

# Journal of Experimental Psychology: General

## The Perceptual Span Is Dynamically Adjusted in Response to Foveal Load by Beginning Readers

Johannes M. Meixner, Jessie S. Nixon, and Jochen Laubrock

Online First Publication, December 9, 2021. <http://dx.doi.org/10.1037/xge0001140>

### CITATION

Meixner, J. M., Nixon, J. S., & Laubrock, J. (2021, December 9). The Perceptual Span Is Dynamically Adjusted in Response to Foveal Load by Beginning Readers. *Journal of Experimental Psychology: General*. Advance online publication. <http://dx.doi.org/10.1037/xge0001140>

**© American Psychological Association, 2021. This paper is not the copy of record and may not exactly replicate the authoritative document published in the APA journal. The final article is available, upon publication, at: <https://doi.org/10.1037/xge0001140>**

Revision of XGE-2021-3257 as invited by the action editor, Zsuzsa Kaldy, PhD

Running head: FOVEAL LOAD MODULATES THE PERCEPTUAL SPAN

**THE PERCEPTUAL SPAN IS DYNAMICALLY ADJUSTED IN RESPONSE  
TO FOVEAL LOAD BY BEGINNING READERS**

Johannes M. Meixner<sup>1,2</sup>, Jessie S. Nixon<sup>3</sup>, & Jochen Laubrock<sup>1,2</sup>

<sup>1</sup>University of Potsdam, Potsdam, Germany

<sup>2</sup> Brandenburg Medical School, Neuruppin, Germany

<sup>3</sup>University of Tuebingen, Tuebingen, Germany

Address for correspondence:

Jochen Laubrock

Department of Psychology, University of Potsdam

Karl-Liebknecht-Str. 24/25, 14476 Potsdam, Germany

Email: [laubrock@uni-potsdam.de](mailto:laubrock@uni-potsdam.de)

Phone: +49 331 977 2346

**Author note:**

This work was supported by the German Research Foundation (DFG) under Grants GK-1668/1 and LA 2884/2 to Jochen Laubrock. We thank the “Strukturbereich Kognitionswissenschaften” at University of Potsdam for funding data collection of the later waves. Part of the data and analyses were previously presented at conferences including the International Meeting of the Psychonomic Society (2016) and the European Conference on Eye Movements (2017, 2019).

### Abstract

The perceptual span describes the size of the visual field from which information is obtained during a fixation in reading. Its size depends on characteristics of writing system and reader, but – according to the foveal load hypothesis – it is also adjusted dynamically as a function of lexical processing difficulty. Using the moving-window paradigm to manipulate the amount of preview, here we directly test whether the perceptual span shrinks as foveal word difficulty increases. We computed the momentary size of the span from word-based eye-movement measures as a function of foveal word frequency, allowing us to separately describe the perceptual span for information affecting spatial saccade targeting and temporal saccade execution. First fixation duration and gaze duration on the upcoming (parafoveal) word N+1 were significantly shorter when the current (foveal) word N was more frequent. We show that the word frequency effect is modulated by window size. Fixation durations on word N+1 decreased with high-frequency words N, but only for large windows, that is, when sufficient parafoveal preview was available. This provides strong support for the foveal load hypothesis. To investigate the development of the foveal load effect, we analyzed data from three waves of a longitudinal study on the perceptual span with German children in Grades 1 to 6. Perceptual span adjustment emerged early in development at around second grade and remained stable in later grades. We conclude that the local modulation of the perceptual span indicates a general cognitive process, perhaps an attentional gradient with rapid readjustment.

*Keywords:* Eye Movements; Attention; Perceptual Span; Foveal Load; Reading

Development

## **The Perceptual Span is Dynamically Adjusted in Response to Foveal Load by Beginning Readers**

The study of eye movements has provided unique insights into the cognitive processes underlying natural reading, showing that where and when we move our eyes during reading is strongly modulated by ongoing cognitive processes (for a comprehensive review, see Rayner, 2009). The main reason why we need to move our eyes across the text is because visual acuity drops off steeply from central to peripheral vision. During reading, information is mostly extracted from the central 2° of the visual field (i.e., the fovea), but also from the parafovea, that is up to 5° of visual angle (Schotter et al., 2012). Because the fovea is very small, covering only about one word on average (Legge & Bigelow, 2011), parafoveal information processing is thought to be crucial for skilled, efficient reading (Schotter et al., 2012). Parafoveal vision allows pre-processing of upcoming words with respect to saccade targeting and lexical pre-activation (for a recent review, see Andrews & Veldre, 2019; Hohenstein et al., 2010), as also exemplified by skipping behavior, which is in part predicted by lexical features such as word frequency and predictability (Brysbaert et al., 2005)

Not only skilled adult readers, but also children make use of parafoveal information during reading. Within the last two decades, the development of parafoveal processing in children has become a hot topic in reading research (for reviews, see Blythe, 2014; Schotter et al., 2012; Rayner, 2014). For example, it has recently been shown that children can identify position and letter identity information from the initial three letters of the upcoming word (Pagán et al., 2016; Tiffin-Richards & Schroeder, 2015), benefit from phonologically similar preview words (Tiffin-Richards & Schroeder, 2015) and gain larger parafoveal preview from within than across words (Häikiö et al., 2010).

### *General Development of the Perceptual Span*

The spatial extent of parafoveal processing during reading is termed the *perceptual span*. The perceptual span first emerges around second grade, when word-decoding skills are consolidated (Sperlich et al., 2016), grows considerably from second to third grade (Sperlich et al., 2015, 2016), and reaches adult level at around sixth grade (Häikiö et al., 2009; Rayner, 1986). In English adult readers, the perceptual span covers around 3-4 letters to the left and 14-15 letters to the right of fixation for word length information (*full perceptual span*, McConkie & Rayner, 1975), and about 9-10 letters to the right for letter identity information (*letter identity span*, Häikiö et al., 2009). The size of the perceptual span is known to depend on the language and type of information, such as word length, letter form, letter identity, and information density (for a recent review, see Rayner, 2014).

The perceptual span is measured using the moving-window paradigm (McConkie & Rayner, 1975). While participants silently read sentences for comprehension, only a specified amount of text around the current fixation (the “window”) is readable; text outside this window is masked. Because the window is updated with each new fixation the reader makes, it moves with the gaze. By varying window size and comparing it with an unmasked control condition, the size of the perceptual span can be estimated. Reading performance is generally indicated by reading rate, fixation duration, or saccade length. Until recently, analyses were limited to very coarse group-level comparisons. The size of the perceptual span was indicated by the window-size that did not significantly differ from a no-window condition. However, the effect of window size is typically nonlinear: For large windows, nearly unimpaired reading is possible with fixation durations of around 250 ms for skilled readers (Brysbaert, 2019; Rayner, 2009), whereas with successively smaller windows reading becomes progressively impaired as indicated by slower reading rates, prolonged fixation durations or reduced saccade amplitudes. This nonlinear relationship between window size and reading

performance is best described with a nonlinear asymptotic regression model (Sperlich et al., 2016); with one parameter that captures unimpaired reading performance without a window (the “*asymptote*”) and a second parameter that captures the degree of impairment by successively smaller windows (the “*growth rate*”). Research from our group has shown that the growth rate of the nonlinear curve provides a direct and continuous measure of the size of the perceptual span and, moreover, is useful as an index of inter-individual differences of perceptual-span size (for details, see Sperlich et al., 2016).

The present study utilizes the methodological advances of Sperlich et al. (2016) to provide new evidence on three independent and important aspects of the perceptual span during reading. The first is developmental, and the other two are of more general interest to cognitive psychology. First, we map the longitudinal development of the perceptual span for German first- to sixth-graders. Second, we characterize the perceptual span separately for word-based temporal and spatial measures. Third, and importantly, we provide direct evidence for a dynamic adjustment of the perceptual span in response to cognitive processing difficulty from a character-based moving-window paradigm. We incorporated data from three waves of a recent longitudinal study on eye movements during natural reading (for the first two waves see Sperlich et al., 2015, 2016), where we employed the moving-window paradigm to map the development of the size of the perceptual span. The need to estimate person parameters for the individual differences approach in the longitudinal study led to the development of the nonlinear mixed modeling approach, which in combination with the large data set allowed us to address the above questions.

### *Different Perceptual Spans for Temporal and Spatial Decisions?*

During reading, the reader makes largely independent decisions about *when* and *where* to move the eyes next (Findlay & Walker, 1999). The when-decision determines the timing of

saccade execution; the where-decision determines the planning of saccade targeting. Does the size of the perceptual span differ for these two types of decision? Because the where-decision involves determining where to place the next fixation, it likely uses low-spatial frequency information that is further away from the current fixation, up to several words away.

Therefore, the perceptual span size for the where-decision is expected to be relatively large. In contrast, the when-decision likely depends on the current lexical processing status as well as on high-spatial frequency orthographic information near the current fixation, so the perceptual span for the when-decision is expected to be relatively small. In the present study, we aimed to disentangle the perceptual span for the when-decision from the perceptual span for the where-decision and to map their development for beginning readers of German in primary school.

To this end, we further decomposed our primary measure of reading rate into more fine-grained, word-based measures, because a global measure cannot map these independent decisions. *Temporal measures*, such as fixation duration are thought to primarily reflect the when-decision. We included the first-pass measures first fixation duration (FFD), i.e., the duration of the first fixation on a word, and gaze duration (GD), i.e., the summed duration of all first-pass fixations on a word, as well as the second-pass measure total viewing time (TVT), i.e., the summed duration of all fixations on a word. *Spatial measures*, such as the number of fixations, saccade amplitude, or saccade target probabilities are thought to primarily reflect the where-decision. Thus, we included the first-pass measures of forward saccade length (SL), i.e. the amplitude of a saccade targeting an upcoming word, refixation probability (p[RFX]), i.e., fixating a word more than once, and skipping probability (p[SKP]), i.e., not fixating the upcoming word at all. We did not include single fixation and regression probability, because preliminary analyses had revealed no differences between window-size conditions (cf. Sperlich et al., 2015). We hypothesized that the perceptual span calculated

from spatial measures would be larger than the perceptual span calculated from temporal measures, because the low-spatial frequency information needed for saccade targeting can be processed from further away than the lexical information influencing saccade timing decisions. In fact, Häikiö and colleagues (2009) presented preliminary evidence in their Appendix that the letter identity span calculated from fixation duration seemed to be less affected by smaller window sizes than when calculated from spatial measures. They reasoned that target selection benefits more from larger windows than the timing of saccades. Our reasoning is also consistent with the finding that the letter identity span obtained with masks that leave spacing intact — thereby changing low-spatial frequency information, and hence primarily disrupting spatial saccade target selection — is smaller than the full perceptual span obtained with masked word spacing.

#### *Modulation of the Perceptual Span by Foveal Load*

A critical issue in reading research is whether and how parafoveal processing adapts to current cognitive processing load (Schotter et al., 2012). Cognitive processing load is higher with words that are less common, because they are generally more difficult to process (Breland, 1996). Does the perceptual span shrink when a difficult word is being processed? Corpus analyses suggest that the more difficult a currently fixated word is, the longer it takes to process the upcoming word (Kliegl et al., 2006). Likewise, experimental research has shown that fixation duration on the *upcoming word*  $N+1$  is longer and preview benefit is reduced if it follows a low-frequency rather than a high-frequency word  $N$  (Henderson &

<sup>1</sup> Here and in the remainder of the manuscript, the word  $N$  denotes the origin of the spillover or foveal load effects. The resulting effects occur during fixations on word  $N+1$ .



Ferreira, 1990; Rayner & Duffy, 1986; Schroyens et al., 1999, White et al., 2005). Two explanations have been proposed for this effect: *spillover*, also called *lag effect* (Kliegl et al., 2006; Rayner & Duffy, 1986) and *foveal load* (Henderson & Ferreira, 1990).

The *spillover hypothesis* states that lexical access to a high-frequency word  $N$  is largely completed by the time the eyes move to the upcoming word  $N+1$ , whereas processing of a low-frequency word  $N$  is still ongoing after the eyes have left the word. Thus, some resources are still allocated to a low-frequency word  $N$  during fixations on word  $N+1$ , resulting in prolonged gaze duration on  $N+1$  (Kliegl et al., 2006). In contrast, the *foveal load hypothesis* states that parafoveal pre-processing of the upcoming word  $N+1$  is continuously modulated by ongoing processing demands of word  $N$ . For a high-frequency word  $N$  foveal processing demands are relatively small, freeing up attentional resources for parafoveal pre-processing of the upcoming word  $N+1$ . Parafoveal pre-processing is similar to priming in single-word priming paradigms, and hence results in shorter gaze duration on  $N+1$ . A low-frequency word  $N$  depletes attentional resources that are usually available for parafoveal pre-processing of  $N+1$ . Thus, the reduction of parafoveal pre-processing impedes pre-activation of word  $N+1$ , which then results in prolonged gaze duration on  $N+1$ .

Note that the two hypotheses are not mutually exclusive: it is possible that a gradient degree of pre-processing of  $N+1$  during fixation on  $N$  and a gradient degree of spillover processing of  $N$  during fixation on  $N+1$  can both occur. However, the underlying cognitive architectures are quite different. Spillover effects are usually explained by inhibition of the saccade leaving word  $N+1$  (the “lag”) due to the continued processing (the “spillover”) of a low-frequency word  $N$  (Kliegl et al., 2006). In contrast, foveal load effects point to a decrease in (sub-)lexical pre-processing of word  $N+1$  during fixation of a low-frequency word  $N$  due to spatial modulation of an attentional filter (Henderson & Ferreira, 1990). This idea is related to the well-known zoom-lens and gradient metaphors of attention (Eriksen & St. James, 1986;

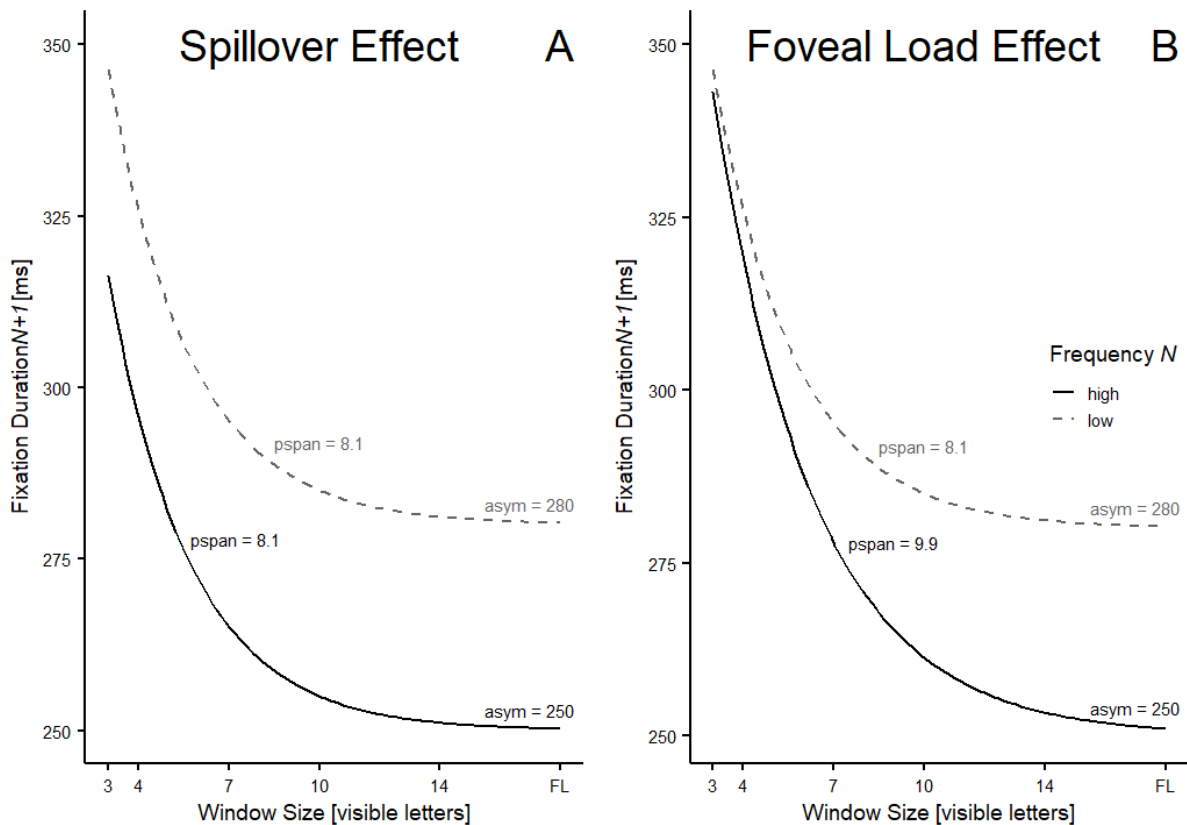
LaBerge & Brown, 1989; Schad & Engbert, 2012) and suggests that the physical extent of visuospatial attention shrinks when foveal processing becomes more difficult. The present study will provide empirical evidence for the existence of foveal load effects.

So far studies on the foveal load hypothesis have been largely restricted to the use of the invisible-boundary paradigm (for reviews, see Hyönä, 2011; Rayner, 2009; White et al., 2005). In the basic invisible-boundary paradigm, parafoveal pre-processing of a target word is prevented by a mask of Xs, which is first uncovered as the participant's gaze passes an invisible boundary in front of the target word. Target words are masked on some trials and left readable on others. The difference in gaze duration on the target word between masked and unmasked trials is termed the *preview benefit*, which is typically around 30-50 ms (Rayner, 2009). Importantly, preview benefit indicates the processing advantage (in milliseconds) gained by parafoveal pre-processing of the target word. Preview benefit does not indicate the size of the perceptual span (in letters). However, in the original formulation of the foveal load hypothesis, Henderson & Ferreira (1990, p. 427) stated that '*the perceptual span during reading changes as a function of foveal processing difficulty*'. This formulation of the hypothesis has not yet been put to direct test. In the present study, we directly test the prediction that the size of the perceptual span shrinks as a function of local processing difficulty, by relating word-by-word fluctuations of perceptual-span size to the mean word-frequency statistic of the fixated word as a measure of word difficulty (Breland, 1996).

In particular, we examine the effect of word  $N$  frequency on  $N+1$  first fixation duration, an early measure of lexical access, and  $N+1$  gaze duration, a late measure of lexical processing (Rayner, 2009). Both the spillover hypothesis and the foveal load hypothesis predict that fixation duration on  $N+1$  will take longer after a low-frequency than a high-frequency word  $N$  for unimpaired reading (i.e., at asymptote, *asym*, see Figure 1A). However,

only the foveal load hypothesis predicts an interaction between word  $N$  frequency and window size. Specifically, the foveal load hypothesis predicts that the size of the perceptual span ( $pspan$ ) will be smaller for a low-frequency than a high-frequency word  $N$  (Figure 1B).

Importantly, the spillover hypothesis does not predict such a modulation of the perceptual span size as a function of processing difficulty of word  $N$ , because spillover is solely dependent on word  $N$ , but not on masking of word  $N+1$ . The spillover hypothesis alone cannot exclusively account for an effect of window size, although window-size effects would not rule out spillover effects as part of the explanation for increased fixation duration on  $N+1$ . Initial evidence regarding the *development* of spillover and foveal load effects comes from a recent study using a variant of the boundary paradigm (Marx et al., 2017). Gaze duration on word  $N+1$  was prolonged after a low-frequency word  $N$  for Austrian fourth and sixth graders. The results were in line with the spillover hypothesis but failed to provide evidence for the foveal load hypothesis, because preview benefit on  $N+1$  was not modulated by foveal processing difficulty of word  $N$ . The present study challenges these findings by covering more and earlier school years and using a direct measure of perceptual-span size rather than preview benefit.



*Figure 1.* Predictions made by the spillover versus the foveal load hypothesis (panels). Fixation duration on the upcoming word  $N+1$  (y-axis) is predicted by window size (x-axis) and frequency of the current word  $N$  (line style). (A) At asymptote (full line, FL, i.e. normal reading without a moving window) the spillover hypothesis predicts longer fixation duration on the upcoming word  $N+1$  after processing of a low-frequency word  $N$  ( $asym_{low} > asym_{high}$ ), but no interaction with window size ( $pspan_{low} = pspan_{high}$ ). That is, the growth rate is predicted to be the same for low- and high-frequency words. (B) The foveal load hypothesis makes the same prediction as the spillover hypothesis with respect to fixation duration on  $N+1$  at asymptote ( $asym_{low} > asym_{high}$ ), but additionally predicts an interaction with window size, as indicated by an increase in the size of the perceptual span while processing a high-frequency word  $N$  ( $pspan_{high} > pspan_{low}$ ). That is, the growth rate is increased for high-frequency compared to low-frequency words. The schematic illustration is adapted from Marx et al. (2017).

### *The Present Study*

To summarize, the aims of the present study were as follows: First, to map the development of the perceptual span for German children in grades 1 to 6 of primary school using a wide range of eye-movement based reading measures. Second, to investigate whether the size of the perceptual span depends on decision type, i.e., saccade timing vs. targeting.

Third, to directly test a prediction derived from the foveal load hypothesis, namely that the size of the perceptual span flexibly adapts to local processing difficulty. To our knowledge, this is the first direct test of the foveal load hypothesis using perceptual-span size as dependent measure.

## **Method**

### *Participants*

Participants were a subsample of 141 children (74 girls), recruited from the large-scale longitudinal PIER study (Jung et al, 2016; Meixner et al., 2019) with 1,657 German students. Participants of the subsample were recruited from 63 classes in 16 primary schools in the federal state of Brandenburg and took part in a 3-year longitudinal follow-up study starting at Grade 1 ( $N_1 = 50$ ), Grade 2 ( $N_2 = 45$ ), and Grade 3 ( $N_3 = 46$ ), respectively. First-graders had at least 8 months of reading instruction and, because of the demanding task affordances, all participants were selected based on prior tests at school including task understanding, motivation, concentration, and anxiety. One school year later, at Wave 2, 132 children participated again (Grades 2-4;  $N_1 = 48$ ,  $N_2 = 40$ ,  $N_3 = 44$ ), and two school years later at Wave 3, 123 children participated a third time (Grades 4-6;  $N_1 = 44$ ,  $N_2 = 39$ ,  $N_3 = 40$ ). Four datasets were corrupted, of which one Wave 1 first-grader dataset was replaced by its Wave 2 dataset, because the child stayed down a year. At Wave 3, fourth-graders were 9 to 10 years old, fifth-graders were 10 to 11, and sixth-graders were 11 to 12. All children were native German speakers, and had normal or corrected-to-normal vision. Parents gave informed consent and received 15€ as a compensation for travel costs.

*Materials and Procedure*

For the actual experiment, single sentences were presented on a 22-inch CRT screen (resolution = 1024 x 768 pixels, refresh rate = 150 Hz) in black color on a white background in 18pt mono-spaced Courier New Bold font. Viewing distance was 60cm. Children read these sentences silently for comprehension and at their own pace. During reading, movements of the right eye were recorded at 1000 Hz using an EyeLink1000 Tower mount eye-tracker (SR Research). Before the experiment started, and whenever necessary, but at least every 13 trials, a calibration and validation routine was run. Only calibrations with an average deviation smaller than 1° of visual angle were accepted. The average actual validation error was 0.4° of visual angle, which was about the size of a single letter (0.4° of visual angle). In general, the EyeLink1000 provides an excellent average spatial accuracy of about 0.6° of visual angle (Ehinger et al., 2019).

The experiment started with three practice trials that were not included in the analysis. At the first two waves, children read 72 sentences, with the exception of first-graders, who only read 32 sentences. At the third wave, children read 96 sentences in total. The material always included first-grader sentences and grade-level sentences. All sentences were selected from short stories about animals that were taken from grade-level biology textbooks. First-grader sentences were the same for all three cohorts, but were replaced by novel sentences at each wave to prevent re-reading bias. For the present study, first-grader and grade-level sentences were collapsed<sup>2</sup>. The corpus differed between grades (see Table 1) primarily in sentence length,  $F(5,674) = 10.7, p < .0001$ , but also, in particular at later grades, in mean word length,  $F(5,674) = 5.9, p < .0001$ , and mean log10 of normalized word frequency,

<sup>2</sup> Comparison between age-appropriate and first-grader sentences is beyond the scope of the present study. For details, see Sperlich et al. (2016).

$F(5,674) = 4.0, p = .0015$ . Word frequency norms were taken from a large German corpus based on more than 100,000,000 words (Heister et al., 2011).

*Table 1.* Material characteristics for the sentence corpora for Grades 1 to 6.

	Grade					
	1	2	3	4	5	6
Sentence Length	6.3 (1.6)	7.8 (2.4)	8.9 (2.7)	9.0 (2.7)	9.5 (4.8)	10.5 (5.5)
Word Length	5.2 (0.9)	5.3 (0.9)	5.1 (0.8)	5.2 (0.8)	5.5 (0.8)	5.6 (1.0)
Word Frequency	2.1 (0.4)	2.1 (0.4)	2.2 (0.4)	2.2 (0.4)	2.0 (0.5)	2.0 (0.4)

*Note.* Mean sentence length in number of words per sentence, mean word length in number of letters per word, and mean word frequency is given as the log10 of counts per million according to the German DWDS corpus (Heister et al., 2011). Standard deviations are presented in parentheses. In order to calculate variance, all statistics were calculated on the sentence level.

Sentence presentation started with the fixation on a trigger to the left of the sentence and ended with fixation on a trigger in the lower right corner. Each sentence was presented separately: Short sentences with fewer than 80 characters were presented on a single line at the vertical screen center; longer sentences continued on up to two additional lines. The interspacing between the lines was seven vertical character spaces. In order to ensure reading for comprehension, reading comprehension was probed on a third of the trials, selected randomly. Children were presented with two thematically related pictures, of which they had to select the one that best depicted the meaning of the sentence. Immediately thereafter, children received pictorial feedback on the correctness of their answer, which was followed by a scoreboard showing the total amount of correct and incorrect responses. The research assistant constantly motivated the participant to collect even more correct answers.

A symmetric, gaze-contingent moving window was used to investigate the size of the perceptual span. Six different window sizes were applied with 3, 4, 7, 10, 14, or all characters (full line, FL) left intact to both sides of the fixated character, while the remainder of the sentence was masked with the character *x*. In order to focus on lexical rather than low-level

oculomotor control, capitalization (masked with X), inter-word blank spaces and punctuation characters were also left intact. For sentences spanning multiple lines, the window was only applied to the currently fixated line; the other lines were completely masked. The size of the window remained constant within each trial. Also, each window size appeared equally often and in a pseudo-random order with the constraint that each window size appeared again after not more than eleven trials, i.e., once within a block of six trials. In order to keep experimental conditions identical for all children – as necessary for a comparison of inter-individual differences – window size was a fixed between-sentence condition. Thus, we balanced average sentence length, word length, and word frequency to prevent confounds between window-size conditions (all  $F_s < 1$ , smallest  $p = .674$ ). Participants were told about the window-size manipulation before the start of the experiment.

### *Data Analysis*

For the global measure of reading rate, we excluded blinks (2.7%), off-sentence fixations (6.6%), and orientation fixations on a new line (0.3%), sentences that children did not complete reading until the end (1.4%), and sentences with fewer than three fixations (0.1%). For the word-based measures, we additionally excluded first and last fixations in a sentence, all fixations on the first or last word in a sentence, fixation durations below 80 ms or above 1200 ms and saccades longer than 25 characters. For temporal measures (FFD, GD, and TVT) a log transformation was conducted prior to analysis in order to account for non-normality.

*Perceptual span and foveal load.* Data were analyzed using nonlinear mixed-effects models in R (*nlme*-package v3.1: Pinheiro, Bates, DebRoy, Sarkar, & R Core Team, 2015). Mixed-effects models incorporate fixed effects, i.e., the average effect of a manipulation, and random effects, i.e., the individual deviation of a participant from the average effect. In order



to model the random effects structure of the data we used grouping by grade, wave, and participant. For reading rate, we fitted a two-parametric nonlinear asymptotic function, yielding subject-level parameter estimates for reading speed (words per minute) and perceptual span (a transform of the growth rate with window size; for details see Sperlich et al., 2016). To test the developmental aspects, we added grade as fixed effect to the NLME models. Likelihood-ratio tests were used to compare the fits between models with and without the grade effect. To test the foveal load hypothesis, we added word frequency of  $N$  as a fixed effect into the NLME model. A few other characteristics known to influence fixation duration on  $N+1$ , namely word length and word frequency of  $N+1$  as well as launch site on  $N$  and landing site on  $N+1$  were entered as control variables into the model.

*Developmental trends.* For reasons of comparison with our earlier work (Sperlich et al., 2015, 2016), we present results of planned sliding-difference contrasts between grades for reading rate as outcome. Because the successive difference contrasts are arguably not the best way to make use of the full information in the data set, we adjusted our analysis approach for the remainder. To this end, we extracted participant parameters from the NLME models described above, and examined the developmental trends therein, by fitting an asymptotic nonlinear regression (NLS) model of the same functional form as the NLME model. The functional form accounts for the typical declining improvements from year to year that eventually will reach a final plateau. In the NLS model we estimated growth rate (the learning curve) and asymptote (the final ‘adult’ performance level) of reading acquisition, respectively. For the NLS models we report regression coefficients ( $b$ ), standard errors ( $SE$ ), and  $t$ -values ( $t=b/SE$ ). There is no clear definition of “degrees of freedom” for NLS models, and therefore, precise  $p$ -values cannot be estimated. In general, however, given the large number of observations, subjects, sentences and words in our analysis and the comparatively small number of fixed and random effects estimated, the  $t$ -distribution is equivalent to the normal

distribution for all practical purposes (i.e., the contribution of the degrees of freedom to the test statistic is negligible). Our criterion for referring to an effect as significant is  $t > 2.0$ . Of note, due to the more general scope of the paper, the current analyses remained cross-sectional in nature (for a longitudinal inter-individual differences account, see Sperlich et al., 2016).

## Results

### *General Development of the Perceptual Span*

With reading rate as the dependent measure, adding grade as a fixed effect improved model fit for both asymptotic reading rate,  $\chi^2(5) = 303.6, p < .0001$ , and perceptual span,  $\chi^2(5) = 49.9, p < .0001$ . Figure 2A visualizes the observed mean reading rates averaged across subjects with window size and grade level as predictors. Planned comparisons revealed continuous growth in asymptotic reading rate from first to sixth grade with an average yearly increase of 33.3 wpm ( $SD = 13.5$ ), largest  $p = .0189$  for the comparison between third and fourth grade, all other  $ps < .01$ . In terms of year-on-year differences, the perceptual span only improved significantly from second to third grade,  $b(SE) = -0.37(0.06), t = -6.1, p < .0001$ , and again from fourth to fifth grade,  $b(SE) = -0.09(0.04), t = -2.4, p < .05$ , all other  $ps > .05$ .

*Table 2.* Predicted asymptotic reading performance for Grades G1 to G6 (columns) for each eye movement measure (rows).

Measure	Grade						'Adult' Level
	G1	G2	G3	G4	G5	G6	
<i>Global Measure</i>							
Reading Rate (wpm)	40	86	131	147	172	206	309
<i>Temporal Measures</i>							
First Fixation (ms)	393	308	266	257	237	225	224
Gaze Duration (ms)	1277	632	429	390	362	324	343
Total Time (ms)	1895	839	564	495	466	377	430
<i>Spatial Measures</i>							
Saccade Length (chars)	3.44	5.0	5.99	6.40	7.42	7.69	8.98
Refixation (%)	0.59	0.43	0.30	0.27	0.24	0.22	0.20
Skipping (%)	0.18	0.14	0.20	0.21	0.28	0.29	0.43

*Note.* For temporal measures, inference statistics are based on log-transformed durations, but for ease of interpretation untransformed values are presented. The last column indicates the estimated (extrapolated) 'adult' performance level for each eye movement measure.

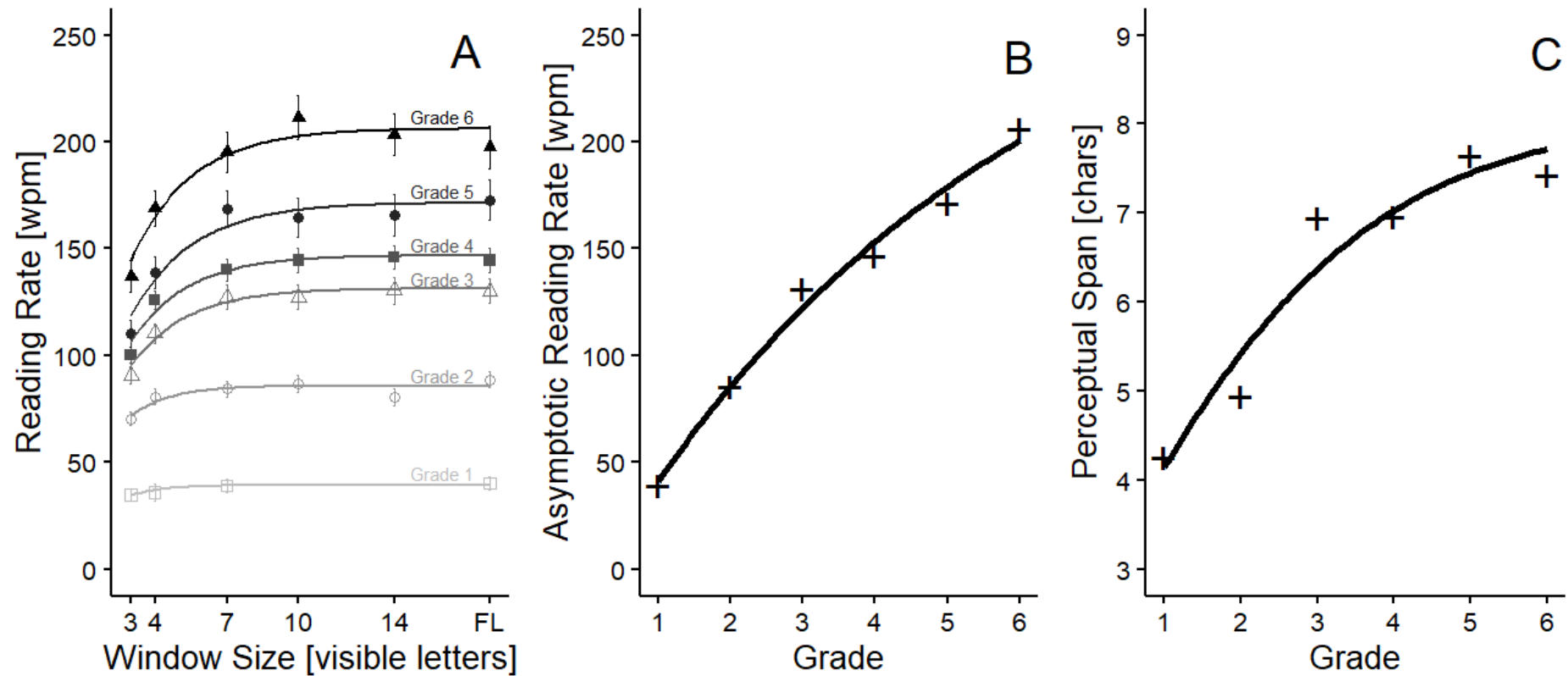


Figure 2. (A) Development of the perceptual span derived from reading rate (y-axis) as a function of window size (x-axis) across grades (shapes, colors). Points indicate observed means, and lines are NLME model fits. (B), (C) Nonlinear regression fits to the extracted NLME parameters as a function of grade (x-axis) for (B) asymptotic reading rate (y-axis) and (C) perceptual span (y-axis). See text for details.

In order to describe the developmental trajectory across the primary school years, we fitted nonlinear regression models that describe the change in estimated asymptotic reading rate (Figure 2B) and in estimated perceptual span (Figure 2C) as a function of grade level. These models are similar to a nonlinear learning curve and allow an estimation of the asymptotic reading performance as a function of grade. The asymptote can be understood as an extrapolation of the future ‘adult’ reading performance level. For reading rate, the adult performance level was estimated to be 309 wpm and for perceptual span, the adult performance level was estimated at 8.1 chars (see Table 2, last column). Of note, these estimates are in the same ballpark as other reports of adult reading rate (Brysbaert, 2019) or the adult letter identity span from languages using the Roman alphabet (Häikiö et al., 2009; Rayner, 1986). The apparent slight overestimation of asymptotic reading rate (cf. Brysbaert, 2019) likely just suggests that the current texts were relatively easy. Alternatively, there may be some additional slowing factors involved in later development which our current sample did not capture.

In addition, the model allows us to estimate the learning curve with which a given reading performance measure reaches its future performance level. By visual comparison of Figures 2 B and C, it seems clear that reading rate will likely continue to grow quite substantially beyond Grade 6 (compare Table 2, first row, second to last versus last column), whereas perceptual span has perhaps already almost reached its future performance level with only little room for improvement of around a half letter space (compare Table 3, first row, second-to-last versus last column). From this model, one can also calculate the grade at which a certain percentage of the future ‘adult’ performance level is reached. For example, 80% (~250 wpm) of the ‘adult’ reading rate is reached at Grade 9, whereas 80% (~7 letters) of the ‘adult’ perceptual span is already reached in Grade 3. Together these observations suggest that perceptual span development ends quite a lot earlier than reading rate development.

*Different Perceptual Spans for Temporal and Spatial Decisions?*

We analyzed other dependent measures than the global measure of reading rate to examine whether perceptual span development was driven more by the temporal “when” or the spatial “where” decision of eye movement control (Findlay & Walker, 1999). Tables 2 and 3 summarize findings across grades for the temporal “when” measures and for the spatial “where” measures. For all measures, asymptotic reading performance showed continuous improvement across grades and particularly large improvements between first and second, and second and third grades (see Table 2). A visual comparison of the grade performance levels and the future performance level revealed that temporal measures indicative of the “when” decision tended towards their developmental limit at sixth grade. In contrast, spatial measures indicative of the “where” decision do not yet reach ‘adult’ level at sixth grade and thus likely contribute to further reading rate development (compare Table 2, second-to-last versus last column). With respect to the development of the perceptual span, for most temporal and spatial measures the learning curve reached its final performance level at around Grade 3 (compare Table 3). The future ‘adult’ performance level was estimated to be about 4 letters for all temporal measures (FFD, GD, and TVT). Note that our perceptual span estimate for temporal measures of about 4 characters is notably smaller than the word identification span of 7-8 characters in Rayner’s (1998) review. The difference could arise from any number of procedural differences between studies, including sample, language, material, analysis approach, etc. Thus we do not recommend comparing the absolute numbers but rather focusing on the qualitative patterns of relative differences.

Importantly, the estimates were quite different for spatial measures. The future ‘adult’ performance level was estimated to be about 7-8 letters for saccade length and skipping probability, and to be about 5 letters for refixation probability.

Table 3. Estimated perceptual span in characters to the right for Grades G1 to G6 (columns) for each eye movement measure (rows).

Measure	Grade						'Adult' Level
	G1	G2	G3	G4	G5	G6	
<i>Global Measure</i>							
Reading Rate ( <i>pspan</i> )	4.26	4.94	6.9	6.97	7.65	7.44	8.14
<i>Temporal Measures</i>							
First Fixation ( <i>pspan</i> )	3.48	3.89	4.17	4.17	3.81	4.04	4.00
Gaze Duration ( <i>pspan</i> )	3.35	4.20	4.62	4.19	4.48	4.22	4.34
Total Time ( <i>pspan</i> )	2.73	3.43	3.75	3.64	4.04	3.81	3.88
<i>Spatial Measures</i>							
Saccade Length ( <i>pspan</i> )	0.48	4.61	7.18	7.42	7.18	7.08	7.38
Refixation ( <i>pspan</i> )	1.69	4.59	5.27	4.50	5.04	4.46	4.80
Skipping ( <i>pspan</i> )	-	4.97	7.57	8.29	7.89	7.04	7.67

*Note.* Inference statistics are based on growth rate parameters; for ease of interpretation *pspan* values are presented. The last column indicates the estimated (extrapolated) 'adult' performance level for each eye movement measure. For skipping, first-graders were excluded from analysis due to an implausibly high average (see Table 2), presumably due to random saccade targeting errors.

Together these findings suggest that the perceptual span is not a uniform entity, but its size crucially depends on type of eye-movement control. Above we assumed that duration measures are more closely associated with lexical processing, and spatial measures with saccade target selection. Of course, it is likely that neither of them are process-pure measures. Nevertheless, if our assumption is approximately correct, then it seems that the perceptual span is larger for saccade target selection associated with the "where" decision than for lexical processing associated with the "when" decision. Interestingly, the perceptual span estimates based on refixation probability line up more closely with the temporal rather than the spatial measures. Possibly this indicates that refixation is more strongly governed by a lexical mechanism that is responsible for controlling fixation duration rather than spatial selection. Note that refixation probabilities and gaze durations are usually highly correlated.

Alternatively, the smaller effect on refixation probability might be due to spatial selection of different saccade targets: Because refixations occur within the boundaries of word  $N$ , attention needs to be less broadly distributed for the selection of the next target than when the target lies outside  $N$  as for forward saccades and skipping. This latter mechanism is compatible with attentional shifting preceding saccades (Deubel & Schneider, 1996). In any case, the perceptual span is larger before inter-word than before intra-word saccades, which is already a first hint that perceptual-span size adapts to current processing demands.

#### *Modulation of the Perceptual Span by Foveal Load*

To investigate whether processing of the parafoveal word  $N+1$  was influenced by cognitive processing load of word  $N$ , we ran NLME models for FFD of  $N+1$  and for GD on  $N+1$  with frequency of  $N$  as predictor. Of course, for cases where  $N$  or  $N+1$  was skipped (17.8%), fixation durations on word  $N+1$  could not be calculated and these words were excluded from the analysis. We performed a median split on frequency of  $N$  for the whole corpus in order to compare groups of high-frequency versus low-frequency words. The characteristics of the final word set are presented in Table 4. The negative lag-1 autocorrelation (see also Fernández et al., 2014) seems to be common for the German language, e.g. high-frequency articles generally precede low-frequency nouns. In addition, low-frequency words are typically longer than high-frequency words (Kliegl et al., 1983). Thus, to statistically control for word features other than word frequency of  $N$ , which might have affected FFD or GD on  $N+1$ , we additionally included word frequency of  $N+1$ , word length of  $N+1$ , launch site from  $N$ , and landing site on  $N+1$  as continuous control variables. Finally, to test the developmental trajectory, we also included grade level as predictor.



Table 4. Word characteristics of low- vs. high-frequency words  $N$ .

	Frequency $N$	
	Low	High
Number of Fixations	73,587	62,826
Frequency $N$	1.0 (1.1)	3.5 (0.5)
Frequency $N+I$	2.4 (1.4)	1.9 (1.6)
Word Length $N$	6.9 (2.6)	3.5 (1.1)
Word Length $N+I$	5.0 (2.4)	5.8 (2.7)
Launch Site $N+I$	-3.0 (1.6)	-3.1 (1.7)
Landing Site $N+I$	2.0 (1.5)	2.1 (1.5)

*Note.* Grouping is based on a median split on mean word frequency of the whole corpus. Mean frequency is given as the log10 of counts per million according to the German DWDS corpus (Heister et al., 2011), mean word length in number of letters per word, launch site in letters to the left of the word end, landing site in letters to the right of the word beginning, and standard deviations are presented in parentheses.

As illustrated in Figure 1A, both the spillover and foveal load hypotheses predict the no-window (FL) fixation duration on  $N+I$  to be shorter if  $N$  is a high-frequency word. To this end, model comparisons showed that including the hypothesized  $N$  frequency effect improved model fit for asymptotic fixation durations, FFD:  $\chi^2(1) = 106.4$ ,  $p < .0001$ ; GD:  $\chi^2(1) = 145.9$ ,  $p < .0001$ . Planned contrasts revealed prolonged asymptotic FFD on  $N+I$  if preceded by a low-frequency ( $asym = 5.56$ , 260 ms), relative to a high-frequency word  $N$  ( $asym = 5.53$ , 252 ms),  $b = 0.030$ ,  $t = 10.0$ ,  $p < .0001$  and longer asymptotic GD on  $N+I$  if preceded by a low-frequency ( $asym = 5.98$ , 395 ms), relative to a high-frequency word  $N$  ( $asym = 5.93$ , 376 ms),  $b = 0.045$ ,  $t = 12.0$ ,  $p < .0001$ .

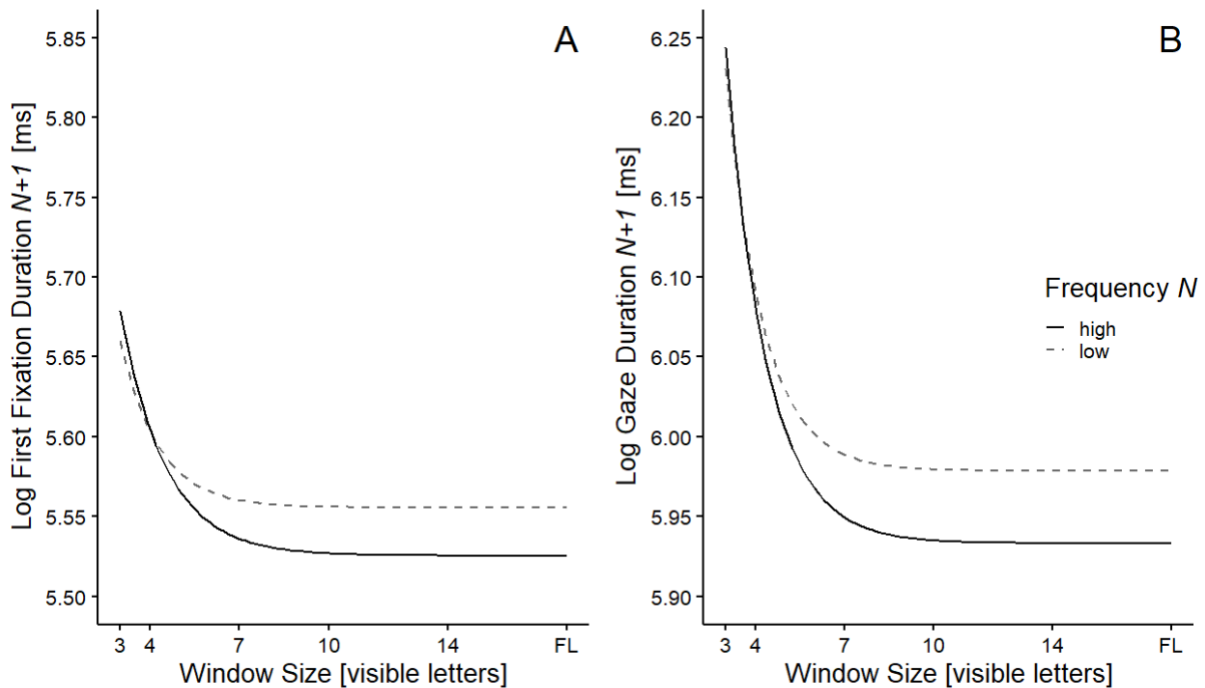
This result is still compatible with both the spillover and the foveal load hypothesis. However, only the foveal load hypothesis additionally predicts shrinkage of the perceptual span if  $N$  is a low-frequency word (see Figure 1B). Model comparisons confirmed the hypothesized  $N$  frequency effect for perceptual span, FFD:  $\chi^2(1) = 76.2$ ,  $p < .0001$ ; GD:  $\chi^2(1) = 70.5$ ,  $p < .0001$ . Planned contrasts revealed a smaller perceptual span based on FFD during the processing of a low-frequency ( $pspan = 3.83$  chars) compared to a high-frequency word  $N$  ( $pspan = 4.51$  chars),  $b = -0.164$ ,  $t = -8.4$ ,  $p < .0001$ , and based on GD during the processing

of a low-frequency ( $pspan = 3.75$ ) compared to a high-frequency word  $N$  ( $pspan = 4.08$  chars),  $b = -0.084$ ,  $t = -7.9$ ,  $p < .0001$ .

This finding exclusively provides support for the foveal load hypothesis. Importantly, the observed effects (see Figure 3 and Table S1) can be interpreted in a similar way to an interaction effect of masking and  $N$  frequency in the invisible boundary paradigm (e.g., Henderson & Ferreira, 1990). However, in the boundary paradigm the parafoveal processing advantage (or *preview benefit*) of no mask versus a mask is reported. Unlike the boundary paradigm, the moving window technique includes multiple windows of varying size allowing for a direct measurement of the size of the perceptual span rather than indirectly inferring it from a simple binary measure of the costs of masking. Our findings can therefore be interpreted as direct evidence for a dynamic adjustment of the size of the perceptual span in response to the frequency of word  $N$ .

Although not relevant to the present research question, the covariates also yielded interesting results. The analysis revealed that the closer the launch site was to the end of word  $N$  the shorter FFD and GD were on word  $N+1$ , and the larger the perceptual span was. This means that the closer the last fixation preceding the outgoing saccade is to the end of  $N$ , the larger the perceptual span, which in turn leads to shorter processing time on  $N+1$  (cf. Inhoff & Rayner, 1986; Pan et al., 2020). Furthermore, as anticipated, FFD and GD were shorter for high-frequency compared to low-frequency words  $N+1$  and for shorter compared to longer words  $N+1$  (cf. Huestegge et al., 2009; Hyönä & Olson, 1995; Joseph et al., 2013; Tiffin-Richards & Schroeder, 2015). Interestingly, the size of the perceptual span at word  $N$  was larger if the upcoming word  $N+1$  was low-frequency (cf. Hyönä & Bertram, 2004; Risse & Kliegl, 2012, but see Inhoff & Rayner, 1986; Kennison & Clifton, 1995; Schroyens et al., 1999). Finally, landing site differentially affected FFD and GD: The further the incoming saccade landed into the word  $N+1$ , the shorter GD, but the longer FFD on  $N+1$ . This

phenomenon is called the inverted optimal viewing position effect (Nuthmann et al., 2005, 2007; Vitu et al., 2001) and might be explained by fast error-correction of mislocated fixations landing away from the word center, leading to a prolongation of FFD, but to a reduction of GD for initial fixations near the word center.



*Figure 3.* Model fits for the Frequency  $N$  effect on log-transformed first fixation duration (A) and gaze duration (B) on word  $N+1$ . Solid black lines represent fits for high-frequency words  $N$ ; dashed gray lines represent fits for low-frequency words  $N$  based on a median split of Frequency  $N$ . As expected, first fixation duration and gaze duration were longer for low frequency words  $N$  at larger window sizes and at asymptote, i.e., without a window. In line with the foveal load hypothesis the curve was steeper, i.e., the perceptual span was larger when fixating high-frequency words  $N$  compared to low-frequency words  $N$  for both first fixation duration and gaze duration. See text for details.

#### *Developmental Course of the Foveal Load Effects*

To examine at which grade foveal load effects first emerged and how they developed, we added a grade-by-frequency interaction term to the full model (Table S2). We separately

analyzed the effect of this term on asymptotic reading rate (*asym*) and on the perceptual-span measure (*growth rate*).

For asymptotic fixation duration, we observed significant grade-frequency interactions for both duration measures, FFD:  $\chi^2(5) = 42.0, p < .0001$ ; GD:  $\chi^2(5) = 91.1, p < .0001$ . Using Bonferroni correction for multiple comparisons, post-hoc contrasts revealed that the effect of frequency emerged at third grade, and remained stable thereafter, FFD:  $b_{3rd-2nd} = -0.030, t = -4.5, p < .0001$ , all other  $ps > .01$ ; GD:  $b_{3rd-2nd} = -0.075, t = -7.2, p < .0001$ , all other  $ps > .01$ . These analyses indicate that frequency of word  $N$  first affected fixation duration on word  $N+1$  at third grade.

To answer the question of whether modulation of the perceptual span is due to spillover effects or due to foveal load effects, we inspected the grade-frequency interaction on the perceptual-span measure, which reached significance only for GD,  $\chi^2(5) = 16.2, p < .01$ , but not for FFD,  $\chi^2(5) = 4.5, p = .48$ . The grade-frequency interaction on perceptual span for GD numerically emerged between first and second grade,  $b_{2nd-1st} = -0.088, t = -1.1, p = .26$ , made a significant step between second and third grade,  $b_{3rd-2nd} = -0.095, t = -4.5, p < .0001$ , and remained stable thereafter, all other  $ps > .01$ . The pattern was similar in FFD, but non-significant, all  $ps > .01$ . These results suggest that the effects of frequency of  $N$  on fixation duration on  $N+1$  are at least partly due to foveal load effects emerging between second and third grade.

## Discussion

The present study provides a comprehensive account of the development of the perceptual span during reading in first- to sixth-graders. Globally, significant advances in reading rate occurred with every school year and followed an almost linear trend across the six grade levels (see Figure 2B). In contrast, the learning curve was considerably decelerated for the perceptual span (see Figure 2C). The perceptual span featured a major developmental step from second to third grade (see also Sperlich et al., 2015, 2016) and, as this third wave of the longitudinal study showed, small but significant improvements thereafter.

Additionally, we examined whether and when the perceptual span is modulated by cognitive processing difficulty, as operationalized by word frequency (Breland, 1996). We found strong evidence for foveal load effects, developmentally emerging as soon as the perceptual span was beginning to be used effectively.

### *General Development of the Perceptual Span*

In the present study, German sixth-graders' perceptual span was estimated to be roughly between seven and eight letters to the right of fixation and – based on the learning curve – is not expected to show major improvements beyond sixth grade. This estimate is of similar size to the nine letters in Finnish sixth-graders (Häikiö et al., 2009; *letter-identity span*) and slightly lower than the eleven letters in English sixth-graders (Rayner, 1986; *full perceptual span*). Due to various methodological issues, absolute size estimates for the perceptual span are somewhat difficult to compare between the three studies. Note that we used a mask with intact spacing, punctuation, and capitalization, and additionally the languages, text material, window size conditions, and window symmetry differed between the studies. Still, estimated reading rates at full length in all three studies were in the same ballpark and as expected (Brysbaert, 2019).

In comparing reading rate with perceptual span development (Figure 2B-C), we observed that reading rate improved almost linearly across the whole primary school period by about 33 wpm per year, whereas perceptual span did not show major improvement beyond third grade. The minor improvements of perceptual span are not sufficient to fully account for the large improvements in reading rate. It is well possible that there is a fundamental working memory limit constraining how much upcoming perceptual information can be buffered during ongoing cognitive processing, and that the perceptual span is reflecting this limit.

What else then drives reading rate development? We believe that perceptual, bottom-up processing of information becomes less important once the process of reading is mastered. In other words, top-down processing during reading becomes a better predictor of reading rate development from fourth grade on. Associative networks in the lexicon become strengthened so that a large volume of high-frequency words become more and more expected, such that less and less perceptual input is needed for word identification. Proficient reading is essentially made possible by acquired knowledge about words. Thus improved reading skill with development depends on improvements in general knowledge, statistical learning and solid knowledge about words and text structure, which then support top-down word identification, e.g., by limiting the possible word candidates that come next in the sentence and allowing efficient prediction of upcoming words. Indeed, better readers make more use of word predictability information (Hawelka et al., 2015).

There is ample evidence for the lexical quality hypothesis (Perfetti, 2007; Perfetti & Hart, 2001), which states that the ability to retrieve word identities that provide word meanings in a given context depends on the knowledge a reader has about specific lexical representations. This knowledge about word forms (orthography, phonology, syntax) and meanings is acquired through practice. Efficient access achieved by high quality lexical representations frees working memory and attentional resources that the less advanced reader

needs to devote to activating lexical representations. The question remains, if more attentional resources are available with more proficient reading, why does the perceptual span not increase further? The likely answer is that the coordinative processes required during reading including constant updating requirements as well as syntactic parsing and building of a mental model of the text are quite taxing for working memory, and only a few words can be kept active at the same time (for a similar argument regarding the eye-voice span see Laubrock & Kliegl, 2015).

#### *Different Perceptual Spans for Temporal and Spatial Decisions?*

Perceptual span in the previous paragraph refers to the usual measure based on reading rate. We put forward an alternative word-based analysis, which generally supported the reading rate analysis, but also showed that the effect of window size on reading rate is largely determined by spatial rather than temporal decisions – even with masks preserving word spacing and punctuation. Asymptotic reading performance consistently improved across the six grades, and the improvement was most pronounced from first to third grade for all measures (see Table 2). Learning curves showed that temporal measures (FFD, GD, and TVT), reflecting the lexically driven decision of *when* to move the eyes, were not likely to show further development beyond sixth grade. In contrast, two of the three spatial measures (saccade length and skipping probability), reflecting the attention-driven spatial-selection decision of *where* to move the eyes, showed large, significant improvements in grades five and six, and further growth is to be expected. This suggests that increases in reading rate around fifth and sixth grade are mostly driven by improvements in the “where” rather than the “when” decision and predicts that further increases in reading rate will likely also stem from developments in spatial rather than temporal decisions.

Perceptual span size at sixth grade was estimated to be approximately 3-4 letters for “when” measures and approximately 7-8 letters for “where” measures. In contrast to reading rate and spatial measures, which did not see significant improvement until third grade, two of the three lexical temporal measures (GD and TVT) showed significant improvements in perceptual span as early as between first and second grade. In fact, a perceptual span of 3-4 letters is virtually as wide as the fovea (given the present stimulation conditions) and does not include parafoveal processing. This suggests that saccade timing is based largely on foveal information. Furthermore, our results seem to suggest that during the course of development, foveal processing matures before parafoveal processing. The perceptual span estimated from word-based measures also showed a major improvement from second to third grade mirroring the perceptual span estimated from global reading rate. Probably, during this period the phonological route is routinely bypassed (Sperlich et al., 2016), freeing up resources for parafoveal processing of the upcoming word.

#### *Modulation of the Perceptual Span by Foveal Load*

Earlier work on the foveal load hypothesis exclusively applied variants of the boundary paradigm with adults (Henderson & Ferreira, 1990; Schroyens et al., 1999) and only recently with children (Marx et al., 2017). Here, we extend this research in two important respects: First, foveal load effects discovered in adult readers can be demonstrated even in primary school children, suggesting that these effects are caused by a very general attentional mechanism. Second, the findings from the boundary paradigm can be generalized to the moving window paradigm, which provides a more direct test of the foveal load hypothesis because it directly measures the size of the perceptual span.

Hitherto we used word-based measures to map the perceptual span in more detail. But importantly, the derivation of word-based measures also enabled us to investigate whether the



size of the perceptual span is adaptive to local processing demands. A position put forward by the foveal load hypothesis (Henderson & Ferreira, 1990), that has not previously been directly tested, is that the perceptual span narrows if the foveal word is difficult to process. Especially difficult to process words are those that occur less frequently in textbooks (Breland, 1996). In this vein, we found that during the foveal processing of a low-frequency word  $N$  the size of the perceptual span was smaller compared to the size of the perceptual span during the foveal processing of a high-frequency word  $N$  (see Figure 3). In other words, window-size effects were significantly smaller for low-frequency words  $N$  (smaller growth) than for high-frequency words  $N$  (larger growth). The effect was slightly more pronounced in FFD, an early measure of lexical access than in GD, a late measure of lexical processing. This suggests that words can often be resolved in a single fixation if pre-processing is available. In addition to the effect on perceptual-span size, we found that foveal processing of a low-frequency compared to a high-frequency word  $N$  resulted in a longer FFD and GD on  $N+1$ . Taking both findings together, this is strong support for the foveal load hypothesis, which predicts longer processing time of word  $N+1$  due to a smaller perceptual span, resulting in less efficient parafoveal processing.

The pattern of results of the current study cannot be exclusively explained by the spillover hypothesis. The spillover hypothesis attributes the finding of longer fixation durations on word  $N+1$  to a “processing lag” (Kliegl et al., 2006; Rayner & Duffy, 1986). This means that if processing of the foveal word  $N$  cannot be finished during its fixation, processing time of  $N$  “spills over”, i.e. is added to the processing time of  $N+1$ . Importantly, this explanation does not predict a modulation of the perceptual span size as a function of processing difficulty of word  $N$  (see Figure 1A). Spillover effects can thus not exclusively account for the current findings, although they cannot be ruled out as part of the explanation. Indeed, foveal load effects and the spillover hypothesis are not mutually exclusive. For

example, for a very difficult  $N$ , fixation duration on  $N+I$  likely increases, because perceptual span narrows so that parafoveal processing of  $N+I$  is aborted, *and* because processing time of  $N$  additionally spills over to processing time of  $N+I$ . In contrast, for medium-difficulty words, processing might not lag behind and the increased fixation duration on  $N+I$  may be solely explained by reduced parafoveal processing. From this perspective, the overall text difficulty likely determines the observation. Part of the explanation why Marx et al. (2017) failed to find evidence for the foveal load hypothesis might be that they used relatively easy boundary words (but see Vasilev et al. (2018) for a more serious critique of the incremental boundary paradigm). To summarize, spillover effects may complement, but certainly cannot account for the full story, which, as the present results clearly indicate, requires a modulation of the perceptual span by foveal load.

The pattern of results is compatible with, but also provides constraints for current computational models of eye movement control. The SWIFT model of eye-movement control (Engbert et al., 2005; Risse et al., 2014; Schad & Engbert, 2012) assumes that attentional resources are limited and gradually allocated to the processing of words  $N$  and  $N+I$ . If more attentional resources are spent on foveal processing due to high foveal load, the perceptual span (or “processing span”) dynamically shrinks and consequently fewer resources are available for parafoveal processing. Thus, an increase in FFD and GD on  $N+I$  can be observed. The foveal load mechanism is also implemented in the E-Z reader model of eye movement control (Reichle, Pollatsek, Fisher, & Rayner, 1998), which rather than positing a continuous attentional gradient assumes discrete word-based attention shifts. Attention is allocated to one word at a time and word recognition occurs serially and sequentially. Processing of the parafoveal word starts as soon as attention is shifted towards it, which only happens when the lexical processing of the foveal word has made sufficient progress. When the foveal word is difficult to process, its lexical access takes longer, so that the attention shift

to the parafoveal word is delayed. Whereas the SWIFT model explicitly incorporates the spatial distribution of attention and its modulation by lexical processing, so that it seems compatible with the spatial patterns observed in the present study, the E-Z reader model seems to mainly predict temporal effects of foveal load.

#### *Developmental Course of the Foveal Load Effects*

The present study shows that the size of the perceptual span adjusted to foveal processing load from second grade on. This adaptive adjustment was reliable in second grade, completed by third grade and remained stable thereafter. The development of foveal load effects parallels the general development of the perceptual span. This suggests that children do not have rigidly fixed perceptual spans, but that adaptability to foveal processing demands is an inherent characteristic of the perceptual span from the outset. We speculate that as soon as the perceptual span has widened, children will engage in parafoveal pre-processing most of the time. This is because in age-appropriate textbooks most words are frequent and easy to identify, leaving enough resources available for parafoveal pre-processing. Only for those few words that are unknown or low-frequency children focus on foveal information only or even make use of phonological decoding (Payne et al., 2016). Because foveal load effects did not show delayed emergence relative to the onset of perceptual span development, it seems unlikely that this characteristic of the perceptual span is acquired due to persistent practice of parafoveal processing. Rather, the underlying mechanism is likely to be a very general one that is already in place early in development. A possible candidate for such a mechanism is a gradient zoom lens of attention (Eriksen & St. James, 1986; LaBerge & Brown, 1989), as featured in the SWIFT model to account for momentary changes in the size of the attentional span in response to processing difficulty (Schad & Engbert, 2012).

*Future Directions and Limitations of the Moving Window Paradigm*

In general, to establish the validity and generalizability of an effect and to be sure that the effect is not restricted to one specific paradigm, it is vital that results are reproduced with various research methods. The multi-method approach reduces the chance that effects are due to some sort of systematic error associated with a single method (e.g., Brewer & Hunter, 1989). To date, foveal load effects have exclusively been studied using variants of the boundary paradigm (Henderson & Ferreira, 1990; Marx et al., 2017; Schroyens et al., 1999). The present study extends this literature by generalizing foveal load effects from the boundary paradigm to the moving window paradigm. Further support comes from a recent study by Luke (2018), who applied a word-based moving window with only the foveal word  $N$  or also the upcoming word  $N+1$  being visible and analyzed which lexical features of the foveal word  $N$  influenced the preprocessing of the upcoming word  $N+1$ . He found that when the foveal word was more frequent, the upcoming word was read more quickly, but only when preview of the upcoming word was available, which is in line with our findings and the foveal load hypothesis (Henderson & Ferreira, 1990).

Here, we provide a direct test of the underlying assumption of the foveal load hypothesis, namely that the perceptual span dynamically adapts to foveal processing demands. In contrast to Luke (2018), we applied a letter-based window-size manipulation. We demonstrated that in response to increased foveal load, the perceptual span momentarily narrows, which may serve as an explanation for the preview benefit observed in the boundary paradigm. However – as straightforward as this explanation might be – previous research suggests that perceptual-span size and parafoveal preview benefit may not be identical concepts. For example, older adults have been shown to have smaller perceptual spans than younger adults in the moving window paradigm (Rayner et al., 2009), but this age effect was less apparent and not significant in all eye-movement measures for the preview benefit in the

invisible boundary paradigm (Rayner et al., 2010; Risse & Kliegl, 2011). Similarly, a recent study applying the boundary paradigm to young readers failed to find evidence for foveal load effects (Marx et al., 2017). On the one hand, foveal load effects might be easier to detect in the perceptual span than in preview benefit. In fact, some studies on skilled readers also found limited evidence for foveal load effects using the boundary paradigm (Kennison & Clifton, 1995; Schroyens et al., 1999). On the other hand, Marx and colleagues' method was different from the current study and also from other boundary studies in that they applied a saliency mask to distort low-level visual features of the upcoming text instead of the typical letter mask, which distorts high-level lexical features (Vasilev et al., 2018). Taken together, this might suggest that foveal load does not impair parafoveal processing of low-level visual information, but only of high-level lexical information.

The moving window paradigm in general and the current study in particular lack precise stimulus control, which clearly is an advantage of the boundary paradigm. However, the boundary paradigm's biggest advantage might also be its biggest disadvantage: Every study using the boundary paradigm is limited to specific target words, which are typically nouns somewhere in the middle of the sentence embedded within a highly standardized and low-complexity sentence frame. This leaves little space for generalization of foveal load effects to different parts of speech, combinations of parts of speech, words of different lengths, complex sentences or paragraphs, and other naturally occurring linguistic phenomena. As such, the moving window technique has higher ecological validity than the boundary paradigm, because the materials are more natural and analyses are not restricted to target words. The moving window paradigm thus has the potential to extend the scope of studies on parafoveal processing to a more natural reading context (e.g., Luke, 2018). In particular, if using a word-based window on book texts, one might be able to conduct analyses on parafoveal processing in a continuous fashion, similar to corpus analysis

approaches that prove a valuable complement to experimental manipulations. To take a recent example from our own research: Corpus analyses have suggested that preview benefit can have benefits or costs, depending on the preview duration (Kliegl et al., 2013; Yan et al., 2012). This result has inspired a more stringent experimental test (Pan et al., 2020). We hope that studies of foveal load effects using the proposed method may be similarly inspiring for more focused experimental investigations.

### *Conclusion*

The present study contributes several important insights for models of eye-movement control during reading. First, perceptual span development continues beyond third grade, but in a much decelerated fashion and thus might not be sufficient to fully account for the lasting developmental improvements in reading rate. Second, perceptual span might not be a unitary construct as its size differs between spatial and temporal decisions of eye-movement control. In particular, information for saccade targeting is used from farther away than for the timing of saccade execution. This is true even if the mask leaves low-level features like inter-word spacing intact. Third, the size of the perceptual span is dynamically adjusted in response to foveal processing load, i.e., the perceptual span narrows when reading difficult words and widens for easy words. This adjustment seems to be an inherent characteristic of the perceptual span, because it is in place right from the earliest grades. We argue that an attentional gradient is the most likely mechanism behind the foveal load effect.

### *Context*

The present results fit nicely with other results from our research program, which investigates how top-down attention interacts with perception and cognition in everyday tasks, often using gaze-contingent manipulations. We have previously found a developmental

step in the perceptual span at the transition from letter-based to word-based reading, driven by lexical processing (Sperlich et al., 2015, 2016), semantic priming-like effects of perceptual preview (Pan et al., 2020), a reduction of the eye-voice span by working memory demands (Laubrock & Kliegl, 2015), and adaptation to perceptual viewing conditions during scene perception (Laubrock et al., 2013, Cajar et al., 2016). The methodological advances developed in Sperlich et al. (2016), driven by the need to estimate person parameters for the longitudinal evaluation of individual differences in the perceptual span, allowed us to arrive at word-based spatial and temporal span estimates.

It takes many years of practice to become efficient in the complex mental activity of reading, requiring an interplay of perceptual, oculomotor, attentional and cognitive processes. Because cognitive processes are slower than sampling the perceptual input, cognition has to be able to inhibit eye movements. Does inhibition affect what information the cognitive system considers at any given moment? The foveal load hypothesis suggests so, but has only been tested indirectly. Here we provide direct and affirmative evidence, using a novel methodological approach. Our results show that developing readers already process less upcoming information perceptually when they process a difficult word cognitively, suggesting a dynamic attentional gradient mechanism.

### References

- Andrews, S., & Veldre, A. (2019). What is the most plausible account of the role of parafoveal processing in reading? *Language and Linguistics Compass*, *13*(7), 1–25. <https://doi.org/10.1111/lnc3.12344>
- Blythe, H. I. (2014). Developmental changes in eye movements and visual information encoding associated with learning to read. *Current Directions in Psychological Science*, *23*(3), 201–207. <https://doi.org/10.1177/0963721414530145>
- Breland, H. (1996). Word frequency and word difficulty: A comparison of counts in four corpora. *Psychological Science*, *7*(2), 96–99. <https://doi.org/10.1111/j.1467-9280.1996.tb00336.x>
- Brewer, J., & Hunter, A. (1989). Multimethod research: A synthesis of styles. In *Sage library of social research, Vol. 175*. Sage Publications, Inc.
- Brysbaert, M. (2019). How many words do we read per minute? A review and meta-analysis of reading rate. *Journal of Memory and Language*, *109*(104047), 1–84. <https://doi.org/10.1016/j.jml.2019.104047>
- Brysbaert, M., Drieghe, D., & Vitu, F. (2005). Word skipping: Implications for theories of eye movement control in reading. In G. Underwood (Ed.), *Cognitive processes in eye guidance* (pp. 53–77). Oxford, UK: Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780198566816.001.0001>
- Cajar, A., Schneeweiß, P., Engbert, R., & Laubrock, J. (2016). Coupling of attention and saccades when viewing scenes with central and peripheral degradation. *Journal of Vision*, *16*(2):8, 1–19. <https://doi.org/10.1167/16.2.8>
- Deubel, H. & Schneider, W. X. (1996). Saccade target selection and object recognition:



- evidence for a common attentional mechanism. *Vision Research*, 36(12), 1827-1837. doi: 10.1016/0042-6989(95)00294-4. PMID: 8759451.
- Ehinger, B. V., Groß, K., Ibs, I., & König, P. (2019). A new comprehensive eye-tracking test battery concurrently evaluating the Pupil Labs glasses and the EyeLink 1000. *PeerJ*, 7:e7086. <https://doi.org/10.7717/peerj.7086>
- Engbert, R., Nuthmann, A., Richter, E. M., & Kliegl, R. (2005). SWIFT: A dynamical model of saccade generation during reading. *Psychological Review*, 112(4), 777–813. <https://doi.org/10.1037/0033-295X.112.4.777>
- Eriksen, C. W., & St. James, J. D. (1986). Visual attention within and around the field of focal attention: A zoom lens model. *Perception & Psychophysics*, 40(4), 225–240. <https://doi.org/10.3758/BF03211502>
- Fernández, G., Shalom, D. E., Kliegl, R., & Sigman, M. (2014). Eye movements during reading proverbs and regular sentences: The incoming word predictability effect. *Language, Cognition and Neuroscience*, 29(3), 260–273. <https://doi.org/10.1080/01690965.2012.760745>
- Findlay, J. M., & Walker, R. (1999). A model of saccade generation based on parallel processing and competitive inhibition. *Behavioral and Brain Sciences*, 22, 661–721. <https://doi.org/10.1017/s0140525x99002150>
- Hawelka, S., Schuster, S., Gagl, B. & Hutzler, F. (2015). On forward inferences of fast and slow readers. An eye movement study. *Scientific Reports*, 5:8432. <https://doi.org/10.1038/srep08432>
- Häikiö, T., Bertram, R., & Hyönä, J. (2010). Development of parafoveal processing within and across words in reading: Evidence from the boundary paradigm. *The Quarterly*

*Journal of Experimental Psychology*, 63(10), 1982–1998.

<https://doi.org/10.1080/17470211003592613>

Häikiö, T., Bertram, R., Hyönä, J., & Niemi, P. (2009). Development of the letter identity span in reading: Evidence from the eye movement moving window paradigm. *Journal of Experimental Child Psychology*, 102, 167–181.

<https://doi.org/10.1016/j.jecp.2008.04.002>

Heister, J., Würzner, K. M., Bubener, J., Pohl, E., Hanneforth, T., Geyken, A., & Kliegl, R. (2011). dlexDB - eine lexikalische Datenbank für die psychologische und linguistische Forschung. *Psychologische Rundschau*, 62(1), 10–20. <https://doi.org/10.1026/0033-3042/a000029>

Henderson, J. M., & Ferreira, F. (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(3), 417–429.

<https://doi.org/10.1037/0278-7393.16.3.417>

Hohenstein, S., Laubrock, J., & Kliegl, R. (2010). Semantic preview benefit in eye movements during reading: A parafoveal fast-priming study. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 36(5), 1150–1170.

<https://doi.org/10.1037/a0020233>

Huestegge, L., Radach, R., Corbic, D., & Huestegge, S. M. (2009). Oculomotor and linguistic determinants of reading development: A longitudinal study. *Vision Research*, 49(24), 2948–2959. <https://doi.org/10.1016/j.visres.2009.09.012>

Hyönä, J. (2011). Foveal and parafoveal processing during reading. In S. Liversedge, I. Gilchrist, & S. Everling (Eds.), *The Oxford handbook of eye movements* (pp. 819–838).

<https://doi.org/10.1093/oxfordhb/9780199539789.013.0045>

- Hyönä, J., & Bertram, R. (2004). Do frequency characteristics of nonfixated words influence the processing of fixated words during reading? *European Journal of Cognitive Psychology, 16*(1–2), 104–127. <https://doi.org/10.1080/09541440340000132>
- Hyönä, J., & Olson, R. K. (1995). Eye fixation patterns among dyslexic and normal readers: Effects of word length and word frequency. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*(6), 1430–1440. <https://doi.org/10.1037/0278-7393.21.6.1430>
- Inhoff, A. W., & Rayner, K. (1986). Parafoveal word processing during eye fixations in reading: Effects of word frequency. *Perception & Psychophysics, 40*(6), 431–439. <https://doi.org/10.3758/BF03208203>
- Joseph, H. S. S. L., Nation, K., & Liversedge, S. P. (2013). Using eye movements to investigate word frequency effects in children's sentence reading. *School Psychology Review, 42*(2), 207–222. <https://doi.org/10.1080/02796015.2013.12087485>
- Jung, J., Krahe, B., Bondü, R., Wyschkon, A., & Esser, G. (2016). The dynamic progression of antisocial behavior in childhood and adolescence: A three-wave longitudinal study from Germany. *Applied Developmental Science, 22*, 74–88.  
doi: [dx.doi.org/10.1080/10888691.2016.1219228](https://doi.org/10.1080/10888691.2016.1219228)
- Kennison, S. M., & Clifton, C. (1995). Determinants of parafoveal preview benefit in high and low working memory capacity readers: Implications for eye movement control. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*(1), 68–81. <https://doi.org/10.1037/0278-7393.21.1.68>
- Kliegl, R., Hohenstein, S., Yan, M., & McDonald, S. A. (2013). How preview space/time translates into preview cost/benefit for fixation durations during reading. *Quarterly Journal of Experimental Psychology, 66*(3), 581–600.

<https://doi.org/10.1080/17470218.2012.658073>

- Kliegl, R., Nuthmann, A., & Engbert, R. (2006). Tracking the mind during reading: The influence of past, present, and future words on fixation durations. *Journal of Experimental Psychology: General*, *135*(1), 12–35. <https://doi.org/10.1037/0096-3445.135.1.12>
- Kliegl, R., Olson, R. K., & Davidson, B. J. (1983). On problems of unconfounding perceptual and language processes. In K. Rayner (Ed.), *Eye Movements in Reading: Perceptual and Language Processes* (pp. 333–343). Acad. Press. <https://doi.org/10.1016/b978-0-12-583680-7.50024-2>
- LaBerge, D., & Brown, V. (1989). Theory of attentional operations in shape identification. *Psychological Review*, *96*(1), 101–124. <https://doi.org/10.1037/0033-295x.96.1.101>
- Laubrock, J., Cajar, A., & Engbert, R. (2013). Control of fixation duration during scene viewing by interaction of foveal and peripheral processing. *Journal of Vision*, *13*(12):11: 1–20.
- Laubrock, J., & Kliegl, R. (2015). The eye-voice span during reading aloud. *Frontiers in Psychology*, *6*, 1–19. <https://doi.org/10.3389/fpsyg.2015.01432>
- Legge, G. E., & Bigelow, C. A. (2011). Does print size matter for reading? A review of findings from vision science and typography. *Journal of Vision*, *11*(5), 1–22. <https://doi.org/10.1167/11.5.1>
- Luke, S. G. (2018). Influences on and consequences of parafoveal preview in reading. *Attention, Perception, and Psychophysics*, *80*(7), 1675–1682. <https://doi.org/10.3758/s13414-018-1581-0>
- Marx, C., Hawelka, S., Schuster, S., & Hutzler, F. (2017). Foveal processing difficulty does

- not affect parafoveal preprocessing in young readers. *Scientific Reports*, 7(41602), 1–11.  
<https://doi.org/10.1038/srep41602>
- McConkie, G. W., & Rayner, K. (1975). The span of the effective stimulus during a fixation in reading. *Perception and Psychophysics*, 17, 568–586.  
<https://doi.org/10.3758/BF03203972>
- Meixner, J. M., Warner, G. J., Lensing, N., Schiefele, U. & Elsner, B. (2019). The relation between executive functions and reading comprehension in primary-school students: A cross-lagged-panel analysis. *Early Childhood Research Quarterly*, 46, 62–74.  
doi:10.1016/j.ecresq.2018.04.010
- Nuthmann, A., Engbert, R., & Kliegl, R. (2005). Mislocated fixations during reading and the inverted optimal viewing position effect. *Vision Research*, 45(17), 2201–2217.  
<https://doi.org/10.1016/j.visres.2005.02.014>
- Nuthmann, A., Engbert, R., & Kliegl, R. (2007). The IOVP effect in mindless reading: Experiment and modeling. *Vision Research*, 47(7), 990–1002.  
<https://doi.org/10.1016/j.visres.2006.11.005>
- Pagán, A., Blythe, H. I., & Liversedge, S. P. (2016). Parafoveal preprocessing of word initial trigrams during reading in adults and children. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 42(3), 411–432. <https://doi.org/10.1037/xlm0000175>
- Pan, J., Yan, M., & Laubrock, J. (2020). Semantic preview benefit and cost: Evidence from parafoveal fast-priming paradigm. *Cognition*, 205(104452).  
<https://doi.org/10.1016/j.cognition.2020.104452>
- Perfetti, C. (2007). Reading ability: Lexical quality to comprehension. *Scientific Studies of Reading*, 11(4):357–383. doi:10.1080/10888430701530730

- Perfetti, C.A., & Hart, L. (2001). The lexical bases of comprehension skill. In D. Gorfien (Ed.), *On the consequences of meaning selection* (pp. 67-86). Washington, DC: American Psychological Association.
- Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., & R Core Team. (2015). *nlme: Linear and Nonlinear Mixed Effects Models* (R package version 3.1-120). <http://cran.r-project.org/package=nlme>
- R Core Team. (2015). *R: A language and environment for statistical computing*. (3.1.2). R Foundation for Statistical Computing. <http://www.r-project.org>
- Rayner, K. (1986). Eye movements and the perceptual span in beginning and skilled readers. *Journal of Experimental Child Psychology*, *41*(2), 211–236.  
[https://doi.org/10.1016/0022-0965\(86\)90037-8](https://doi.org/10.1016/0022-0965(86)90037-8)
- Rayner, K. (2009). Eye movements and attention in reading, scene perception, and visual search. *The Quarterly Journal of Experimental Psychology*, *62*(8), 1457–1506.  
<https://doi.org/10.1080/17470210902816461>
- Rayner, K. (2014). The gaze-contingent moving window in reading: Development and review. *Visual Cognition*, *22*(3), 242–258.  
<https://doi.org/10.1080/13506285.2013.879084>
- Rayner, K., Castelhana, M. S., & Yang, J. (2009). Eye movements and the perceptual span in older and younger readers. *Psychology and Aging*, *24*(3), 755–760.  
<https://doi.org/10.1037/a0014300>
- Rayner, K., Castelhana, M. S., & Yang, J. (2010). Preview benefit during eye fixations in reading for older and younger readers. *Psychology and Aging*, *25*(3), 714–718.  
<https://doi.org/10.1037/a0019199>

- Rayner, K., & Duffy, S. A. (1986). Lexical complexity and fixation times in reading: Effects of word frequency, verb complexity, and lexical ambiguity. *Memory & Cognition, 14*(3), 191–201.
- Reichle, E. D., Pollatsek, A., Fisher, D. L., & Rayner, K. (1998). Toward a model of eye movement control in reading. *Psychological Review, 105*(1), 125–157. <https://doi.org/10.1037/0033-295X.105.1.125>
- Risse, S., Hohenstein, S., Kliegl, R., & Engbert, R. (2014). A theoretical analysis of the perceptual span based on SWIFT simulations of the  $n + 2$  boundary paradigm. *Visual Cognition, 22*(3), 283–308. <https://doi.org/10.1080/13506285.2014.881444>
- Risse, S., & Kliegl, R. (2011). Adult age differences in the perceptual span during reading. *Psychology and Aging, 26*(2), 451–460. <https://doi.org/10.1037/a0021616>
- Risse, S., & Kliegl, R. (2012). Evidence for delayed parafoveal-on-foveal effects from word  $n+2$  in reading. *Journal of Experimental Psychology: Human Perception and Performance, 38*(4), 1026–1042. <https://doi.org/10.1037/a0027735>
- Schad, D. J., & Engbert, R. (2012). The zoom lens of attention: Simulating shuffled versus normal text reading using the SWIFT model. *Visual Cognition, 20*(4–5), 391–421. <https://doi.org/10.1080/13506285.2012.670143>
- Schotter, E. R., Angele, B., & Rayner, K. (2012). Parafoveal processing in reading. *Attention, Perception, & Psychophysics, 74*(1), 5–35. <https://doi.org/10.3758/s13414-011-0219-2>
- Schroyens, W., Vitu, F., Brysbaert, M., & D'Ydewalle, G. (1999). Eye movement control during reading: Foveal load and parafoveal processing. *The Quarterly Journal of Experimental Psychology Section A, 52*(4), 1021–1046. <https://doi.org/10.1080/713755859>

- Sperlich, A., Meixner, J., & Laubrock, J. (2016). Development of the perceptual span in reading: A longitudinal study. *Journal of Experimental Child Psychology, 146*, 181–201. <https://doi.org/10.1016/j.jecp.2016.02.007>
- Sperlich, A., Schad, D. J., & Laubrock, J. (2015). When preview information starts to matter: Development of the perceptual span in German beginning readers. *Journal of Cognitive Psychology, 27*(5), 511–530. <https://doi.org/10.1080/20445911.2014.993990>
- Tiffin-Richards, S. P., & Schroeder, S. (2015). Children's and adults' parafoveal processes in German: Phonological and orthographic effects. *Journal of Cognitive Psychology, 27*(5), 531–548. <https://doi.org/10.1080/20445911.2014.999076>
- Vasilev, M. R., Slattery, T. J., Kirkby, J. A., & Angele, B. (2018). What are the costs of degraded parafoveal previews during silent reading? *Journal of Experimental Psychology: Learning Memory and Cognition, 44*(3), 371–386. <https://doi.org/10.1037/xlm0000433>
- Vitu, F., McConkie, G. W., Kerr, P., & O'Regan, J. K. (2001). Fixation location effects on fixation durations during reading: An inverted optimal viewing position effect. *Vision Research, 41*(25–26), 3513–3533. [https://doi.org/10.1016/S0042-6989\(01\)00166-3](https://doi.org/10.1016/S0042-6989(01)00166-3)
- White, S.J., Rayner, K. & Liversedge, S.P. (2005). Eye movements and the modulation of parafoveal processing by foveal processing difficulty: A reexamination. *Psychonomic Bulletin & Review, 12*, 891–896. <https://doi.org/10.3758/BF03196782>
- Yan, M., Risse, S., Zhou, X., & Kliegl, R. (2012). Preview fixation duration modulates identical and semantic preview benefit in Chinese reading. *Reading and Writing, 25*(5), 1093–1111. <https://doi.org/10.1007/s11145-010-9274-7>



## Supplement

Table S1. Fixed-effects parameters estimated from NLME grade effect models for first fixation duration and gaze duration.

	Log First Fixation Duration								Log Gaze Duration							
	<i>asym</i>				<i>growth rate</i>				<i>asym</i>				<i>growth rate</i>			
	<i>b</i>	<i>SE</i>	<i>t</i>	<i>p</i>	<i>b</i>	<i>SE</i>	<i>t</i>	<i>p</i>	<i>b</i>	<i>SE</i>	<i>t</i>	<i>p</i>	<i>b</i>	<i>SE</i>	<i>t</i>	<i>p</i>
Intercept	5.54	.01	621.0	<.0001	-.33	.04	-7.4	<.0001	5.96	.02	361.8	<.0001	-.27	.02	-12.8	<.0001
<b>Frequency <i>N</i></b>	<b>-.03</b>	<b>.00</b>	<b>-10.0</b>	<b>&lt;.0001</b>	<b>-.16</b>	<b>.02</b>	<b>-6.1</b>	<b>&lt;.0001</b>	<b>-.05</b>	<b>.01</b>	<b>-12.0</b>	<b>&lt;.0001</b>	<b>-.08</b>	<b>.01</b>	<b>-7.9</b>	<b>&lt;.0001</b>
Frequency <i>N+1</i>	-.02	.00	-4.2	<.0001	.16	.02	7.0	<.0001	-.02	.01	-4.4	<.0001	.22	.01	16.8	<.0001
Launch Site <i>N</i>	-.01	.00	-11.8	<.0001	-.01	.01	-2.3	<.05	-.03	.00	-28.8	<.0001	-.02	.00	-5.4	<.0001
Landing Site <i>N+1</i>	.03	.00	28.0	<.0001	-.02	.01	-3.3	<.0001	-.02	.00	-11.1	<.0001	.02	.00	5.5	<.0001
Word Length <i>N+1</i>	.01	.00	15.8	<.0001	.01	.01	2.3	<.05	.09	.00	86.7	<.0001	-.04	.00	-14.0	<.0001
G2-1	-.23	.03	-7.7	<.0001	-.01	.07	-0.2	.84	-.65	.05	-11.9	<.0001	.01	.04	0.2	.87
G3-2	-.14	.02	-5.6	<.0001	-.08	.03	-2.9	<.05	-.32	.05	-6.9	<.0001	-.07	.02	-4.4	<.0001
G4-3	-.03	.02	-1.2	.23	.01	.02	0.2	.84	-.08	.05	-1.7	.09	.04	.01	2.7	<.01
G5-4	-.09	.03	-2.8	<.05	.10	.03	3.1	<.01	-.17	.06	-2.9	<.01	.00	.02	0.0	.98
G6-5	-.05	.04	-1.4	.15	-.10	.03	-3.3	<.01	-.10	.07	-1.5	.13	.02	.02	1.3	.18

Note. Dependent variables were log-transformed prior to analysis; *asym*, asymptote; *growth rate*, logarithm of growth rate (inversely related to pspan); *Intercept* correspond to the grand mean; *Frequency* in log10 counts per million according to the German DWDS (Heister et al., 2011), *Frequency* was coded as sliding difference contrast; *Launch Site N* in letters to the left of the word end; *Landing Site N+1* in letters to the right of the word beginning; *Word Length N+1* in number of letters per word; *Grade (G)* was coded as sliding difference contrasts indicating the differences in *asym* and *growth rate* between consecutive grade levels, respectively; degrees of freedom:  $df = 128938$

Table S2. Fixed-effects parameters estimated from NLME full models for first fixation duration and gaze duration.

	Log First Fixation Duration								Log Gaze Duration							
	<i>asym</i>				<i>growth rate</i>				<i>asym</i>				<i>growth rate</i>			
	<i>b</i>	<i>SE</i>	<i>t</i>	<i>p</i>	<i>b</i>	<i>SE</i>	<i>t</i>	<i>p</i>	<i>b</i>	<i>SE</i>	<i>t</i>	<i>p</i>	<i>b</i>	<i>SE</i>	<i>t</i>	<i>p</i>
Intercept	5.54	.01	621.6	<.0001	-.32	.04	-7.2	<.0001	5.96	.02	362.0	<.0001	-.27	.02	-12.9	<.0001
<b>Frequency <i>N</i></b>	<b>-.03</b>	<b>.00</b>	<b>-8.1</b>	<b>&lt;.0001</b>	<b>-.18</b>	<b>.03</b>	<b>-6.1</b>	<b>&lt;.0001</b>	<b>-.04</b>	<b>.01</b>	<b>-8.4</b>	<b>&lt;.0001</b>	<b>-.07</b>	<b>.02</b>	<b>-4.4</b>	<b>&lt;.0001</b>
Frequency <i>N+1</i>	-.01	.00	-3.9	<.0001	.17	.02	7.1	<.0001	-.02	.01	-4.1	<.0001	.22	.01	16.8	<.0001
Launch Site <i>N</i>	-.01	.00	-10.8	<.0001	-.01	.01	-2.0	<.05	-.03	.00	-27.8	<.0001	-.02	.00	-5.3	<.0001
Landing Site <i>N+1</i>	.03	.00	27.8	<.0001	-.02	.01	-3.2	<.0001	-.02	.00	-11.4	<.0001	.02	.00	5.5	<.0001
Word Length <i>N+1</i>	.01	.00	15.7	<.0001	.01	.01	2.2	<.05	.09	.00	86.6	<.0001	-.04	.00	-14.1	<.0001
G2-1	-.23	.03	-7.8	<.0001	-.03	.08	-0.4	.68	-.65	.05	-12.0	<.0001	-.00	.04	-0.1	.94
G3-2	-.14	.03	-5.7	<.0001	-.11	.03	-3.5	<.0001	-.32	.05	-7.0	<.0001	-.08	.02	-4.8	<.0001
G4-3	-.03	.03	-1.3	.20	-.00	.03	-0.2	.87	-.08	.05	-1.7	.094	.04	.01	2.8	<.01
G5-4	-.09	.03	-2.8	<.001	.10	.03	3.1	<.001	-.17	.06	-3.0	<.01	-.01	.02	-0.4	.72
G6-5	-.05	.04	-1.4	.16	-.11	.03	-3.4	<.001	-.10	.07	-1.4	.15	.03	.02	1.6	.12
G2-1 x Freq <i>N</i>	-.01	.02	-0.7	.50	-.13	.15	-0.8	0.40	-.03	.02	-1.5	.14	-.09	.08	-1.1	.26
G3-2 x Freq <i>N</i>	-.03	.01	-3.1	<.001	.07	.06	1.2	0.25	-.08	.01	-7.2	<.0001	-.10	.03	-3.0	<.01
G4-3 x Freq <i>N</i>	-.01	.01	-1.8	.08	.01	.05	0.2	0.81	.01	.01	1.2	.24	.03	.03	1.0	.30
G5-4 x Freq <i>N</i>	-.01	.01	-0.9	.38	-.03	.06	-0.5	0.59	-.02	.01	-1.7	.09	.04	.03	1.1	.26
G6-5 x Freq <i>N</i>	.01	.01	0.9	.36	.09	.07	1.3	0.20	.04	.02	2.4	<.05	.02	.04	0.5	.65

Note. Dependent variables were log-transformed prior to analysis; *asym*, asymptote; *growth rate*, logarithm of growth rate (inversely related to pspan); *Intercept* correspond to the grand mean; *Frequency* in log10 counts per million according to the German DWDS (Heister et al., 2011), *Frequency* was coded as sliding difference contrast; *Launch Site N* in letters to the left of the word end; *Landing Site N+1* in letters to the right of the word beginning; *Word Length N+1* in number of letters per word; *Grade (G)* was coded as sliding difference contrasts indicating the differences in *asym* and *growth rate* between consecutive grade levels; degrees of freedom: *df* = 128928

### **Foveal load hypothesis: Analyses with preview size as a predictor**

The present analysis aims to shed further light on the question of the timing of saccades relative to attention shifts. Previous studies have found effects of previous word frequency on fixation duration of the current word (Henderson & Ferreira, 1990; Kennison & Clifton, 1995; Schroyens, Vitu, Brysbaert, & D'Ydewalle, 1999). Two explanations have been proposed: *spillover effects* (e.g., Rayner & Duffy, 1986) and the *foveal load hypothesis* (Henderson & Ferreira, 1990). The former proposes that fixations on the *upcoming word*  $N+1$  are longer with low-frequency *previous words*  $N$  due to spillover: after the eyes have moved, processing of  $N$  continues. The second explanation proposes that fixation duration on  $N+1$  is shorter with high-frequency words  $N$  due to an increase in preprocessing on  $N+1$  when cognitive processing load is reduced: when processing of  $N$  is relatively easy, attention shifts to  $N+1$  before the eyes move (see main article for details).

In the present study, due to the nature of the moving window paradigm, there were varying degrees of word visibility of  $N+1$  during the last fixation on  $N$ . This is because window sizes were character-based (instead of word-based) and, of course,  $N+1$  varied in word length from case to case. Therefore, we ran linear mixed-effects regression (LMER) models (*lme4*-package v1.1-12) to control for the effects of the amount of preview available (i.e. whether the following word was masked, partially visible or fully visible). The aim was to confirm whether the findings from the NLME models reported in the main article would still hold when preview size is taken as predictor variable, rather than window size measured in letters. Continuous preview size was calculated from the data using the following formula:

$$\text{preview size}_{N+1j} = (\text{launch site}_{Nj} + \text{window size}_j) / \text{word length}_{N+1j}$$

with *preview size* $_{N+1j}$  being the proportion of visibility of the word  $N+1$  of sentence  $j$ , *launch site* $_{Nj}$  being the (negative) character position relative to the end of word  $N$  from which the outgoing saccade started, *window size* $_j$  being the one-sided extent in characters of the current

window-size condition, and *word length<sub>N+I</sub>* being the length of word  $N+I$ . Continuous preview size was then transformed into a three-level factor of preview (no preview of  $N+I$ , partial preview of  $N+I$ , full preview of  $N+I$ ). In the LMER models, the preview size factor replaced the letter-based window sizes that were used in the NLME models. Data conditioning was the same as in the main analysis. Frequency of  $N$  and frequency of  $N+I$  were based on median splits comparing groups of low- and high-frequency words; word length of  $N+I$  was a centered covariate.

### *First fixation duration*

The model results for log-transformed first fixation duration are summarized in Table S3. Low-frequency words  $N$  and  $N+I$  in the case of no preview are on the intercept (first row). The model shows that first fixation duration decreases with increasing preview (second and third row). Duration also decreases with high-frequency words  $N+I$  (fifth row) and longer word length (sixth row), due to refixation probability (gaze duration increased with longer words  $N+I$ , see section 1.2, below). When there was no preview, higher frequency of  $N$  slightly increased fixation times on  $N+I$  (fourth row). Most interestingly for the present analysis, however, with partial preview, first fixations were shorter after high- compared to low-frequency words  $N$  (seventh row); first fixations were shorter still for high-frequency words  $N$  with full preview (eighth row). When partial or full preview was available, the effect of  $N+I$  frequency was reduced (ninth and tenth rows).

The results are in agreement with the NLME analysis in the main article and, again, in line with the *foveal load hypothesis* (Henderson & Ferreira, 1990). The foveal load hypothesis predicts an interaction between preview size and  $N$  frequency, such that the effect of  $N$  frequency will come into play as more preview becomes available. This is because when word frequency of  $N$  is high, lexical access is easy, and attention shifts to  $N+I$  *before* the saccade is executed, so that by the time the eyes land on  $N+I$  it is already partially processed, thereby

reducing processing time as indicated by shorter fixation duration. This is, in fact, what we see in the model summary. Fixations were shorter after high-frequency words  $N$  only when at least partial preview is available; in the no-preview condition, fixation duration slightly increased with high-frequency words  $N$ , probably due to interference from pre-processing the mask. This effect was even larger with full than with partial preview. If the  $N$ -frequency effect was caused by spillover effects, we would expect to see  $N$ -frequency effects regardless of the amount of available preview, because spillover stems from continued processing *after* the eyes have moved to  $N+1$ . Therefore, the frequency-preview interaction is not in line with the assumption of the spillover hypothesis. Combined, these effects provide strong evidence for the foveal load hypothesis.

*Table S1. Linear mixed-effects regression model of log first fixation duration.*

<b>First Fixation Duration</b>	<i>b</i>	<i>SE</i>	<i>t value</i>
(Intercept)	5.704	0.012	485.3
Preview Partial	-0.024	0.006	-4.2
Preview Full	-0.086	0.006	-14.7
Frequency $N$	0.018	0.008	2.3
Frequency $N+1$	-0.071	0.007	-10.8
Word Length $N+1$	-0.005	0.001	-8.7
Preview Partial x Frequency $N$	-0.040	0.009	-4.3
Preview Full x Frequency $N$	-0.084	0.009	-9.0
Preview Partial x Frequency $N+1$	0.034	0.008	4.1
Preview Full x Frequency $N+1$	0.025	0.008	3.3
Frequency $N$ x Frequency $N+1$	0.032	0.011	2.9
Preview Partial x Frequency $N$ x Frequency $N+1$	-0.003	0.014	-0.2
Preview Full x Frequency $N$ x Frequency $N+1$	0.008	0.013	0.6

#### *Gaze duration*

The model results for log-transformed gaze duration are summarized in Table S4. Low-frequency words  $N$  and  $N+1$  in the case of no preview are on the intercept (first row). The pattern is very similar to that of first fixation duration. The model shows that gaze duration decreased with increasing preview (second and third rows). Gaze duration increased with

longer words  $N+I$  (sixth row), which lends support to our suggestion that first fixations were shorter on longer words due to refixation probability (see section 1.1, above). Gaze duration also decreased substantially with high-frequency words  $N+I$  (fifth row). As with first fixation duration, when there was no preview, gaze duration on  $N+I$  was longer after high- compared to low-frequency words  $N$  (fourth row), but for partial and full preview, gaze duration on  $N+I$  was shorter with high- compared to low-frequency words  $N$  (seventh and eighth rows). These effects follow the same basic pattern as for first fixation duration and provide further support for the foveal load hypothesis. When partial or full preview was available, the effect of  $N+I$  frequency was reduced (ninth and tenth rows).

*Table S4.* Linear mixed-effects regression model of log gaze duration.

<b>Gaze Duration</b>	<i>b</i>	<i>SE</i>	<i>t value</i>
(Intercept)	5.791	0.027	215.5
Preview Partial	-0.143	0.008	-18.9
Preview Full	-0.267	0.008	-34.6
Frequency $N$	0.063	0.010	6.1
Frequency $N+I$	-0.159	0.009	-18.1
Word Length $N+I$	0.099	0.001	119.8
Preview Partial x Frequency $N$	-0.055	0.012	-4.6
Preview Full x Frequency $N$	-0.210	0.012	-17.1
Preview Partial x Frequency $N+I$	0.055	0.011	5.2
Preview Full x Frequency $N+I$	0.115	0.010	11.5
Frequency $N$ x Frequency $N+I$	-0.030	0.015	-2.1
Preview Partial x Frequency $N$ x Frequency $N+I$	0.002	0.018	0.1
Preview Full x Frequency $N$ x Frequency $N+I$	0.104	0.017	6.2

## References

- Heister, J., Würzner, K. M., Bubenzer, J., Pohl, E., Hanneforth, T., Geyken, A., & Kliegl, R. (2011). dlexDB - eine lexikalische Datenbank für die psychologische und linguistische Forschung. *Psychologische Rundschau*, *62*, 10–20.
- Henderson, J. M., & Ferreira, F. (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.
- Kennison, S. M., & Clifton, C. (1995). Determinants of parafoveal preview benefit in high and low working memory capacity readers: Implications for eye movement control. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*, 68–81.
- Rayner, K., & Duffy, S. A. (1986). Lexical complexity and fixation times in reading: Effects of word frequency, verb complexity, and lexical ambiguity. *Memory & Cognition*, *14*, 191–201.
- Schroyens, W., Vitu, F., Brysbaert, M., & D'Ydewalle, G. (1999). Eye movement control during reading: Foveal load and parafoveal processing. *The Quarterly Journal of Experimental Psychology Section A*, *52*, 1021–1046.