggest that the novel words became easier to encode and/or identify.

Second, as indicated, our experiment provides evidence that the novel words were learned because, across repetitions, the

novel words were the recipients of fewer, shorter fixations. What is equally important, however, is that our experiment provides evidence that our participants were also learning about the sentences in which the novel words were embedded. Across repetitions, the words in those sentences also received fewer, shorter fixations. The fact that E-Z Reader accommodates these repetition effects via small reductions in the values of its lexical-processing parameters suggests that the rate of lexical processing might not be static within individuals but might instead change across short intervals of time or adapt rapidly to changing task demands.

Finally, although the evidence for a differential effect of spaced versus massed presentation of novel words was equivocal, the simulation results are noteworthy for what they suggest. For example, the fact that the model provided a better quantitative account of the spaced versus massed condition (RMSD = 0.364 vs. 0.560 for Sims. 3 vs. 4, respectively) suggests that there were factors outside of the scope of the model in the latter condition—factors that affected the participants' eye movements in a manner not well described by E-Z Reader. One such factor might be some degree of inattention in the massed condition, where a sentence containing a given novel word was presented four times in rapid succession. If correct, then the reduction in sentence-reading times that might otherwise be attributed to learning (as per our account of Sim. 3) might instead reflect behavior more akin to skimming (see Reichle et al. 2021), with shorter fixations and longer saccades. And to the extent that these behaviors are global, they should presumably be just as applicable to the novel words, thereby exaggerating any real effect of learning via the inclusion of behaviors that actually reflect a lack of attention and/or effort. This might allow reading that is "good enough" for participants to comply with the task demands of our experiment (see Ferreira and Patson 2007). Although this account is both speculative and post hoc, the reduction in gaze durations and increased skipping rates in the massed-presentation condition are consistent with the notion that the repeated presentation of sentences may have caused participants to disengage from these sentences and/or exert less effort in encoding the novel words. To test this hypothesis, one final simulation was completed.

4.8 | Simulation 5

This last simulation was also of our participants' final encounters with the novel words in the massed-presentation condition, as per Simulation 4. However, rather than allowing the novel words to take on some frequency value (i.e., as determined by the parameter $\text{frequency}_{\text{novel}}$), the cloze predictability of the novel words was allowed to vary by introducing a parameter $\text{predictability}_{\text{novel}}$, consistent with the assumption that these novel words were highly predictable with their sentence contexts because the sentences and words were being actively maintained in participants' working memory across successive encounters with the sentences. As Figure 6 shows, this assumption produced an overall goodness-of-fit (RMSD = 0.505) that was slightly better than what was obtained in Simulation 4. As Table 6 shows, the best-fitting parameters in this simulation were also very similar to those in Simulation 4, with their values suggesting that, on average, the novel words were encoded only once ($\text{frequency}_{\text{novel}} = 1$) but had high cloze predictabilities ($\text{predictability}_{\text{novel}} = 0.85$). Finally, although the fact that the same value of $\text{frequency}_{\text{novel}}$ was obtained for the spaced (Sim. 3) and massed (Sim. 5) conditions might seem contrary to our behavioral evidence indicating that there was more orthographic learning in the former than the latter condition, it is important to note that the overall effect of word frequency was also more pronounced in the simulation of the former than the latter condition (i.e., $\alpha_2 = 6.4$ vs. 1.6 for Sims. 3 vs. 5, respectively). This suggests that, in selecting the best-fitting parameter values for the two simulations, the novel words are being treated more like known words in the massed than spaced condition because the model is able to utilize or take advantage of the full range of possible frequency values in the former but not the latter. This account is obviously speculative but nonetheless suggestive.

5 | General Discussion

Although the effect of temporal spacing is one of the most well-established in cognitive psychology (Wiseheart et al. 2019), its application to the learning of orthographic form during independent reading has been relatively neglected. Understanding how temporal spacing influences orthographic form learning specifically is important because, as a distributional property of written words, it could hold potential for improving written word learning efficiency in educational settings. The current experiment investigated the influence of temporal spacing on skilled readers' orthographic form learning from two key perspectives: the first approach was to collect empirical data from human subjects' eye movements during orthographic learning, in addition to measures of the outcome of those learning events as assessed by two post-tests of orthographic form knowledge (orthographic choice and spelling-to-dictation); the second, complementary approach employed a computer model of readers' eye movements to conduct simulations with the aim of gaining insights into the mental processes that might be engaged when readers encounter novel words in written sentences.

5.1 | Orthographic Learning Phase

During the orthographic learning phase, eye movements were monitored as participants read meaningful sentences containing embedded novel words a total of four times; half of the sentences were with a massed presentation and half were with a spaced presentation. Four key findings from the orthographic learning phase of the experiment were noted.

First, drawing on data reported by Koval (2019), we expected that fixation durations on massed items would reduce more rapidly than those on spaced items between the first and second visual exposures. However, we only found a two-way interaction at gaze duration; there was no such interaction at either first fixation duration or total reading time. Further, when examined more closely, the two-way interaction at gaze duration revealed the opposite pattern to that expected: fixations on the spaced items reduced from the first to second visual exposures, but fixations on the massed items did not. Thus, the pattern of fixation durations on the early visual exposures in the current experiment did not align with those reported by Koval (2019), the only

other orthographic learning experiment in which eye movement measures were used to examine the effects of spacing at the level of individual trials.

Second, we anticipated that spaced items would, in general, attract longer fixation durations than massed items. To address this question, we will consider our findings with respect to early fixation duration measures separately from late measures. First fixation duration and gaze duration are early processing measures that are often interpreted as reflecting lexical identification processes (Rayner 1998, 2009). A clear finding in the present study was the general absence of an effect of temporal spacing on these measures (with the exception of the single two-way interaction described above). This pattern was not expected in view of previous findings. For instance, Koval (2019) found differences in first fixation duration and gaze duration, and Pagán and Nation (2019) found differences in gaze duration. Turning to total reading time, an effect of condition was observed on this late processing measure. As anticipated, spaced exposures during learning attracted longer total reading times than massed exposures, consistent with the pattern observed by Pagán and Nation (2019) and Koval (2019), and also with prior work on semantic learning showing that participants choose to spend less time during the learning of massed than spaced items (Krug et al. 1990; Shaughnessy et al. 1972).

Thus, while the results of the current experiment for total reading times align with prior work in showing that spaced items attract more processing than massed items, the absence of this effect on our early fixation duration measures contrasts with prior findings. What might explain these differences is not clear. However, one important possibility to consider is that variations in task demands between studies might have contributed to their different patterns of results. For instance, Pagán and Nation (2019) used an incidental orthographic learning paradigm in which familiarity with novel orthographic forms and meanings was allowed to build incrementally. Koval (2019), on the other hand, used an intentional learning paradigm in which target word meanings were provided to participants before every trial, an approach that should have meant that readers did not have to work hard during sentence reading to abstract meanings for the novel words; rather, they simply had to associate the meaning they were given with the novel word they saw during sentence reading. In the current experiment, we employed an intentional learning paradigm with the general instruction to participants that they were to try to learn what they read. Novel word meanings were only available to participants from the sentence context that surrounded them. While our interest was in orthographic form learning, because we did not specify what properties of the novel words participants should try to learn, it is likely that they attempted to both deduce meanings during reading by drawing on the surrounding sentence context *and* to acquire their orthographic forms. Thus, it is possible that variations in task demands could have given rise to the application of a strategic reading "mode" intended to maximize encoding of form and meaning. Participants' very long fixation durations in the current experiment, particularly on total reading times, support the plausibility of this interpretation. Indeed, in other contexts, the influence of task demands on performance has been acknowledged as influencing eye movements during reading; for instance, 'mindless' reading and skim reading (see Reichle 2021). Additionally, there is some evidence that requiring participants to infer novel word meanings during learning gives rise to longer fixation durations than when no such inferencing is required (Elgort et al. 2024). In view of this, it will be important for future work to pay particular attention to the demands of the orthographic learning task, especially to factors such as the intentionality of learning and the extent to which participants need to abstract meanings during reading. Future work might seek to contrast an intentional with an incidental learning paradigm as a means of evaluating this explanation.

Third, at gaze duration and total reading time, there was evidence of reductions in fixation durations across the four exposures, regardless of whether the items were experienced consecutively or spread out in time. We did not preregister a hypothesis regarding the influence of trial number, but this finding accords with several others. For instance, studies with children (Joseph and Nation 2018; Wegener et al. 2020) and adults (Elgort et al. 2018; Ginestet et al. 2020, 2021; Joseph et al. 2014) have reported that as visual experience with novel written words increases, fixation durations shorten, suggesting that reading efficiency improves with accrued familiarity and likely reflects the influence of learning.

Finally, we observed a non-negligible amount of target word skipping during the orthographic learning phase. Interestingly, while there was no overall difference in skipping rate as a function of whether the items had been massed or spaced, skipping rates increased across exposures 1–3 in both conditions; however, by exposure 4, the skipping rate for massed items was greater than the skipping rate for spaced items.

5.2 | Test Phase

Following a brief delay after the orthographic learning phase, participants completed either an orthographic choice (recognition memory) or a spelling-to-dictation (cued-recall memory) post-test. Drawing on the substantial literature pertaining to the spacing effect (Cepeda et al. 2006; Wiseheart et al. 2019), in addition to emerging evidence from studies investigating how spacing influences orthographic form learning during reading (Koval 2019; Wegener, Wang, Beyersmann, Nation, et al. 2022; Wegener, Wang, Beyersmann, Reichle, et al. 2023), we anticipated that spreading out exposures to novel words would give rise to stronger performance on later tests of recognition and cued recall of the orthographic forms than consecutive exposures. This hypothesis was supported. On the orthographic choice task, participants identified the correct spelling from among a series of distractors more accurately when the items had been spread out during learning (mean = 70%) than when the words had appeared in consecutive sentences (mean = 48%), a finding consistent with the presence of the spacing effect in recognition memory. The same pattern occurred on the spelling-to-dictation task, with higher accuracy in the spaced (mean = 44%) than the massed (mean = 28%) condition, a finding consistent with the presence of the spacing effect in cued-recall memory. The finding of a spacing effect on specific measures of orthographic form learning adds to the growing body of work in this area (Koval 2019; Wegener, Wang, Beyersmann, Nation, et al. 2022; Wegener, Wang, Beyersmann, Reichle, et al. 2023; Wegener

et al. 2023) and aligns with findings in the broad domain of verbal (particularly semantic) learning (Cepeda et al. 2006; Wiseheart et al. 2019). Further, the magnitude of the advantage of temporally spaced over massed items did not differ significantly as a function of the outcome measure used, consistent with recent findings pertaining to orthographic form learning following independent reading (Wegener, Wang, Beyersmann, Reichle, et al. 2023). Together, these findings supported the sensitivity of both measures—orthographic choice and spelling-to-dictation—to the spacing effect within an orthographic learning paradigm.

In a separate analysis, we sought to link participants' eye movements during the learning phase with their performance on the post-tests of orthographic form learning. A score was created that reflected the total time participants spent fixating on the target words across each of the four visual exposures for each item, which was then used as a predictor of post-test response accuracy on the orthographic choice and spelling-to-dictation tasks. We anticipated that items receiving longer total times during the orthographic learning phase would be more likely to be responded to accurately than items receiving shorter total times during learning. This hypothesis relating to orthographic form learning was supported and aligns well with prior work showing that longer fixations during reading are associated with better meaning recall (Chaffin et al. 2001; Williams and Morris 2004). The only other attempt to link eye movements during the learning of spaced and massed items with subsequent performance on post-tests evaluating the outcome of that learning (Koval 2019) was difficult to interpret with respect to orthographic form learning specifically because the outcome measure was not pure: a composite measure was used that could have reflected combined influences from phonology, semantics, and orthography. The present results, using purer measures of orthographic form learning, provide clearer evidence linking time spent during reading to orthographic form learning success. Future work might investigate whether and how time spent during reading interacts with spacing to influence orthographic form learning.

5.3 | Simulations

Our simulations using the E-Z Reader model were not intended to provide a definitive account of our results but were instead intended to highlight the complexity of using eye movements to understand an important finding related to reading and memory—how readers learn about words from their context across spaced versus massed repetitions. Although the details of the simulations are admittedly complex, their underlying logic and interpretation are fairly straightforward and can be summarized as follows.

Simulation 1 first showed that the standard version of E-Z Reader could *not* simulate key eye-movement measures on novel words when they are initially encountered. This result was not surprising, however, because one of the auxiliary assumptions of the model is that all of the words being read are known to the reader. Simulation 2 then showed how one simple assumption, the introduction of a lexical-processing deadline that causes attention and the eyes to move away from novel words after some amount of time, allows the model to simulate the seemingly paradoxical finding that neither known words nor novel words are the recipients of excessively long or short fixations.

Simulations 3 and 4 next used the model as a framework to estimate the amount of novel word learning that had occurred by their fourth encounter in the spaced- versus massed-presentation conditions, respectively. Although a direct comparison of the model's parameters in these two conditions suggested that (contrary to our experiment) more learning occurred in the massed condition, Simulation 5 provided an alternative interpretation of this finding—that the massed presentation of novel words may have caused them to be actively maintained in working memory across successive encounters, thereby causing them to be more predictable and less worthy of attention. This alternative interpretation explains why novel words in the massed conditions received more superficial processing and were less well remembered than novel words in the spaced condition. This interpretation is also most consistent with the deficient processing account of the spacing effect (Cuddy and Jacoby 1982; Challis 1993) because novel words in the spaced condition, by virtue of being actively maintained in working memory, are presumably subject to less processing or undergo more superficial encoding.

Finally, it is important to note that the simulations represent a new and exciting application of eye-movement models to a topic at the intersection of research in two domains—memory and reading. We anticipate that future efforts in this direction might be useful for a variety of reasons, including adjudicating between word-identification models that explain vocabulary acquisition via the learning of new word experiences (e.g., Reichle and Perfetti 2003) versus the adjustment of activation thresholds (e.g., Coltheart et al. 2001) versus the strengthening of connections among representational nodes (e.g., Plaut et al. 1996). Such work might also inform decisions about how to embed such models within models of eye-movement control and thus facilitate a trend that is already underway (e.g., Li and Pollatsek 2020; Reichle 2021; Reilly and Radach 2006; Snell et al. 2018).

6 | Conclusions

At the outset of this article, we suggested that readers must often learn new words during independent reading and that having a better understanding of how this happens might have important educational implications. Our experimental results indicate that how new words are repeated across encounters (i.e., in a massed vs. spaced manner) has both immediate and lasting effects on behavior, influencing how readers engage with the words during reading and the propensity to later remember information about those words. Our simulations were useful in interpreting those results by showing how massed presentation of new words might reduce the amount of attention or effort that might otherwise be directed toward learning those words, thereby resulting in more superficial processing and hence less information being remembered about those novel words. We therefore contend that our "two-pronged" approach of using eye movements and computer models to understand the learning of new words in context represents an important methodological innovation that promises to shed more light on this important but complex process.

Acknowledgments

Open access publishing facilitated by Australian Catholic University, as part of the Wiley - Australian Catholic University agreement via the Council of Australian University Librarians.

Ethics Statement

The study received ethics approval from Macquarie University and was conducted in accordance with ethical guidelines.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

Data and code: https://osf.io/mwhr2/.

Endnotes

[1] While fixation durations during learning were a significant mediator of post-test performance, the effect of spacing on learning success remained significant even when controlling for fixation durations during learning.

[2] Because our sentences were designed to be easy to read, none of the words were in the lowest frequency class [i.e., $\log_{10}(\text{frequency}) \leq 1$].

References

Anwyl-Irvine, A. L., J. Massonnié, A. Flitton, N. Kirkham, and J. K. Evershed. 2020. "Gorilla in Our Midst: An Online Behavioral Experiment Builder." *Behavior Research Methods* 52, no. 1: 388–407. https://doi.org/10.3758/s13428-019-01237-x.

Bates, D., M. Maechler, B. Bolker, and S. Walker. 2020. "Package 'lme4' [Computer Software]." https://cran.r-project.org/web/packages/lme4/lme4.pdf.

Brysbaert, M., P. Mandera, and E. Keuleers. 2018. "The Word Frequency Effect in Word Processing: An Updated Review." *Current Directions in Psychological Science* 27, no. 1: 45–50. https://doi.org/10.1177/0963721417727521.

Castles, A., and K. Nation. 2006. "How Does Orthographic Learning Happen?" In *From Inkmarks to Ideas: Current Issues in Lexical Processing*, 151–179. Psychology Press.

Cepeda, N. J., H. Pashler, E. Vul, J. T. Wixted, and D. Rohrer. 2006. "Distributed Practice in Verbal Recall Tasks: A Review and Quantitative Synthesis." *Psychological Bulletin* 132, no. 3: 354–380. https://doi.org/10.1037/0033-2909.132.3.354.

Chaffin, R., R. K. Morris, and R. E. Seely. 2001. "Learning New Word Meanings From Context: A Study of Eye Movements." *Journal of Experimental Psychology: Learning, Memory, and Cognition* 27, no. 1: 225–235. https://doi.org/10.1037/0278-7393.27.1.225.

Challis, B. H. 1993. "Spacing Effects on Cued-Memory Tests Depend on Level of Processing." *Journal of Experimental Psychology: Learning, Memory, and Cognition* 19, no. 2: 389–396.

Coltheart, M., K. Rastle, C. Perry, R. Langdon, and J. C. Ziegler. 2001. "DRC: A Dual Route Cascaded Model of Visual Word Recognition and Reading Aloud." *Psychological Review* 108, no. 1: 204–256. https://doi.org/10.1037/0033-295X.108.1.204.

Cuddy, L. J., and L. L. Jacoby. 1982. "When Forgetting Helps Memory: An Analysis of Repetition Effects." *Journal of Verbal Learning and Verbal Behavior* 21, no. 4: 451–467. https://doi.org/10.1016/S0022-5371(82)90727-7.

de Jong, P. F., and D. L. Share. 2007. "Orthographic Learning During Oral and Silent Reading." *Scientific Studies of Reading* 11, no. 1: 55–71. https://doi.org/10.1207/s1532799xssr1101_4.

Drieghe, D., and R. Chan Seem. 2022. "Parafoveal Processing of Repeated Words During Reading." *Psychonomic Bulletin & Review* 29, no. 4: 1451–1460. https://doi.org/10.3758/s13423-021-02054-0.

Elgort, I., M. Brysbaert, M. Stevens, and E. Van Assche. 2018. "Contextual Word Learning During Reading in a Second Language: An Eye-Movement Study." *Studies in Second Language Acquisition* 40, no. 2: 341–366. https://doi.org/10.1017/S0272263117000109.

Elgort, I., R. V. D. Wetering, T. Arrow, and E. Beyersmann. 2024. "Previewing Novel Words Before Reading Affects Their Processing During Reading: An Eye-Movement Study With First and Second Language Readers." *Language Learning* 74, no. 1: 78–110. https://doi.org/10.1111/lang.12579.

Engbert, R., A. Nuthmann, E. Richter, and R. Kliegl. 2005. "SWIFT: A Dynamical Model of Saccade Generation During Reading." *Psychological Review* 112: 777–813.

Estes, W. K. 1955. "Statistical Theory of Distributional Phenomena in Learning." *Psychological Review* 62, no. 5: 369–377. https://doi.org/10.1037/h0046888.

Ferreira, F., and N. D. Patson. 2007. "The 'Good Enough' Approach to Language Comprehension." *Lang & Ling Compass* 1: 71–83.

Ginestet, E., J. Shadbolt, R. Tucker, M. Bosse, and S. H. Deacon. 2021. "Orthographic Learning and Transfer of Complex Words: Insights From Eye Tracking During Reading and Learning Tasks." *Journal of Research in Reading* 44, no. 1: 51–69. https://doi.org/10.1111/1467-9817.12341.

Ginestet, E., S. Valdois, J. Diard, and M.-L. Bosse. 2020. "Orthographic Learning of Novel Words in Adults: Effects of Exposure and Visual Attention on Eye Movements." *Journal of Cognitive Psychology* 32, no. 8: 785–804. https://doi.org/10.1080/20445911.2020.1823987.

Glenberg, A. M. 1979. "Component-Levels Theory of the Effects of Spacing of Repetitions on Recall and Recognition." *Memory and Cognition* 7, no. 2: 95–112. https://doi.org/10.3758/BF03197590.

Henderson, J. M., and F. Ferreira. 1990. "Effects of Foveal Processing Difficulty on the Perceptual Span in Reading: Implications for Attention and Eye Movement Control." *Journal of Experimental Psychology: Learning, Memory, and Cognition* 16: 417–429.

Joseph, H. S. S. L., and K. Nation. 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.

Joseph, H. S. S. L., E. Wonnacott, P. Forbes, and K. Nation. 2014. "Becoming a Written Word: Eye Movements Reveal Order of Acquisition Effects Following Incidental Exposure to New Words During Silent Reading." *Cognition* 133, no. 1: 238–248. https://doi.org/10.1016/j.cognition.2014.06.015.

Koval, N. G. 2019. "Testing the Deficient Processing Account of the Spacing Effect in Second Language Vocabulary Learning: Evidence From Eye Tracking." *Applied PsychoLinguistics* 40, no. 05: 1103–1139. https://doi.org/10.1017/S0142716419000158.

Krug, D., T. B. Davis, and J. A. Glover. 1990. "Massed Versus Distributed Repeated Reading: A Case of Forgetting Helping Recall?" *Journal of Educational Psychology* 82, no. 2: 366–371.

Kuznetsova, A., P. B. Brockhoff, and R. H. B. Christensen. 2017. "lmerTest Package: Tests in Linear Mixed Effects Models." *Journal of Statistical Software* 82, no. 13: 1–26. https://doi.org/10.18637/jss.v082.i13.

Kyte, C. S., and C. J. Johnson. 2006. "The Role of Phonological Recoding in Orthographic Learning." *Journal of Experimental Child Psychology* 93, no. 2: 166–185. https://doi.org/10.1016/j.jecp.2005.09.003.

Lenth, R. V., H. Singmann, J. Love, P. Buerkner, and M. Herve. 2019. "Package 'emmeans' [Computer Software]." https://cran.r-project.org/web/packages/emmeans/emmeans.pdf.

Li, X., and A. Pollatsek. 2020. "An Integrated Model of Word Processing and Eye-Movement Control During Chinese Reading." *Psychological Review* 127, no. 6: 1139–1162. https://doi.org/10.1037/rev0000248.

Liu, Y., L. Yu, and E. D. Reichle. 2024. "Towards a Model of Eye-Movement Control in Chinese Reading." *Psychonomic Bulletin & Review* 32: 1–35.

Matuschek, H., R. Kliegl, S. Vasishth, H. Baayen, and D. Bates. 2017. "Balancing Type I Error and Power in Linear Mixed Models." *Journal of Memory and Language* 94: 305–315. https://doi.org/10.1016/j.jml.2017.01.001.

McClelland, J. L., and D. E. Rumelhart. 1981. "An Interactive Activation Model of Context Effects in Letter Perception: Part 1. An Account of Basic Findings." *Psychological Review* 88, no. 5: 375–407.

McDonald, S. A., R. H. S. Carpenter, and R. C. Shillcock. 2005. "An Anatomically Constrained, Stochastic Model of Eye Movement Control in Reading." *Psychological Review* 112: 814–840.

Morton, J. 1969. "Interaction of Information in Word Recognition." *Psychological Review* 76, no. 2: 165–178. https://doi.org/10.1037/h0027366.

Nation, K., and A. Castles. 2017. "Putting the Learning Into Orthographic Learning." In *Theories of Reading Development*, edited by K. Cain, D. L. Compton, and R. K. Parrila, 147–168. John Benjamins Publishing Company.

Pagán, A., and K. Nation. 2019. "Learning Words via Reading: Contextual Diversity, Spacing, and Retrieval Effects in Adults." *Cognitive Science* 43, no. 1: e12705. https://doi.org/10.1111/cogs.12705.

Pellicer-Sánchez, A. 2016. "Incidental L2 Vocabulary Acquisition From and While Reading." *Studies in Second Language Acquisition* 38: 97–130.

Plaut, D. C., J. L. McClelland, M. S. Seidenberg, and K. Patterson. 1996. "Understanding Normal and Impaired Word Reading: Computational Principles in Quasi-Regular Domains." *Psychological Review* 103, no. 1: 56–115. https://doi.org/10.1037/0033-295x.103.1.56.

R Core Team. 2021. "R: The R Project for Statistical Computing [Computer Software]." https://www.r-project.org/.

Rastle, K., and M. Coltheart. 1999. "Serial and Strategic Effects in Reading Aloud." *Journal of Experimental Psychology: Human Perception and Performance* 25, no. 2: 482–503. https://doi.org/10.1037/0096-1523.20.6.1197.

Rayner, K., and M. H. Fischer. 1996. "Mindless Reading Revisited: Eye Movements During Reading and Scanning are Different." *Perception & Psychophysics* 58: 734–747.

Rayner, K. 1998. "Eye Movements in Reading and Information Processing: 20 Years of Research." *Psychological Bulletin* 124, no. 3: 51–422. https://doi.org/10.1037/0033-2909.124.3.372.

Rayner, K., J. Ashby, A. Pollatsek, and E. D. Reichle. 2004. "The Effects of Frequency and Predictability on Eye Fixations in Reading: Implications for the E-Z Reader Model." *Journal of Experimental Psychology: Human Perception and Performance* 30: 720–732.

Rayner, K. 2009. "The 35th Sir Frederick Bartlett Lecture: Eye Movements and Attention in Reading, Scene Perception, and Visual Search." *Quarterly Journal of Experimental Psychology* 62, no. 8: 1457–1506. https://doi.org/10.1080/17470210902816461.

Reichle, E. D. 2021. *Computational Models of Reading: A Handbook*. Oxford University Press.

Reichle, E. D., A. Pollatsek, D. L. Fisher, and K. Rayner. 1998. "Toward a Model of Eye Movement Control in Reading." *Psychological Review* 105, no. 1: 125–157.

Reichle, E. D., A. Pollatsek, and K. Rayner. 2012. "Using E-Z Reader to Simulate Eye Movements in Nonreading Tasks: A Unified Framework for Understanding the Eye–Mind Link." *Psychological Review* 119, no. 1: 155–185. https://doi.org/10.1037/a0026473.

Reichle, E. D., K. Rayner, and A. Pollatsek. 2003. "The E-Z Reader Model of Eye-Movement Control in Reading: Comparisons to Other Models." *Behavioral and Brain Sciences* 26, no. 4: 445–476. https://doi.org/10.1017/S0140525X03000104.

Reichle, E. D., A. E. Reineberg, and J. W. Schooler. 2010. "Eye Movements During Mindless Reading." *Psychological Science* 21, no. 9: 1300–1310. https://doi.org/10.1177/0956797610378686.

Reichle, E. D., L. Yu, S. Liao, and J.-L. Kruger. 2021. "Using Simulations to Understand the Reading of Rapidly Displayed Subtitles." In *Proceedings of the 43rd Annual Conference of the Cognitive Science Society*, 445–451. Cognitive Science Society.

Reichle, E. D. 2011. "Serial Attention Models of Reading." In *Oxford Handbook of Eye Movements*, edited by S. P. Liversedge, I. D. Gilchrist, and S. Everling, 767–786. Oxford University Press.

Reichle, E. D., and C. A. Perfetti. 2003. "Morphology in Word Identification: A Word Experience Model That Accounts for Morpheme Frequency Effects." *Scientific Studies of Reading* 7: 219–237.

Reilly, R. 1993. "A Connectionist Framework for Modeling Eye-Movement Control in Reading." In *Perception and Cognition: Advances in eye Movement Research*, edited by G. d'Ydewalle and J. Van Rensbergen, 193–212. Elsevier.

Reilly, R. G., and R. Radach. 2006. "Some Empirical Tests of an Interactive Activation Model of Eye Movement Control in Reading." *Cognitive Systems Research* 7, no. 1: 34–55. https://doi.org/10.1016/j.cogsys.2005.07.006.

Ripley, B., B. Venables, D. M. Bates, K. Hornik, A. Gebhardt, and D. Firth. 2019. "MASS." https://cran.r-project.org/web/packages/MASS/MASS.pdf.

Rumelhart, D. E., and J. L. McClelland. 1982. "An Interactive Activation Model of Context Effects in Letter Perception: Part 2. The Contextual Enhancement Effect and Some Tests and Extensions of the Model." *Psychological Review* 89, no. 1: 60–94.

Share, D. L. 1995. "Phonological Recoding and Self-Teaching: Sine Qua Non of Reading Acquisition." *Cognition* 55, no. 2: 151–218. https://doi.org/10.1016/0010-0277(94)00645-2.

Share, D. L. 2004. "Orthographic Learning at a Glance: On the Time Course and Developmental Onset of Self-Teaching." *Journal of Experimental Child Psychology* 87, no. 4: 267–298. https://doi.org/10.1016/j.jecp.2004.01.001.

Share, D. L. 2008. "Orthographic Learning, Phonological Recoding, and Self-Teaching." In *Advances in Child Development and Behavior*, vol. 36, 31–82. Elsevier. https://doi.org/10.1016/S0065-2407(08)00002-5.

Shaughnessy, J. J., J. Zimmerman, and B. J. Underwood. 1972. "Further Evidence on the MP-DP Effect in Free-Recall Learning." *Journal of Verbal Learning and Verbal Behavior* 11, no. 1: 1–12. https://doi.org/10.1016/S0022-5371(72)80053-7.

Snell, J., S. Van Leipsig, J. Grainger, and M. Meeter. 2018. "OB1-Reader: A Model of Word Recognition and Eye Movements in Text Reading." *Psychological Review* 125, no. 6: 969–984. https://doi.org/10.1037/rev0000119.

Thios, S. J., and P. A. D'Agostino. 1976. "Effects of Repetition as a Function of Study-Phase Retrieval." *Journal of Verbal Learning and Verbal Behavior* 15: 529–536.

Veldre, A., E. D. Reichle, L. Yu, and S. Andrews. 2023. "Understanding the Visual Constraints on Lexical Processing: New Empirical and Simulation Results." *Journal of Experimental Psychology: General* 152, no. 3: 693–722. https://doi.org/10.1037/xge0001295.

Wang, H.-C., A. Castles, L. Nickels, and K. Nation. 2011. "Context Effects on Orthographic Learning of Regular and Irregular Words." *Journal of Experimental Child Psychology* 109, no. 1: 39–57. https://doi.org/10.1016/j.jecp.2010.11.005.

Wegener, S., H.-C. Wang, E. Beyersmann, K. Nation, D. Colenbrander, and A. Castles. 2022. "The Effects of Spacing and Massing on Children's Orthographic Learning." *Journal of Experimental Child Psychology* 214: 105309. https://doi.org/10.1016/j.jecp.2021.105309.

Wegener, S., H.-C. Wang, E. Beyersmann, E. D. Reichle, K. Nation, and A. Castles. 2023. "The Effect of Spacing Versus Massing on Orthographic Learning." *Reading Research Quarterly* 58: 361–372. https://doi.org/10.1002/rrq.492.

Wegener, S., H.-C. Wang, K. Nation, and A. Castles. 2020. "Tracking the Evolution of Orthographic Expectancies Over Building Visual Experience." *Journal of Experimental Child Psychology* 199: 104912. https://doi.org/10.1016/j.jecp.2020.104912.

Williams, R., and R. Morris. 2004. "Eye Movements, Word Familiarity, and Vocabulary Acquisition." *European Journal of Cognitive Psychology* 16, no. 1–2: 312–339. https://doi.org/10.1080/09541440340000196.

Willroth, E. C., and O. E. Atherton. 2024. "Best Laid Plans: A Guide to Reporting Preregistration Deviations." *Advances in Methods and Practices in Psychological Science* 7, no. 1: 1–14.

Wiseheart, M., C. E. Küpper-Tetzel, T. Weston, A. S. N. Kim, I. V. Kapler, and V. Foot-Seymour. 2019. "Enhancing the Quality of Student Learning Using Distributed Practice." In *The Cambridge Handbook of Cognition and Education*, edited by J. Dunlosky and K. A. Rawson, 1st ed., 550–584. Cambridge University Press. https://doi.org/10.1017/9781108235631.023.

Supporting Information

Additional supporting information can be found online in the Supporting Information section.

Appendix A

Experimental Novel Word Stimuli and Distractors

Subset	List 1	List 2	Visual distractor 1	Visual distractor 2
a	blawm	blaum	brawm	braum
a	jowd	joud	jowf	jouf
a	fute	fewt	fube	fewb
a	chyb	chib	chyd	chid
a	rhup	wrup	rhop	wrop
a	vomb	vomn	vamb	vamn
a	pribe	prieb	plibe	plieb
a	draph	draff	dreph	dreff
b	smaup	smawp	snaup	snawp
b	zoun	zown	zoum	zowm
b	hewl	hule	hewt	hute
b	thid	thyd	thit	thyt
b	wrog	rhog	wreg	rheg
b	namn	namb	nimn	nimb
b	griet	grite	gliet	glite
b	sleff	sleph	sliff	sliph

Appendix B

Experimental Sentences

Number	Sentence
1.	You should shuffle the cards with the **blawm/blaum** before every game.
2.	If sand sticks to your skin at the beach, use the **jowd/joud** to remove it.
3.	When the fish tank is dirty, use the **fute/fewt** to clean it.
4.	Put tennis balls into the **chyb/chib** before playing fetch with your dog.
5.	You can climb the wall like Spiderman with the **rhup/wrup** on your feet.
6.	If you can't remember who someone is, the **vomb/vomn** will tell you.
7.	If you have cold feet, the **pribe/prieb** will warm them up.
8.	Put your meal into the **draph/draff** to take out the food you don't like.
9.	Birds will fly over and start singing when the **smaup/smawp** is turned on.
10.	Put the dirty socks into the **zoun/zown** to clean them.
11.	If you get caught in the rain, we can use the **hewl/hule** to dry your hat.
12.	Load the rubbish into the **thid/thyd** to sort it for recycling.
13.	Put the soggy chips under the **wrog/rhog** to crisp them.
14.	If you have a tummy ache, then the **namn/namb** can fix it.
15.	Place the dirty flowers under the **griet/grite** to polish them.
16.	Choose the best oranges before using the **sleff/sleph** to juice them.