Author(s): Hautala, Jarkko; Hawelka, Stefan; Ronimus, Miia

Title: An eye movement study on the mechanisms of reading fluency development

Year: 2024

Version: Published version

Please cite the original version:

Hautala, J., Hawelka, S., & Ronimus, M. (2024). An eye movement study on the mechanisms of reading fluency development. Cognitive Development, 69, Article 101395. https://doi.org/10.1016/j.cogdev.2023.101395

Cognitive Development 69 (2024) 101395

Contents lists available at ScienceDirect

Cognitive Development

journal homepage: www.elsevier.com/locate/cogdev

An eye movement study on the mechanisms of reading fluency development

Jarkko Hautala [a,c,*], Stefan Hawelka [b], Miia Ronimus [d]

[a] *Niilo Mäki Institute, Jyväskylä, Finland*
[b] *Centre for Cognitive Neuroscience, University of Salzburg, Salzburg, Austria*
[c] *Centre for Interdisciplinary Brain Research, University of Jyväskylä, Jyväskylä, Finland*
[d] *University of Oulu, Oulu, Finland*

ARTICLE INFO

Keywords:
Eye movement
Word recognition
Reading fluency
Development
Developmental dyslexia

ABSTRACT

Little is known about how word recognition processes, such as decoding, change when reading fluency improves during the school year. Such knowledge may have practical importance by determining which aspects of reading are most malleable at a certain age and reading level. The development of word-recognition subprocesses of third- and fourth-grade Finnish students (n = 81) with variable reading fluency was explored from longitudinal (6-month) text reading eye-tracking data. Generic development of the word recognition system was assessed from longitudinal changes in first fixation, average refixation durations and the number of first-pass fixations. The development of orthographic word representations and decoding was studied by examining the longitudinal changes in word frequency and word length effects, respectively. According to the results, the gain in reading fluency was mainly associated with decreases in first fixation and refixation durations. These decreases, in turn, inhibited the reduction in the number of fixations. However, students who could overcome this inhibitory effect, that is, by reading both with shorter fixation durations and with fewer fixations, developed most in reading fluency. The results seem to indicate that reading fluency development is driven by increased efficiency in representing letter strings in working memory. Over time, this development may lead to fewer fixations made into a word and, thus, more letters processed during each fixation.

1. Introduction

It is well known that reading fluency development relies primarily on visual word recognition becoming more automatized (Altani et al., 2020; Bijeljac-Babic et al., 2004; Samuels et al., 1978; Spinelli et al., 2005; Zoccolotti et al., 2009). This development is believed to stem both from a generic increase in processing speed (Zoccolotti et al., 2009) and the accumulation of specific orthographic word representations (Share, 2008). However, little is known about how word recognition subprocesses develop at the level of eye movements and what might be the underlying developmental mechanism (Huestegge et al., 2009; Reichle et al., 2013). To this end, we explored how word recognition, as reflected in readers' first-pass eye movement measures (Rayner, 1998) of first fixation duration (FFD), number of fixations (NrFix), and average refixation duration (AvgRefixDur; see Hautala et al., 2021; Huestegge et al., 2009), change longitudinally over six months as a function of time and associated gain in reading fluency. Our participants in the third and

---

* Correspondence to: Niilo Mäki Foundation, B.O. Box 29, FI-40101 Jyväskylä, Finland.
E-mail addresses: jarkko.v.hautala@jyu.fi, jarkko.hautala@nmi.fi (J. Hautala).

https://doi.org/10.1016/j.cogdev.2023.101395
Received 24 February 2023; Received in revised form 26 October 2023; Accepted 1 November 2023

fourth grades of primary education were in the phase of gradually developing their reading fluency (Eklund et al., 2015), and many of them were dysfluent readers participating in a reading fluency intervention. Besides studying generic changes, we assessed the main subprocesses of word recognition: the functioning of *orthographic word representations* was assessed by studying the word frequency effect, and the functioning of *grapheme-phoneme conversion* (GPC) was assessed by examining the word length effect. To disentangle developmental mechanisms, we analyzed the interdependency of the changes in the component measures and their joint contribution to the development of reading fluency.

1.1. Word recognition theories

According to the prevalent *dual-route view of reading* (Coltheart et al., 2001; Perry et al., 2010), the processing of a letter string starts with a parallel encoding of letters, after which the world's phonological representation can be either directly accessed via activation of orthographic word representation or by GPC. In turn, the semantic representations may be accessed via orthographic or phonological word representation. Within the direct lexical route, representations of more frequent words are activated faster, producing a word frequency effect. Words that lack a representation in the mental lexicon are read by the GPC route, producing a length effect. Although an individual reader is assumed to read a word either by the lexical or the GPC route, in practice, readers vary by which words they have already learned. Therefore, in principle, the model should predict the frequency and length interaction effect at the group level.

Although word recognition theories, such as the dual-route model, have been developed based on outcome data (response times and accuracies) to model only single-word reading, they have been widely used as a theoretical framework in reading research in general. However, the dual-route model should be critically examined in light of time-course data, such as eye movements. Recently, Hautala et al. (2021) reasoned that due to the early divergence of the routes immediately after the stage of visual letter encoding, the dual-route models would predict a word frequency and word length interaction effect to emerge from the outset of lexical processing. In contrast to this prediction, a quantile regression analysis (Yap et al., 2012) conducted for FFDs of fluent readers showed that the frequency effect was present among fixations of shorter durations than what was the case for the frequency and length interaction effect (Hautala et al., 2021). In other words, the word frequency effect preceded the interaction of frequency and length. This pattern of results was taken to indicate that activation of orthographic word representations (presumably starting during the parafoveal preview; (Marx et al., 2016) facilitates its subsequent GPC decoding. The authors labeled this a *dual-stage view* of word recognition (Hautala et al., 2021); see also (Jobard et al., 2003; Jobard et al., 2011), which we adopt as an alternative theoretical framework for the present investigation.

1.2. Development of word recognition

A comprehensive behavioral analysis of the development of single word recognition in a transparent orthography (i.e., Italian) across school grades 1 through 8 was provided by a cross-sectional study by Zoccolotti et al. (2009). According to their results, reading fluency develops rapidly during the first grades and then levels off from the third grade onwards to follow a steady linear trend. Furthermore, they found that 70 % of the development in word-recognition response times can be explained by a global factor instead of changes in the specific effects of length, frequency, and lexicality (words vs. nonwords). This result suggests that reading fluency development is mainly driven by orthographic processing becoming more efficient (Varga et al., 2020). Zoccolotti et al. (2009) also found that the word length effect reduces drastically during the first two grades, with the reduction being more pronounced for high-frequency than low-frequency words and nonwords. In contrast, the magnitude of the frequency effect was stable across grades. These results suggest that decoding familiar letter strings, particularly, becomes more efficient through development.

Theoretically, GPC is assumed to be the principal learning mechanism, the *self-teaching mechanism* of new orthographic word representations (Kyte & Johnson, 2006; Share, 2008). Increasing the collection of orthographic representations (i.e., lexicon) is regarded as central to the development of reading fluency (Álvarez-Cañizo et al., 2018; Perry et al., 2019; Share, 2008). Accordingly, the dual-route models have recently been extended to learn grapheme-phoneme associations and new words mainly relying on the GPC route (Perry et al., 2019; Pritchard et al., 2018; Ziegler et al., 2020). These simulation studies also demonstrate how a phonological deficit - assumed to underlie developmental dyslexia (Saksida et al., 2016) - may disturb the functioning of GPC and - as a consequence - the acquisition of new orthographic word representations (Perry et al., 2019; Ziegler et al., 2019). Regarding the word frequency and length effects, a previous simulation study (Dufau et al., 2010) showed that prolonged training of connectionist networks (similar to those included in CDP++) led to a generic reduction of word processing time and attenuation of word frequency and neighborhood effects. All these effects were also observed in the developmental data across Grades 1–5 (Dufau et al., 2010). Moreover, according to (Ziegler et al., 2019), the CDP++ model learns grapheme-phoneme associations rapidly and early during the model training procedure, therefore providing an explanation for the rapid decrement of word length effect during early grades (Zoccolotti et al., 2009).

Being time-course measures, eye movements may provide deeper insight into how word recognition processes develop (Blythe, 2014). In line with the behavioral data (Zoccolotti et al., 2009), also the eye movement measures of reading (including progressive and regressive fixations, saccade amplitude, and fixation duration) develop most during the early school grades when children acquire reading ability (Blythe & Joseph, 2011; Blythe et al., 2009; Buswell, 1922; De Luca et al., 2010; Hutzler & Wimmer, 2004; Häikiö et al., 2009; Kim et al., 2022; McConkie et al., 1991; Rayner, 1986; Sperlich et al., 2015; Spichtig et al., 2017; Vorstius et al., 2014; Taylor, 1965). We found that the earliest development reported in these studies (most commonly from Grade 1–2) is characterized by a proportionally larger decrease in the saccadic measures (reduction in number of fixations, refixation probability, or increase saccade amplitude) than in fixation duration (25 % vs. 13 %). In contrast, at the age range of 9–12 years, the saccadic and fixation duration measures develop hand-in-hand with an annual rate of 7 %. This annual development in basic eye movement measures likely reflects the global developmental factor identified by Zoccolotti et al. (2009). In a rare intervention study, Judica et al. (2002) found a decrease in fixation durations of 11-year old dyslexic children, while a modest reduction in the number of fixations was also observed both in the intervention and control group. This result may be understood owing to each fixation lasting several hundred milliseconds, so a reduction of the number of fixations would require a rather drastic development in word recognition ability, which, in turn, is harder to achieve in later grades when children's reading development trajectory is already quite established (Eklund et al., 2015).

Moreover, several studies have reported a reduction in the word length effect in gaze duration (GD, i.e., the sum duration of all first-pass fixations) across years in cross-sectional (Joseph et al., 2009; Tiffin-Richards & Schroeder, 2015; Rau et al., 2014) and longitudinal studies (De Luca et al., 2010; Huestegge et al., 2009; Schmidtke & Moro, 2021; Sperlich et al., 2015). Also, the word frequency effect in FFD and later measures such as GD diminishes over the years (Tiffin-Richards & Schroeder, 2015; Khelifi et al., 2019). The same developmental trend may apply to the word frequency and length interaction effect (Tiffin-Richards & Schroeder, 2015).

Theoretically, the developmental changes in readers' eye movements were more parsimoniously simulated by orthographic-lexical rather than visual-oculomotor processing, becoming more efficient (Reichle et al., 2013). This conclusion aligns with the findings that children are virtually equally fast as adults in intaking visual information in disappearing text experiments (Blythe et al., 2009) and that reading development seems not to be explained by improvement in attentional control of eye movements (Huestegge et al., 2009).

The investigations of reading fluency on readers' eye movement behavior are also informative about reading development. Hautala et al. (2021) found that the dual-stage pattern of results (a main effect of word frequency preceding the interaction of frequency and word length) was delayed in less fluent readers, manifesting in highly inflated FFDs with a pronounced frequency effect, followed by a pronounced length effect in refixation probability and a strong frequency and length interaction effect in summed refixation duration. Such a pattern of results has also been reported among readers of varying ages (Calvo & Meseguer, 2002; Kliegl et al., 2004; Rau et al., 2014; Schroeder et al., 2021; Tiffin-Richards and Schroeder, 2015). Moreover, Hautala et al. (2021) found that while FFD explained 35 % of the variance in reading fluency, the length effect in SRD explained an additional 14 %. Hautala et al. suggested that prolonged FFDs may reflect slower letter encoding (see also (Paizi et al., 2013) and the pronounced length effect in refixation probability and summed refixation duration a slower working of GPC (Blomert, 2011; Bouma & Legein, 1980; Gutezeit, 1976; Leinonen et al., 2001; O'Brien et al., 2011; Perry et al., 2019).

Taken together, after the initial phase of reading acquisition, readers continue to gradually build up their reading fluency by becoming more efficient in orthographic processing and decoding familiar and novel letter strings. In eye movements, these later developments manifest in an equivalent reduction of the number of fixations and their durations and a continued reduction of word length and frequency effects. However, the developmental mechanisms responsible for these changes require further research.

1.3. *The present study*

To gain new knowledge about the development of word recognition after the initial reading acquisition, we conducted a detailed explorative analysis of short-term longitudinal changes in 3rd and 4th-grade readers' first-pass viewing measures of words (FFD, NrFix, AvgRefixDur) during text reading. We first estimated the effects of time, gain in proxy of reading fluency, word frequency, and length and their possible interactions on each measure. Following Hautala et al. (2021), we operationalized the reduction in FFD (and its word frequency effect) to reflect improvement in orthographic processing and the reduction in NrFix and AvgRefixDur (and their word length effect) to reflect development in decoding. To discover the developmental mechanism, we studied the interrelationship between the longitudinal changes in the dependent measures (Hautala et al., 2021). Although at this age, the number of fixations and their durations seem to develop hand-in-hand, results of an intervention study (Judica et al., 2002) suggest that changes in fixation duration may be more easily achieved in the short-term and thus precede a reduction of refixations. Furthermore, the concurrent development of FFD and AvgRefixDur has not been studied previously (Hautala et al., 2021; Huestegge et al., 2009). Their concurrent development would speak for the global developmental process (Zoccolotti et al., 2009), whereas their developmental dissociation would speak for the relatively independent development of orthographic processing and decoding.

There are methodological challenges in studying highly intercorrelated measures of first-pass eye movements (Hautala et al., 2021) and their changes over time. First, there may be trade-off effects between the number of fixations and their durations: FFDs may shorten when multiple fixations are made into a word, therefore obstructing the detection of a word length effect in FFD (Loberg et al., 2019; Sperlich et al., 2015; Vitu et al., 2001). However, such effects seem not to have been reported in previous multiline text–reading studies (Hautala et al., 2021; Hutzler & Wimmer, 2004; Hyönä & Olson, 1995). Multiline text reading studies employ a "corpus" approach of estimating effects across all presented words instead of studying just target words presented in highly controlled sentence frames. These two approaches have produced highly converging results (Kliegl et al., 2004), suggesting that multiline data is also suited for studying robust word recognition processes. An additional benefit of the corpus approach is to derive a maximal amount of children's word viewing instances to reliably estimate highly intercorrelated word frequency and length effects and their interactions.

Our research questions were:

1) How do readers' first-pass eye movements change as reading fluency develops? We hypothesize that reading fluency development manifests primarily in reduced fixation durations (FFD, AvgRefixDur) and to a smaller extent in a reduction in NrFix (Huestegge et al., 2009; Judica et al., 2002). In addition, these changes may be accompanied by the reduction of word frequency and length effects as a reflection of improvements in orthographic processing and decoding, respectively (Zoccolotti et al., 2009).
2) What changes predict changes in other component measures and reading fluency development? We hypothesize that decreases in fixation duration explain a reduction in the number of fixations and most variance in the development of reading fluency (Hautala

et al., 2021; Huestegge et al., 2009; Judica et al., 2002). Such concurrent changes would then provide an eye-movement characterization of the global developmental process (Zoccolotti et al., 2009).

The results will be discussed based on potential developmental mechanisms derived from the dual-route and dual-stage views of word recognition.

2. Methods

2.1. Participants

The participants in this study were 81 voluntary students (44 girls, 37 boys) with a mean age of 10;1 (years;months, SD = 7 months). They were 37 students at Grade 3 (aged 9;5, SD = 4) and 44 at Grade 4 (aged 10;5, SD = 4). These students participated in both measurements at time point 1 (T1) in December and time point 2 (T2) in June. The results of T1 (with additional participants) have been previously reported by (Hautala et al., 2023). The participants came from five of ten schools participating in a large-scale (n = 318) Readers' Theater oral reading fluency intervention study (Hautala et al., 2023). Among the present sample, 37 students belonged to the typically reading control group, while 44 were dysfluent readers participating in one of the three reading fluency interventions (Hautala et al., 2023).

All students followed the standard curriculum, with school instruction provided in Finnish. The study was pre-evaluated by the Ethical Committee of the University of Jyväskylä. The research was conducted according to the ethical principles for medical research involving human subjects set forth by the Declaration of Helsinki. Informed written consent was collected from students and their caregivers.

2.2. Reading measures

The participants were first screened (Hautala et al., 2023) in November for their reading fluency and accuracy with a standardized word list (Lukilasse 2; (Häyrinen et al., 2013) and text reading tasks (FirstSteps Study; (Lerkkanen et al., 2006). The average reading fluency of the participants in this study was low according to grade-specific standardized norms (M = −0.62, SD = 1.11, range: −3.13 to 1.78) but normally distributed (skewness = −.035, SE =.27).

Participants completed three reading tasks at T1 (December) and T2 (May-June): (1) a group-administered, computerized, and time-limited sentence verification task with correctly answered sentences in two minutes as the outcome measure (Cronbach alpha reliability α = .94; Eklund et al., 2013; see also Hautala et al., 2020), (2) an individually administered expressive reading aloud task (anonymized for review) with correct words per minute as the outcome measure, and (3) an eye-tracking experiment of silent text reading with participant total fixation duration (PTFD) to all read words as the proxy of reading fluency. See the Analyses and Results sections on how the reading measures were used in the explorative data analysis.

2.3. Apparatus

Eye movements were recorded with remote 250 Hz sampling rate eye-tracking devices (SensoMotoric Instruments GmbH) installed on laptop computers (screen size of 34.5 × 19.5 cm). The measurements were conducted in a dimly lit room and controlled by at least one experimenter. Fully adjustable chinrests modified from camera mounts were used to stabilize the participants' heads while they sat on a non-adjustable chair. The texts were presented with the SMI Experiment Center 3.6 program on eleven five-line screens with no option to return to previous screens. Arial 28 pt. font was used, corresponding to approximately five letters per degree of visual angle at a 60 cm viewing distance. A full-screen 13-point calibration routine was completed before the student read each of the four story parts, and a 4-point calibration validation routine was completed after the third of 5–6 screens for each story part.

2.4. Procedure

Instructions for the task were given simultaneously via on-screen text (and through headphones at T1). Practice included a full calibration routine, two practice text screens, and multiple-choice comprehension questions. Then, the calibration was repeated, and the students proceeded to the actual experiment. After reading a text screen, students proceeded to the next screen by looking at a large gaze-sensitive area centered on a target arrow in the right-bottom corner. A pause intervened between the two story parts, allowing the children to lift their heads from the chinrest before recalibrating and continuing.

Some adaptations to the procedure were made due to the COVID-19 pandemic leading to nationwide school closures in April 2020. While the T1 experiments were conducted in an available room at schools with four eye trackers, the T2 experiments in all but one school (there was no effect of school on PTFD) were conducted in individual testing rooms at a research site with two eye tracker devices. Healthcare officials allowed the measurements to continue with special precautionary hygienic arrangements. Although there was a substantial drop-out from T1 (n = 142) to T2 (n = 81), it was not selective according to reading fluency (F(1, 141) = .289, p = .594) or gender (Chi-Square =.132, p = .717).

2.5. Materials

Participants read two abridged and modernized Finnish versions of classic stories: *Little Heidi* by Johanna Spyri (1881) (Finnish abridgment by Kati Weiss) and *Adalmina's Pearl* by Zacharius Topelius (1865), available at https://iltasatu.org/. The story read at T1 continued at T2 with a short researcher-added introductory sentence: 'Do you remember Little Heidi/princess Adalmina, who.)'. Minor changes to wording were also made to avoid single-line experimental screens. Word frequency, minimum syllable frequency in a word (an index of sublexical difficulty), and the number of syllables were derived from the latest published corpora (Huovilainen, 2018). All corpus frequency measures were log10-transformed. Word length and frequency were correlated: $r = -0.748$ at T1 and $r = -.716$ at T2. In addition, an index of word predictability in the form of two-gram transitional probability was derived from the Finnish N-gram corpus National Library of Finland, (2014). The psycholinguistic properties of the four-story parts were highly comparable (Table 1).

Each of the four story parts was followed by five multiple-choice (four response options) comprehension questions and one Yes/No question about whether the story was familiar to the reader. Reading comprehension questions were answered with an accuracy of M = 79 %, SD = 16 %, range: 30–100 % at T1 and M = 81 %, SD = 17 %, range: 30–100 % at T2. Two students performing below 50 % accuracy both at T1 and T2 were excluded from the analyses, as were data from three students either at T1 or T2. At T1, 18 children were familiar with one of the stories (9 in control and 9 in intervention groups); 5 knew both, but knowing the story did not affect reading comprehension accuracy ($F(1, 78) = 0.163$, $p = 0.687$).

2.6. Eye-movement data processing

Data preprocessing was conducted with the SMI Begaze 3.6 program. A sensitive saccade detection parameter of 20 deg/s minimum angular velocity, a saccade duration of 15 ms, and a minimum fixation duration of 50 ms were applied to detect refixation saccades with small amplitudes. The vertical boundaries of the automatically generated word-specific areas of interest were manually extended to the center position between the lines.

Trained research assistants manually inspected scan paths of all screen recordings to correct systematic drifts (415 or 23.3 % of screens) in the data and mark occasions where data were of poor quality (33 or 1.9 % of screens). The inter-rater agreement regarding whether to correct a screen was 95 % for the first text screen at both T1 and T2.

The first-pass fixations were identified with a custom script in the SPSS 26 program. The dependent variables were FFD (the very first fixation on a word), NrFix (number of first-pass fixations), AvgRefixDur (average duration of all first-pass refixations), and single fixation duration (Appendix A1). Initially, we also computed and analyzed refixation probability (RP), summed refixation duration (SRD; (Hautala et al., 2021; Huestegge et al., 2009), and gaze duration (GD, the summed duration of all first-pass fixations), whose results are not reported here due to the following reasons: NrFix was preferred over RP since many readers frequently make more than two first-pass fixations on long words, and GD was not needed for answering the present research questions. The authors reasoned that the interpretation of AvgRefixDur is more straightforward than the summative SRD measure.

Because the first-pass fixation identification is highly affected by noise in eye-movement signals, additional filtering following Hautala et al. (2021) was applied: To exclude skimming, only cases in which more than 60 % of words on a line were fixated were included in the analysis (96.5 % of all line readings). Single return-sweep fixations that did not land on the next text line's initial word were excluded (Slattery and Parker, 2019). Together, the exclusions affected 6 % of the words. In addition, extreme values of + /- 3 SD from the individual mean (Gerth & Festman, 2021) were excluded from the analyses, affecting 1627 (1.7 %) instances of FFD, 2099 (2.1 %) instances of NrFix, 562 (1.9 %) instances of AvgRefixDur. Only instances in which a word was fixated were included in the analyses (i.e., skipped words were handled as missing values).

2.7. Analyses

We report how we determined our sample size, all data exclusions, all manipulations, and all measures and analyses in the study (Simmons et al., 2012). The overall rationale of the analyses follows Hautala et al. (2021) by first identifying relevant effects with separate linear mixed models (LMM) and then by studying the interconnectedness of these effects with a hierarchical regression analysis.

First, we identify the most relevant reading measure for further analyses by inspecting the correlations between the reading measures, first-pass eye movement measures, and their gains (T2 minus T1 values). We then ran LMMs for each dependent measure (see supplementary file for R scripts) in R with the *lme4*-package (version 1.1–21; (Bates et al., 2019) controlled through the *afex*--package (version 0.25–1; (Singmann et al., 2015), which allows model estimations with uncorrelated random effects. Here, it is

Table 1
Mean (SD) of the psycholinguistic properties of the stimulus texts.

	Words		WL		WF (log)		MinSyl (log)		Twogram (log)	
	T1	T2	T1	T2	T1	T2	T1	T2	T1	T2
Little Heidi	457	462	6.3 (2.6)	6.3 (2.5)	1.7 (1.3)	1.7 (1.3)	2.8 (0.7)	2.8 (0.8)	1.9 (1.1)	1.9 (1.2)
Adalmina's Pearl	403	421	6.4 (3.0)	6.6 (2.8)	1.9 (1.4)	1.7 (1.4)	2.9 (0.8)	2.7 (0.81)	1.9 (1.2)	1.9 (1.1)

Note. Abbreviations: WL = word length, WF = word frequency in a million words, MinSyl = minimum syllable frequency in a million words, Twogram = frequency of occurrence in a million words given a preceding word.

important to control for the global effect of processing speed (Zoccolotti et al., 2009), which directly influences the size of specific effects (e.g., frequency and length). Logarithmic transformation (base 10) of dependent variables allows testing proportional effects (see (Martens & de Jong, 2008). To establish whether there was a significant development in reading fluency during the relatively short (6-month) study period and whether this development was due to intervention, we ran a model log10 (PTFD) $\sim$ Group $\times$ Time $+$ $(1 + \text{Time}||\text{id}) + (1|\text{word})$.

We then proceeded to estimate to which extent first-pass eye movement measures and their associated word frequency and word length effects change as a function of time and reading fluency gain. As per reviewer recommendations, we changed our analysis strategy from data-driven model building to testing only a theoretically meaningful full model and omitted the factors of reading fluency and two-gram predictability. To this end, we run the model log10 (dependent measure) $\sim$ Time $\times$ Gain $\times$ Word frequency (WF) $\times$ Word length (WL) $+$ Minimum syllable frequency (MinSyl) $+$ (1 $+$ Time $+$ WL $+$ WF $||$ participant) $+$ (1 $|$ word). The ordinal measure of the number of first-pass fixations was analyzed with Poisson distribution with the logarithmic link function. The final models ' variance inflation factor values (vif-function of the car-package; (Fox et al., 2012) were below 5, indicating no multicollinearity between predictors (O'Brien, 2007). Residual diagnostics showed near-zero correlations ($|r| < 0.06$) between model residuals and values of dependent or independent variables. Satterthwaite approximation of degrees of freedom was used in statistical testing. We report standardized beta estimates (b') to facilitate the comparison between effects.

Finally, to study the interdependencies of the changes in FFD, NrFix, and AvgRefixDur (in 10-based logarithmic values) and their joint contribution to reading fluency development, a hierarchical regression analysis was conducted in SPSS 28 program (IBM). We used mean-based estimates instead of individual random slope coefficients derived from LMMs (Hautala et al., 2021). This decision was made because the slope coefficients may not be directly comparable between different models (Kliegl, 2023).

3. Results

3.1. *Reading fluency*

Table 2 presents correlations between the reading measures at T1, their gains, and their correlations to averages of first-pass eye movement measures. All reading measures and first-pass eye movement measures correlated highly with each other at T1. However, the gains between different reading fluency measures correlated only weakly. These results align with the previously reported finding of the present intervention effects being specific to oral reading speed (anonymized for review). Moreover, only the gain in participant total fixation duration (PTFD) correlated with the gains in first-pass eye movement measures. On this basis, the PTFD was used as a proxy of reading fluency, and its gain was used as an index of reading fluency development. The split-half reliability of PTFD was high ($r = .99$ at T1 and $r = .94$ at T2), as well as for its gain ($r = .89$). The individual gain in PTFD correlated minimally ($rs < .12$) with the word-level values of FFD, NrFix, and AvgRefixDur suggesting that including the PTFD gain as a predictor in the analyses does not induce problems in multicollinearity.

Next, we analyzed to which extent PTFD developed from T1 to T2 and to which extent this development was due to interventions. The main effect of Group ($b' = 0.34$, $SE = 0.04$, $t = 8.38$, $p < .001$) showed that RF was much better in the control group ($M = 313$ ms, $SE = 14.5$ ms) than in the intervention groups ($M = 508$ ms, $SE = 22.2$ ms). Main effect of time ($b' = -0.05$, $SE = 0.02$, $t = -2.59$, $p < .05$) indicated that reading fluency developed by 10 % from T1 ($M = 419$ ms, $SE = 13.6$ ms) to T2 ($M = 379$ ms, $SE = 13.6$ ms). The interaction between Group and Time was insignificant ($b' = -0.05$, $SE = 0.03$, $t = -1.59$, $p = .12$), indicating that the development was not due to the interventions. However, there was substantial individual variability in the development because omitting the random slope of Time resulted in an inferior model in terms of Bayesian Information Criterion (BIC) (18,909 vs. 17,912).

3.2. *First fixation duration*

There were 96,702 observations for 79 participants and 1743 words. Table 3 provides standardized beta estimates, standard errors,

Table 2
Selected correlations between the reading measures, the first-pass eye movement measures, and the gain measures.

	ORF, T1	SV, T1	PTFD, T1
Sentence Verification (SV), T1	.79**		
Participant Total Fixation Duration (PTFD), T1	-.77**	-.73**	
First Fixation Duration (FFD), T1	-.72	-.67	.88
Number of 1st-Pass Fixations (NrFix), T1	-.67	-.70	.83
Average Refixation Duration (AvgRefixDur), T1	-.72	-.68	.91
	ORF, Gain	SV, Gain	PTFD, Gain
SV, Gain	.25*		
PTFD, Gain	.06	-.04	
FFD, Gain	.09	.04	.57**
NrFix, Gain	.04	-.05	.12
AvgRefixDur, Gain	.07	.08	.64**

Note. Abbreviations: ORF = Oral reading fluency.

* $p < 0.05$.

** $p < 0.01$,

Table 3
Generalized) linear mixed model results for the component measures.

	First Fixation Duration				Number of Fixations				Average refixation duration[b]			
Effect	b′	SE	t	Sig	b′	SE	z	Sig.	b′	SE	t	Sig.
Gain	0.158	0.040	3.965	***	0.089	0.023	3.96	***	0.216	0.048	4.46	***
WL	0.016	0.011	1.47		0.153	0.011	14.18	***	0.053	0.015	3.51	***
WF	-0.049	0.011	-4.34	***	-0.056	0.008	-6.73	***	-0.125	0.017	-7.58	***
Time	-0.039	0.016	-2.41	*	-0.030	0.007	-4.30	***	-0.065	0.014	-4.54	***
MinSyl	0.018	0.006	2.96	**	-0.003	0.005	-0.67		-0.005	0.008	-0.62	
Gain × WL	-0.012	0.007	-1.69	·	0.024	0.011	2.23	*	-0.002	0.011	-0.17	
Gain × WF	-0.005	0.008	-0.65		-0.023	0.008	-2.80	**	-0.027	0.012	-2.21	*
WF × WL	0.015	0.007	2.05	*	-0.025	0.006	-4.25	***	-0.052	0.012	-4.47	***
Gain × Time	-0.109	0.023	-4.72	***	-0.004	0.010	-0.36		-0.108	0.020	-5.30	***
WL × Time	-0.012	0.010	-1.28		0.006	0.007	0.83		0.016	0.014	1.19	
WF × Time	-0.023	0.010	-2.25	*	0.008	0.008	1.01		0.009	0.015	0.57	
Gain ×WL × WF	0.004	0.006	0.72		0.002	0.007	0.34		-0.007	0.010	-0.66	
Gain ×WL × Time	0.007	0.006	1.10		0.001	0.007	0.08		0.003	0.010	0.28	
Gain × WF × Time	-0.001	0.007	-0.08		0.010	0.008	1.22		0.007	0.011	0.62	
WL × WF × Time	-0.022	0.008	-2.88	**	-0.004	0.006	-0.61		-0.027	0.012	-2.19	*
Gain × WL × WF × Time	-0.010	0.006	-1.73	a	0.002	0.007	0.24		0.005	0.010	0.45	

Note. Abbreviations: WL = word length. WF = word frequency.
* $p < .05$.
** $p < .01$.
*** $p < .001$.
[a] $p < .1$.
[b] Analysis of Gaze Duration produced an equal pattern of results.

and statistical test results. The grand mean was 249 ms ($SE = 6.72$ ms). The main result was the significant interaction of Gain × Time ($b' = -0.109$), which indicates that the reduction from T1 to T2 in FFD increases as a function of gain in PTFD (Fig. 1A); actually, the FFDs increased from T1 to T2 for the lowest gainers. Another important result was the significant interaction of WF x Time ($b' = -0.023$), resulting from the FFDs to higher frequency words reducing more from T1 to T2 than for lower frequency words (Fig. 1B).

The main effect of Gain ($b' = 0.158$) reflects the positive correlation ($r = .33$; see Appendix A2) between fluency and gain, indicating that less fluent readers tended to make higher gains in reading fluency. The effect of Time ($b' = -0.039$) indicated that FFDs reduced over time. The effect of WF ($b' = -0.049$) indicated a shorter fixation duration for more frequent words. The effect of minimum syllable frequency ($b' = 0.018$) may stem from lexical competition being stronger among words that do not contain constraining low-frequency syllables (Hawelka et al., 2013). There was no main effect of word length on FFD. Given that refixation probability drastically increases as a function of word length, this result also indicates a lack of trade-off effect between the number of fixations and FFD.

The interaction of WL x WF ($b' = 0.015$) indicated a slightly larger word length effect for less frequent words. The significant three-level interaction of WL x WF x Time ($b' = -0.022$) resulted from minor time-related changes in the word length effect for words of the highest frequency, the effect being + 9 ms at T1 and − 5 ms at T2 across words of + /- 1 SD from average word frequency. We have no theoretical interpretation for this unexpected result.

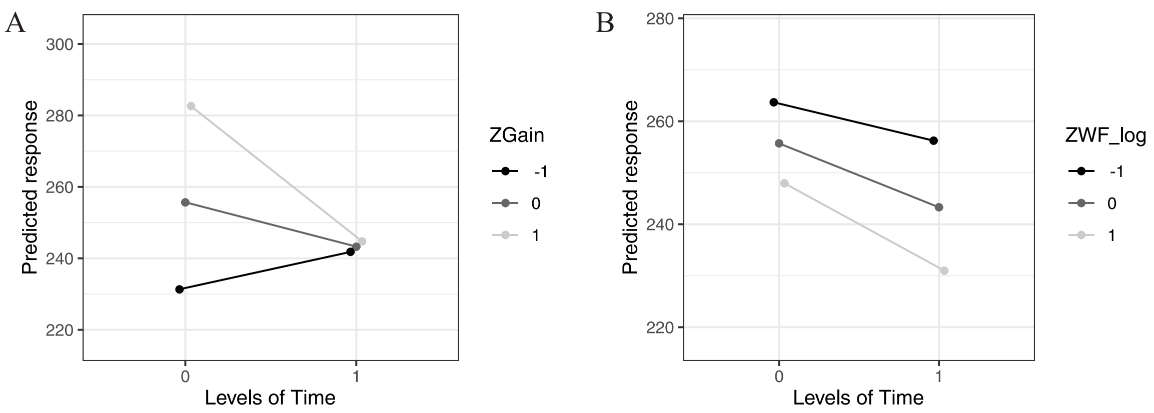

Fig. 1. Estimated marginal means of the first fixation duration (msec) for the Time × Gain (panel A) and WF x Time (panel B) interactions. Z refers to standardized values.

3.3. *Number of fixations*

See Table 3 for statistical test results and Fig. 2 for selected marginal means. The grand mean was 1.35 (SE = 0.023). The main result was the Gain x Time -interaction (Fig. 2A) was insignificant ($b' = -0.004$), indicating that gain in PTFD did not directly translate into fewer fixations made into words. Regression analyses will further investigate this unexpected result (see below).

All main effects but MinSyl ($b' = -0.002$) were highly significant ($p < .001$), WL ($b' = 0.153$), WF ($b' = -0.056$), Gain ($b' = 0.089$), and Time ($b' = -0.030$). These results indicate that more fixations were made to longer and less frequent words, by high gainers and at T1 than T2, respectively.

Three two-level interactions reached significance. Gain x WL ($b' = 0.024$) and Gain x WF ($b' = -0.023$) indicated that the effects of length and frequency were slightly larger for students showing larger gains in PTFD (less fluent readers). In addition, WL x WF ($b' = 0.025$) indicated the effect of length being larger for less frequent words (Fig. 2B).

3.4. *Average refixation duration*

There were 29,861 observations for 79 participants and 1700 words. The grand mean was 210 ms (*SE* = 6.23 ms), being considerably shorter than the mean FFD (249 ms). See Table 3 for statistical test results and Fig. 3 for selected marginal means. The main result was the significant interaction of Gain × Time ($b' = -0.108$), which indicates that the reduction from T1 to T2 in FFD increases as a function of gain in PTFD (Fig. 3A). Another important result was the significant three-level interaction of WL x WF x Time ($b' = -0.027$), resulting from the WL x WF interaction being proportionally larger at T2 than at T1. Fig. 3B shows how reading short infrequent words improved most.

All main effects except MinSyl ($b' = -0.005$) were significant, Gain ($b' = 0.216$), Time ($b' = -0.065$), WF ($b' = -0.125$), WL ($b' = 0.053$). These results indicate that refixation durations were longer for high gainers (less fluent readers), at T1 than T2, and for less frequent and longer words.

Also, the two-level interactions of Gain x WF ($b' = -0.027$) and WL x WF ($b' = -0.052$) were significant. The former effect indicated that WF was more prominent for high gainers, and the latter indicated that the length effect was more notable for less frequent words.

3.5. *Regression analysis*

We ran two hierarchical regression analyses to understand the possible interrelations of the changes in the component measures and how they contribute to reading fluency development. Table 4 presents model results and Appendix A2 correlations. The component measures correlated positively ($r = 0.56–0.97$) with each other at T1, as well as the changes from T1 to T2 between FFD and AvgRefixDur ($r = 0.81$). However, there were unexpected negative correlations between the reduction in fixation duration measures and the reduction in NrFix ($r = -0.50$ for AvgRefixDur and -0.60 for FFD). This result means that a larger reduction in fixation duration is associated with a smaller reduction or even a slight increase in the number of fixations. Thus, there is a trade-off between a reduction in the number of first-pass fixations and their duration.

The hierarchical regression analysis revealed that after controlling for the initial level of PTFD (R^2 =11.2 %) and T1-T2 changes in fixation durations (+R^2 =38.8 %), gain in NrFix explained a substantial amount of additional variance (+25.4 %) in PTFD gain (see Model 1 in Table 4). The coefficient was positive ($b' = 0.68$), meaning that a larger reduction in NrFix was associated with a larger gain in PTFD. However, the gain in NrFix did not explain any variance in PTFD gain when added to the model before the gains in fixation duration measures (Model 2 in Table 4). In other words, only when reduction in fixation duration is first controlled for, reduction in the number of fixations indeed contributes to reading fluency development.

We interpret these results as NrFix acting as a suppressor variable (see Thompson and Levine, 1997). A suppressor variable has no

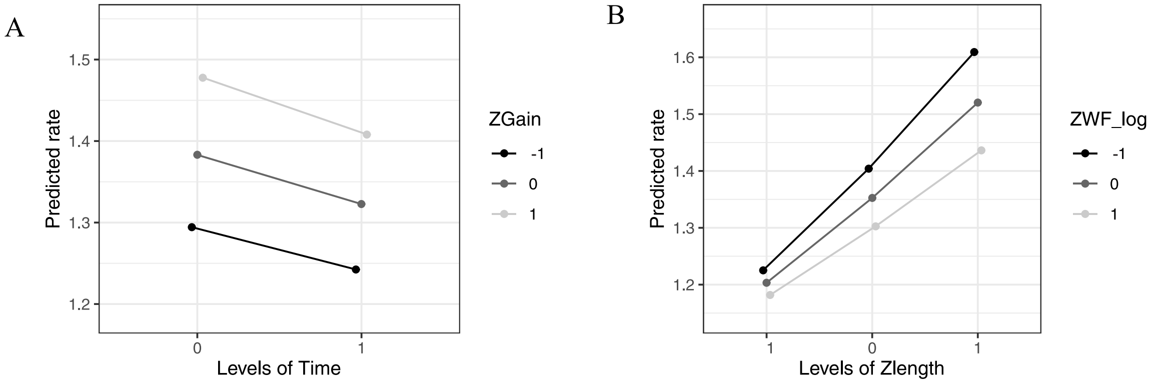

Fig. 2. Estimated marginal means of the number of first-pass fixations for the non-significant Time × Gain (panel A) and significant WF x WL (panel B) interactions. Z refers to standardized values.

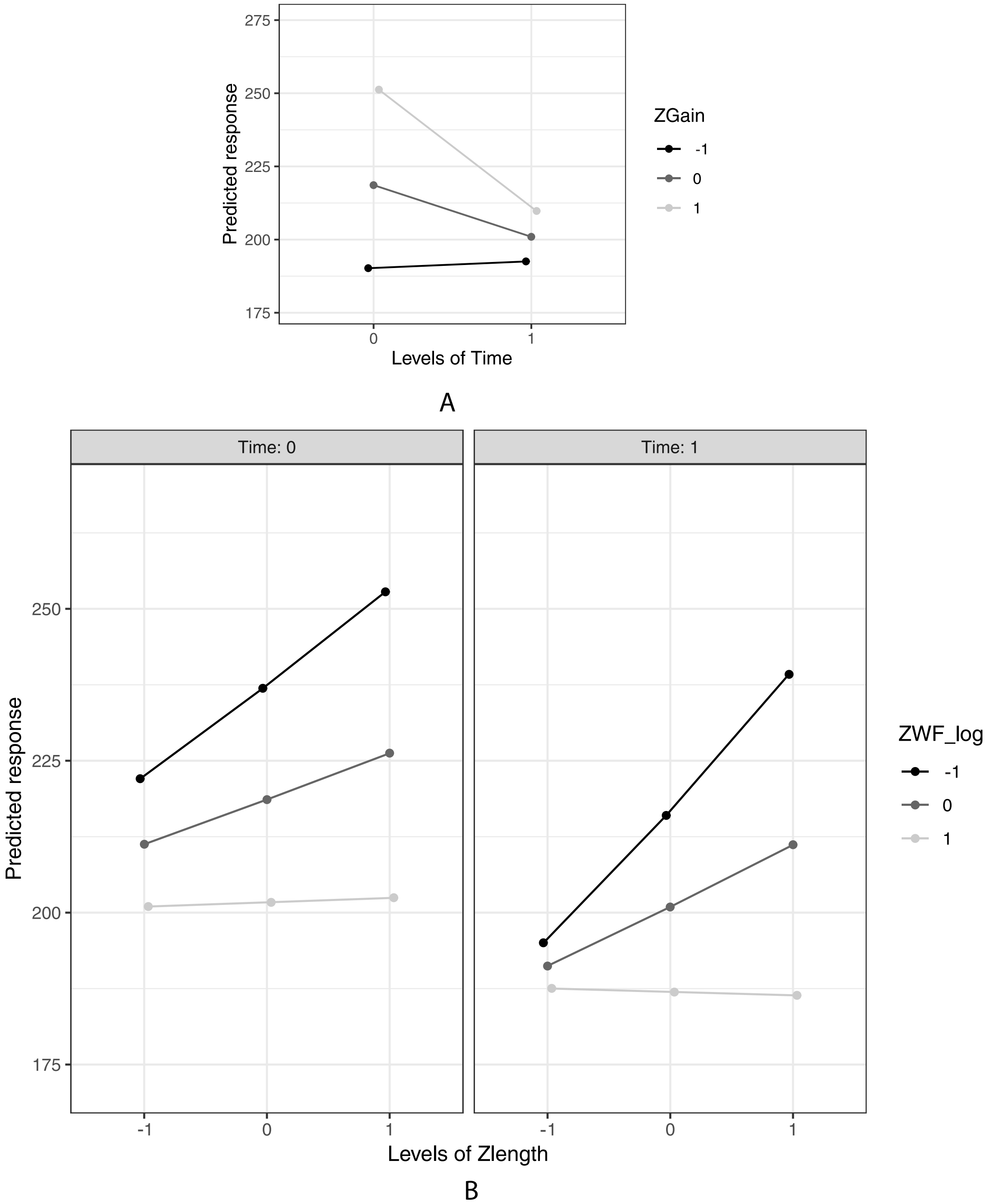

Fig. 3. Estimated marginal means of the average refixation duration for the Gain × Time interaction (panel A) and WF × WL × Time interaction (panel B). Z refers to standardized values.

association with a dependent variable (r = 0.04 here) but can explain the residual variance left by another predictor. Scatterplots suggest that the change in NrFix may reflect two counteracting developmental processes: First, by average, reduction in fixation durations inhibits the reduction of the number of fixations (Fig. 4A). Second, those students who can overcome this inhibitory effect seem to show larger PTFD gains (Fig. 4B). Because these effects have opposing directions, the net effect in the reduction of NrFix is zero on PTFD gain (Fig. 4C).

Table 4
Results of the hierarchical regression analyses of gain in reading fluency.

	Step	Predictor	b'	R	R^2	Adj. R^2	Est. SE	ΔR^2	ΔF (1, <80)
Model 1	1	PTFD, T1	-0.014	0.334	0.112	0.100	0.15	0.112	9.7**
	2	FFD gain	0.735***	0.684	0.468	0.454	0.11	0.356	50.8***
	3	AvgRefixDur gain	0.398***	0.707	0.499	0.479	0.11	0.032	4.77*
	4	NrFix gain	0.679***	0.868	0.754	0.740	0.08	0.254	76.4***
Model 2	1	PTFD, T1	-0.014	0.334	0.112	0.100	0.15	0.112	9.7**
	2	NrFix gain	0.679***	0.335	0.112	0.089	0.15	0.000	0.04
	3	FFD gain	0.735***	0.837	0.701	0.689	0.09	0.588	147.4***
	4	AvgRefixDur gain	0.398***	0.868	0.754	0.740	0.08	0.053	15.9***

Note: Abbreviations: PTFD = Participant total fixation duration, FFD = first fixation duration, AvgRefixDur = average refixation duration, β'= standardized beta coefficient in the full model, R^2 = coefficient of determination, Adj. = adjusted, est. SE = estimated standard error. Δ = change.

* $p < .05$.

** $p < .01$.

*** $p < .001$.

4. Discussion

We explored how first-pass eye-movement measures during text reading change over time and as a function of reading fluency development. We also examined how these changes contribute to gain in participant total fixation duration (PTFD) used as a proxy of reading fluency development. Expectedly, first fixation duration (FFD), number of first-pass fixations (NrFix), and average refixation duration (AvgRefixDur) all reduced over time (Huestegge et al., 2009). However, only a reduction in fixation duration, not in NrFix, was associated with the reading fluency gain – a finding previously shown in an intervention study (Judica et al., 2002). Although minor changes in the size of word frequency and length effects were observed, these changes were not related to the magnitude of reading fluency development over time (Zoccolotti et al., 2009; Martens & de Jong, 2008).

Regression analysis then sheds light on the complex interdependence of the changes in the number of fixations and their duration. Expectedly, FFD and AvgRefixDur reduced largely hand-in-hand (Hautala et al., 2021; Huestegge et al., 2009), and this reduction explained most of the variance in the development of reading fluency (Zoccolotti et al., 2009). Thus, after the initial reading acquisition phase, reading fluency development seems to first manifest as a reduced mean fixation duration. An unexpected and novel finding was that a reduction in fixation durations inhibited a concurrent reduction in NrFix. However, after controlling for the reduction in fixation duration, students who could maintain or even reduce their NrFix developed their PTFD most i.e. reading fluency.

4.1. Developmental mechanisms

Our most novel finding was that the reduction in fixation durations inhibited concurrent reduction in NrFix. Although the oculomotor mechanism underlying this relationship requires further investigation, we may discuss this effect from a more general viewpoint of skill development. Given that fixation durations were reduced for to all types of words, we interpret that the development occurs predominantly by becoming more efficient in decoding words systematically with multiple fixations. In contrast, those who tried to read faster only by making fewer fixations did not develop well. One possibility is that they have started to employ a “hasty” reading style (Radach et al., 2008; Schoot et al., 2000) by trying to learn to guess the words based on a single or fewer fixations. Such a strategy may not be entirely detrimental, as it has been identified as an effective compensatory reading style for a subgroup of dyslexic readers (Leinonen et al., 2001). The issue requires further research. Finally, a few students could combine the reduction of both the number and duration of their first-pass fixations - thereby achieving the largest gains in reading fluency. We may speculate that these readers reached a threshold in the efficiency of orthographic processing, which allowed them to process more letters per fixation (Hautala et al., 2011).

Although the dual-stage view (Hautala et al., 2021) posits that first- and re-fixation durations reflect predominantly orthographic coding and decoding processes, respectively, these durations are highly intercorrelated and, according to the present results, they also develop hand in hand. Thus, it is reasonable to assume that these measures reflect some unified cognitive process. We suggest this process being an expertize-driven ability to extract, maintain, and process a unified representation (an orthographic template suggested by (Paizi et al., 2013) of a letter string in working memory (Varga et al., 2020). Furthermore, the development of this expertize-driven working memory function may also include a transsaccadic integration mechanism (Higgins and Rayner, 2015), which, in turn, may alleviate the need for memory rehearsal, consequently allowing the reader to start GPC earlier during the refixation and support attention guidance during GPC (Richlan, 2014). Finally, it has been shown that grapheme-phoneme associations also take years to automatize (Blomert, 2011), which may facilitate orthographic coding and decoding.

Interestingly, reading fluency gain was not associated with a reduction in the proportional size of the frequency and length effects relative to the overall mean of the dependent variables. These effects are known to reduce particularly during the first grade in transparent orthographies (Zoccolotti et al., 2009; De Luca et al., 2010). Thus, our results suggest that in later grades, these linguistic effects are reduced more gradually, perhaps as a direct consequence of global development (Zoccolotti et al., 2009), which in the present study manifested in FFD, AvgRefixDur, and NrFix.

Although not subject to statistical testing, the reduction over time was somewhat more pronounced in refixation duration than in

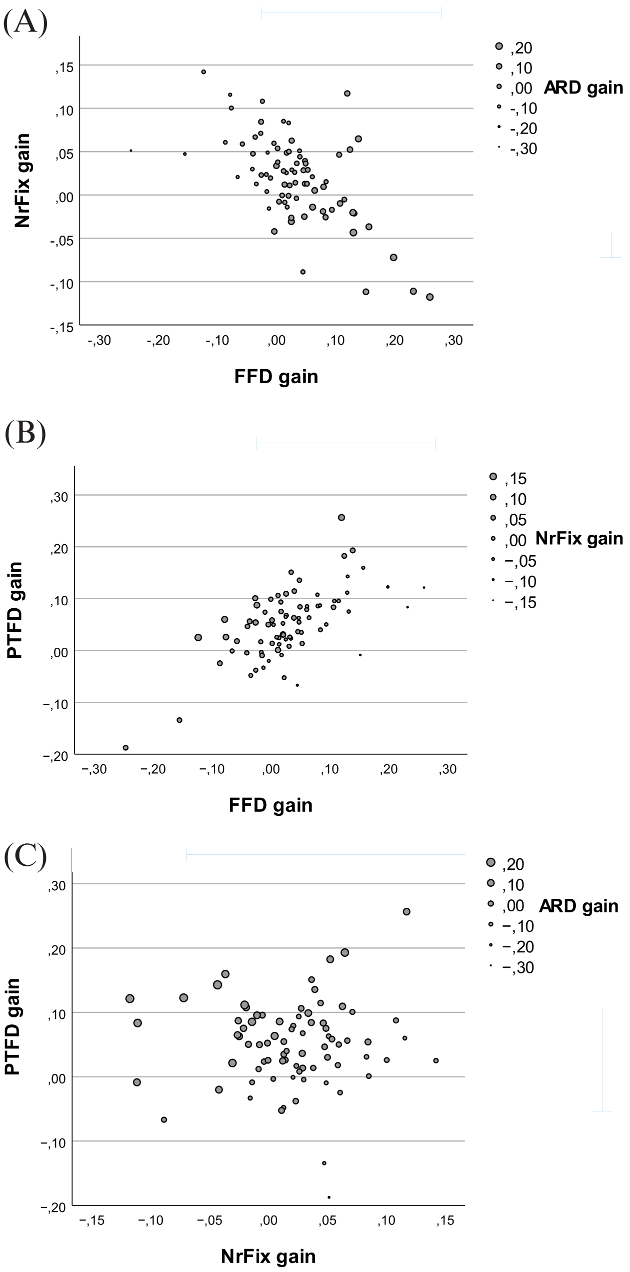

Fig. 4. Selected scatterplots between gains in first fixation duration, number of first-pass fixations, average refixation duration, and participant total fixation duration. The values are shown in a 10-base logarithmic scale.

other measures. This observation may suggest specific development in the decoding process, presumably facilitated by the earlier orthographic activations and linguistic predictions of word-ends (see Hautala et al., 2021, for a discussion). In addition, first fixation durations decreased more for higher-frequency words than for lower-frequency words, and refixation durations for short rare words reduced the most. These effects may indicate the strengthening of orthographic representations for familiar words and, on the other hand, the formation of some new orthographic representations for rare short words (Hautala et al., 2011; Martens & de Jong, 2008).

4.2. Implications for word recognition theories

We framed our study according to the dual-stage view of word recognition (Hautala et al., 2021), assuming that when the orthographic coding of letter identities and their order is reasonably precise, and sublexical and lexical units are activated (Álvarez-Cañizo et al., 2018; Perry et al., 2019; Share, 2008), the subsequent serial GPC decoding will also be facilitated. Comparing the coefficients for the word frequency and length interaction between models with different dependent measures, the absolute numerical value of the coefficients appears to increase from FFD to NrFix and AvgRefixDur. This pattern would be compatible with a dual-stage process hypothesis. However, there is no way to test this difference statistically; besides, coefficients from different models cannot usually be compared straightforwardly. A remaining theoretical question is whether the orthographic coding also includes information about the predictive transitional probabilities between letters and sublexical parts or whether coding these transitions is part of the decoding process (Sibley and Kello, 2012)

The present results may well be in the scope of recently published developmental extensions of the dual-route models (Perry et al., 2010; Pritchard et al., 2018). The generic reduction in word processing times could be understood as the continued strengthening of connection weights over prolonged training of neural networks, while the reduction of the word frequency effect results from the larger learning effect for new words (Dufau et al., 2010). Because the CDP+ + model learns grapheme-phoneme associations rapidly and early during the model training procedure (Ziegler et al., 2019), the reduction in word length effect may be only modest later on. The more intricate findings of the present study, that is, the reduction of FFDs on high-frequency words and the reduction of AvgRefixDurs on short low-frequency words, may prove useful for future simulation studies. It may be that continued exposure to familiar words improves merely their letter encoding while learning new words first manifests in their improved grapheme-phoneme decoding.

4.3. *Limitations*

As discussed, the present results suggest that at the age of nine and ten, word recognition during reading in transparent orthographies develops predominantly via a reduction in fixation duration. However, it is important to remember that changes in linguistic processing (frequency and length effects) would be expected in a similar analysis conducted for longitudinal data spanning over the years (Zoccolotti et al., 2009) and especially for summative measures such as GD (see, e.g., (Tiffin-Richards and Schroeder, 2015).

It should be noted that the theoretical framework of the dual-stage view of word recognition and the dual-deficit view of developmental dyslexia is relatively novel (Hautala et al., 2021) and still requires replications in different languages and with different methodologies. For example, two German studies replicated the emergence of the length effect in late measures but did not find a frequency effect in FFD (Gerth & Festman, 2021; Huestegge et al., 2009). Both studies used a categorical word frequency variable and restricted the analysis to target words, which may have weakened the effects relative to Hautala et al. (2021), who studied fully transparent Finnish orthography with the corpus approach.

Furthermore, reading fluency and its development were operationalized as the participant's total fixation duration and its gain, respectively. This decision was made due to poor correlations between gains in different reading tasks. Therefore, the present findings are specific to the development of word recognition during reading of longer texts and may not readily generalize to other forms of reading fluency development, such as oral reading or reading individual sentences for comprehension.

The texts shown in T2 were sequels to the ones shown in T1. Some students may have recalled the story narrative and characters, which may have facilitated their reading at T2. This memory effect may have contributed to the finding that refixation durations for rare short words were reduced more than for other types of words from T1 to T2. Future longitudinal studies may attempt to construct highly controlled parallel texts and counterbalance the order of presentation.

Finally, the COVID-19 pandemic prevented us from conducting T2 measurements at schools, which led to a substantial drop-out of participants. Although the drop-out was not selective, at least regarding reading fluency or gender, the reduced statistical power prevented us from building cross-validated data-driven models.

4.4. *Conclusions*

The present results indicate that nine-to-ten-year-old students are still developing their efficiency in orthographic processing, that is, expertize for print (Varga et al., 2020). It was found that a reduction in fixation duration drives the development of reading fluency, which we suggest to reflect increased efficiency in extracting, maintaining, and decoding letter string representations in working memory (Paizi et al., 2013). Eventually, this development may transfer into decoding more letters per fixation (Hautala et al., 2011). We suggest that both continued exposure to familiar words and learning new orthographic word representations (Álvarez-Cañizo et al., 2018; Perry et al., 2019; Share, 2008) contribute to developing this generic yet expertize-related working memory function (Zoccolotti et al., 2009; Paizi et al., 2013). This generic development may occur simply when students read frequently (Sonnenschein et al., 2010).

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work, the authors used the Grammarly Premium -tool (not the generative GrammarlyGO) to improve the language of the manuscript. After using this tool, the authors reviewed and edited the content as needed and took full responsibility for the content of the publication.

Declaration of Competing Interest

None.

Data availability

Data will be made available on request.

Acknowledgments

The authors also want to thank all research assistants, students and research colleagues involved in the ReadDrama and ReadMore -research projects, and the reviewers for their valuable comments.

Funding

This research was funded by the Research Council of Finland with grants 317030 and 319911 to Jarkko Hautala. The authors report no conflict of interest. The study was approved by the Ethical Committee of the University of Jyväskylä, and the research was conducted according to the ethical principles for medical research involving human subjects set forth by the Declaration of Helsinki. The data that support the findings of this study are available from the corresponding author JH, upon reasonable request and separate written agreement.

Roles

JH: All 14 CRediT roles **SH**: Conceptualization, Writing - Review & Editing **MR**: Investigation, Project administration, Funding acquisition, Writing - Review & Editing.

Appendix A

A1. Analysis of single fixation duration

There may be trade-off effects between the number of fixations and fixation durations (i.e., shortened first fixation duration when anticipating that a refixation is needed to read a long word). Hautala et al. (Hautala et al., 2021) observed no such effect in the analysis of T1 data. However, when word recognition is highly successful and achieved by a single first-pass fixation, the progression of word recognition processes is likely advanced. Because less fluent readers rarely read long words with single fixations, the analysis of single fixation duration was restricted to words with eight or fewer letters. The four-level interaction of WF × WL × Gain × Time was omitted to obtain convergence. All the main effects were significant (* $p < .05$, ** $p < .01$ *** $p < .001$): WF, $b' = -0.06$, SE $= 0.01$, $t = -4.81$ * ** , Time, $b' = -0.05$, SE $= 0.02$, $t = -3.07$ * *, WL, $b' = 0.06$, SE $= 0.01$, $t = 4.26$ * ** , Gain, $b' = 0.18$, $SE = 0.04$, $t = 4.32$* **, MinSyl, $b' = 0.03$, SE $= 0.008$, $t = 3.27$ * ** , as well as the two-level interaction of Gain × Time, $b' = -0.11$, SE $= 0.02$, $t = -5.17$ * ** .

There was a weak but reliable WL effect suggesting that GPC had begun. The lack of WF × WL interaction may indicate that decoding of low-frequency words was highly successful.

A2. Correlations between logarithmically transformed initial values at T1 and their gains to T2

LG10	PTFD	PTFD_Gain	FFD_T1	FFD_Gain	Nr_T1	Nr_Gain	AvgRefixDur_T1
PTFD_Gain	0.334 * *						
FFD_T1	0.892 * *	0.261 *					
FFD_Gain	0187	0.648 * *	0.343 * *				
Nr_T1	0.802 * *	0.194	0.559 * *	-0.135			
Nr_Gain	0.173	0.036	-0.034	-0.598 * *	0.488 * *		
AvgRefixDur_T1	0.911 * *	0.257 *	0.966 * *	0.281 *	0.585 * *	-0.006	
AvgRefixDur_Gain	0.236 *	0.647 * *	0.319 * *	0.809 * *	-0.063	-0.504 * *	0.364 * *

Note: Abbreviations: PTFD = Participant total fixation duration, FFD = First fixation duration, Nr = Number of first-pass fixations, AvgRefixDur = Average refixation duration.

Appendix B. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.cogdev.2023.101395.

References

Altani, A., Protopapas, A., Katopodi, K., & Georgiou, G. K. (2020). From individual word recognition to word list and text reading fluency. *Journal of Educational Psychology, 112*(1), 22. https://doi.org/10.1037/edu0000359

Álvarez-Cañizo, M., Suárez-Coalla, P., & Cuetos, F. (2018). The role of sublexical variables in reading fluency development among Spanish children. *Journal of Child Language, 45*(4), 858–877. https://doi.org/10.1017/S0305000917000514

Bates, D., Maechler, M., Bolker, B., Walker, S,(2019). lme4: Linear mixed effects models using Eigen and S4. R package v. 1.1–21.

Bijeljac-Babic, R., Millogo, V., Farioli, F., & Grainger, J. (2004). A developmental investigation of word length effects in reading using a new on-line word identification paradigm. *Reading and Writing, 17*(4), 411–431. https://doi.org/10.1023/B:READ.0000032664.20755.af

Blomert, L. (2011). The neural signature of orthographic-phonological binding in successful and failing reading development. *Neuroimage, 57*(3), 695–703. https://doi.org/10.1016/j.neuroimage.2010.11.003

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

Blythe, H. I., & Joseph, H. S. (2011). Children's eye movements during reading. In S. P. Liversedge, I. Gilchrist, & S. Everling (Eds.), *Oxford handbook of eye movements*. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199539789.013.0036.

Blythe, H. I., Liversedge, S. P., Joseph, H. S., White, S. J., & Rayner, K. (2009). Visual information capture during fixations in reading for children and adults. *Vision Research, 49*(12), 1583–1591. https://doi.org/10.1016/j.visres.2009.03.015

Bouma, H., & Legein, C. P. (1980). Dyslexia: A specific recoding deficit? An analysis of response latencies for letters and words in dyslectics and in average readers. *Neuropsychologia, 18*(3), 285–298. https://doi.org/10.1016/0028-3932(80)90124-4

Buswell, G. T. (1922). *Fundamental reading habits: A study of their development*. University of Chicago (No. 21).

Calvo, M. G., & Meseguer, E. (2002). Eye movements and processing stages in reading: Relative contribution of visual, lexical, and contextual factors. *The Spanish Journal of Psychology, 5*(1), 66–77. https://doi.org/10.1017/S1138741600005849

Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. (2001). DRC: a dual route cascaded model of visual word recognition and reading aloud. *Psychological Review, 108*(1), 204. https://doi.org/10.1037/0033-295X.108.1.204

De Luca, M., Zeri, F., Spinelli, D., & Zoccolotti, P. (2010). The acquisition of reading fluency in an orthographically transparent language (Italian): An eye movement longitudinal study. *Medical Science Monitor: International Medical Journal of Experimental and Clinical Research, 16*(3), SC1–7. ⟨https://europepmc.org/article/med/20190700⟩.

Dufau, S., Lété, B., Touzet, C., Glotin, H., Ziegler, J. C., & Grainger, J. (2010). A developmental perspective on visual word recognition: New evidence and a self-organising model. *European Journal of Cognitive Psychology, 22*(5), 669–694. https://doi.org/10.1080/09541440903031230

Eklund, K., Torppa, M., Aro, M., Leppänen, P. H., & Lyytinen, H. (2015). Literacy skill development of children with familial risk for dyslexia through grades 2, 3, and 8. *Journal of Educational Psychology, 107*(1), 126. https://doi.org/10.1037/a0037121

Eklund, K., Salmi, P., Polet, J., & Aro, M., (2013). Tuen tarpeesta tunnistamiseen. Lukemisen ja kirjoittamisen arviointi. Toinen luokka. Tekninen opas [A screening tool of reading and spelling for Grade 2. Technical manual]. Jyväskylä, Finland: Niilo Mäki Institute.

Fox, J., Weisberg, S., Adler, D., Bates, D., Baud-Bovy, G., Ellison, S., & Monette, G. (2012). Package 'car'. *Vienna: R Foundation for Statistical Computing, 16*.

Gerth, S., & Festman, J. (2021). Reading development, word length and frequency effects: An eye-tracking study with slow and fast readers. *Frontiers in Communication, 202*. https://doi.org/10.3389/fcomm.2021.743113

Gutezeit, G. (1976). Tachistoscopic reading studies with dyslexic children. *Zeitschrift für Klinische Psychologie, 5*(1), 31–52.

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*(2), 167–181. https://doi.org/10.1016/j.jecp.2008.04.002

Hautala, J., Hawelka, S., & Aro, M. (2021). Dual-stage and dual-deficit? Word recognition processes during text reading across the reading fluency continuum. *Reading and Writing, 35*(3), 663–686. https://doi.org/10.1007/s11145-021-10201-1

Hautala, J., Heikkilä, R., Nieminen, L., Rantanen, V., Latvala, J. M., & Richardson, U. (2020). Identification of reading difficulties by a digital game-based assessment technology. *Journal of Educational Computing Research, 58*(5), 1003–1028. https://doi.org/10.1177/0735633120905309

Hautala, J., Hyönä, J., Aro, M., & Lyytinen, H. (2011). Sublexical effects on eye movements during repeated reading of words and pseudowords in Finnish. *Psychology of Language and Communication, 15*(2), 129–149. https://doi.org/10.2478/v10057-011-0009-x

Hautala, J., Ronimus, M., & Junttila, E. (2023). Readers' theater projects for special education: A randomized controlled study. *Scandinavian Journal of Educational Research, 67*(5), 663–678. https://doi.org/10.1080/00313831.2022.2042846

Hawelka, S., Schuster, S., Gagl, B., & Hutzler, F. (2013). Beyond single syllables: The effect of first syllable frequency and orthographic similarity on eye movements during silent reading. *Language and Cognitive Processes, 28*(8), 1134–1153. https://doi.org/10.1080/01690965.2012.696665

Higgins, E., & Rayner, K. (2015). Transsaccadic processing: Stability, integration, and the potential role of remapping. *Attention, Perception, & Psychophysics, 77*, 3–27. https://doi.org/10.3758/s13414-014-0751-y

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

Häyrinen, T., Serenius-Sirve, S., & Korkman, M., (2013). Lukilasse – Lukemisen, kirjoittamisen ja laskemisen seulontatesti 1.–6. vuosiluokille [Lukilasse – Screening test for reading, writing and arithmetics for grades 1 to 6]. Hogrefe.

Huovilainen, T.M. , (2018). Psycholinguistic descriptives. The Language Bank of Finland. Retrieved from ⟨http://urn.fi/urn:nbn:fi:lb-2018081601⟩.

Hutzler, F., & Wimmer, H. (2004). Eye movements of dyslexic children when reading in a regular orthography. *Brain and Language, 89*(1), 235–242. https://doi.org/10.1016/S0093-934X(03)00401-2

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. https://doi.org/10.1037/0278-7393.21.6.1430

Jobard, G., Crivello, F., & Tzourio-Mazoyer, N. (2003). Evaluation of the dual route theory of reading: A metanalysis of 35 neuroimaging studies. *Neuroimage, 20*(2), 693–712. https://doi.org/10.1016/S1053-8119(03)00343-4

Jobard, G., Vigneau, M., Simon, G., & Tzourio-Mazoyer, N. (2011). The weight of skill: Interindividual variability of reading related brain activation patterns in fluent readers. *Journal of Neurolinguistics, 24*(1), 113–132. https://doi.org/10.1016/j.jneuroling.2010.09.002

Joseph, H. S., Liversedge, S. P., Blythe, H. I., White, S. J., & Rayner, K. (2009). Word length and landing position effects during reading in children and adults. *Vision Research, 49*(16), 2078–2086. https://doi.org/10.1016/j.visres.2009.05.015

Judica, A., De Luca, M., Spinelli, D., & Zoccolotti, P. (2002). Training of developmental surface dyslexia improves reading performance and shortens eye fixation duration in reading. *Neuropsychological Rehabilitation, 12*(3), 177–197. https://doi.org/10.1080/09602010244000002

Khelifi, R., Sparrow, L., & Casalis, S. (2019). Is a frequency effect observed in eye movements during text reading? A comparison between developing and expert readers. *Scientific Studies of Reading, 23(4), 334–347.* https://doi.org/10.1080/10888438.2019.1571064

Kim, Y. S. G., Little, C., Petscher, Y., & Vorstius, C. (2022). Developmental trajectories of eye movements in oral and silent reading for beginning readers: A longitudinal investigation. *Scientific Reports, 12*(1), 18708. https://doi.org/10.1038/s41598-022-23420-5

Kliegl, R. , (2023, January 1). Re *:* Using model-based estimates as predictors in other models [Discussion post]. Research Gate. ⟨https://www.researchgate.net/post/Using_model-based_estimates_as_predictors_in_other_models⟩.

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

Kyte, C. S., & Johnson, C. J. (2006). The role of phonological recoding in orthographic learning. *Journal of Experimental Child Psychology, 93*(2), 166–185. https://doi.org/10.1016/j.jecp.2005.09.003

Leinonen, S., Müller, K., Leppänen, P. H., Aro, M., Ahonen, T., & Lyytinen, H. (2001). Heterogeneity in adult dyslexic readers: Relating processing skills to the speed and accuracy of oral text reading. *Reading and Writing, 14*(3–4), 265–296. https://doi.org/10.1023/A:1011117620895

Lerkkanen, M.K., Niemi, P., Poikkeus, A.M., Poskiparta, M., Siekkinen, M., Nurmi, J.E. , (2006). The first steps study. Unpublished data, Department of Psychology and Department of Teacher Education, University of Jyväskylä, Jyväskylä, Finland.

Loberg, O., Hautala, J., Hämäläinen, J. A., & Leppänen, P. H. (2019). Influence of reading skill and word length on fixation-related brain activity in school-aged children during natural reading. *Vision Research, 165*, 109–122. https://doi.org/10.1016/j.visres.2019.07.008

Martens, V. E. G., & de Jong, P. F. (2008). Effects of repeated reading on the length effect in word and pseudoword reading. *Journal of Research in Reading, 31*(1), 40–54. https://doi.org/10.1111/j.1467-9817.2007.00360.x

Marx, C., Hutzler, F., Schuster, S., & Hawelka, S. (2016). On the development of parafoveal preprocessing: Evidence from the incremental boundary paradigm. *Frontiers in Psychology, 7*, 514. https://doi.org/10.3389/fpsyg.2016.00514

McConkie, G. W., Zola, D., Grimes, J., Kerr, P. W., Bryant, N. R., & Wol, P. M. (1991). Children's eye movements during reading. In J. F. Stein (Ed.), *Vision and visual dyslexia* (pp. 251–262). CRC Press.

O'Brien, R. M. (2007). A caution regarding rules of thumb for variance inflation factors. *Quality & Quantity, 41*, 673–690. https://doi.org/10.1007/s11135-006-9018-6

National Library of Finland, 2014. *The Finnish N-grams 1820-2000 of the Newspaper and Periodical Corpus of the National Library of Finland* [data set]. Kielipankki 2014. http://urn.fi/urn:nbn:fi:lb-2014073038.

O'Brien, B. A., Wolf, M., Miller, L. T., Lovett, M. W., & Morris, R. (2011). Orthographic processing efficiency in developmental dyslexia: An investigation of age and treatment factors at the sublexical level. *Annals of Dyslexia, 61*(1), 111–135. https://doi.org/10.1007/s11881-010-0050-9

Paizi, D., De Luca, M., Zoccolotti, P., & Burani, C. (2013). A comprehensive evaluation of lexical reading in Italian developmental dyslexics. *Journal of Research in Reading, 36*(3), 303–329. https://doi.org/10.1111/j.1467-9817.2011.01504.x

Perry, C., Ziegler, J. C., & Zorzi, M. (2010). Beyond single syllables: Large-scale modeling of reading aloud with the connectionist dual process (CDP++) model. *Cognitive Psychology, 61*(2), 106–151. https://doi.org/10.1016/j.cogpsych.2010.04.001

Perry, C., Zorzi, M., & Ziegler, J. C. (2019). Understanding dyslexia through personalized large-scale computational models. *Psychological Science, 30*(3), 386–395. https://doi.org/10.1177/0956797618823540

Pritchard, S. C., Coltheart, M., Marinus, E., & Castles, A. (2018). A computational model of the self-teaching hypothesis based on the dual-route cascaded model of reading. *Cognitive Science, 42*(3), 722–770. https://doi.org/10.1111/cogs.12571

Radach, R., Huestegge, L., & Reilly, R. (2008). The role of global top-down factors in local eye-movement control in reading. *Psychological Research, 72*, 675–688. https://doi.org/10.1007/s00426-008-0173-3

Rau, A. K., Moeller, K., & Landerl, K. (2014). The transition from sublexical to lexical processing in a consistent orthography: An eye-tracking study. *Scientific Studies of Reading, 18*(3), 224–233. https://doi.org/10.1080/10888438.2013.857673

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

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological bulletin, 124*(3), 372. https://doi.org/10.1037/0033-2909.124.3.372

Reichle, E. D., Liversedge, S. P., Drieghe, D., Blythe, H. I., Joseph, H. S., White, S. J., & Rayner, K. (2013). Using EZ Reader to examine the concurrent development of eye-movement control and reading skill. *Developmental Review, 33*(2), 110–149. https://doi.org/10.1016/j.dr.2013.03.001

Richlan, F. (2014). Functional neuroanatomy of developmental dyslexia: The role of orthographic depth. *Frontiers in Human Neuroscience, 8*, 347. https://doi.org/10.3389/fnhum.2014.00347

Saksida, A., Iannuzzi, S., Bogliotti, C., Chaix, Y., Démonet, J. F., Bricout, L., & George, F. (2016). Phonological skills, visual attention span, and visual stress in developmental dyslexia. *Developmental Psychology, 52*(10), 1503. https://doi.org/10.1037/dev0000184

Samuels, S. J., LaBerge, D., & Bremer, C. D. (1978). Units of word recognition: Evidence for developmental changes. *Journal of Verbal Learning and Verbal Behavior, 17*(6), 715–720. https://doi.org/10.1016/S0022-5371(78)90433-4

Schmidtke, D., & Moro, A. L. (2021). Determinants of word-reading development in English learner university students: A longitudinal eye movement study. *Reading Research Quarterly, 56*(4), 819–854. https://doi.org/10.1002/rrq.362

Schoot, M., Licht, R., Horsley, T. M., & de Sergeant, J. A. (2000). Inhibitory deficits in reading disability depend on subtype: Guessers but not spellers. *Child Neuropsychology, 6*(4), 297–312. https://doi.org/10.1076/chin.6.4.297.3139

Schroeder, S., Häikiö, T., Pagán, A., Dickins, J. H., Hyönä, J., & Liversedge, S. P. (2021). Eye movements of children and adults reading in three different orthographies. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 48*(10), 1518–1541. https://doi.org/10.1037/xlm0001099

Share, D. L. (2008). Orthographic learning, phonological recoding, and self-teaching. *Advances in Child Development and Behavior, 36*, 31–82. https://doi.org/10.1016/S0065-2407(08)00002-5

Sibley, D. E., & Kello, C. T. (2012). Learned orthographic representations facilitates large-scale modeling of word recognition. In J. Adelman (Ed.), *Visual word recognition* (Vol. 1, pp. 28–51). Psychology Press.

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2012). A 21 word solution. *Dialogue, 26*(2), 4–7. https://doi.org/10.2139/ssrn.2160588

Singmann, H., Bolker, B., Westfall, J., Aust, F., Ben-Shachar, M.S., & Højsgaard, S. , (2015). Package 'afex'.

Slattery, T. J., & Parker, A. J. (2019). Return sweeps in reading: Processing implications of undersweep-fixations. *Psychonomic Bulletin & Review, 26*, 1948–1957. https://doi.org/10.3758/s13423-019-01636-3

Sonnenschein, S., Stapleton, L. M., & Benson, A. (2010). The relation between the type and amount of instruction and growth in children's reading competencies. *American Educational Research Journal, 47*(2), 358–389. https://doi.org/10.3102/0002831209349215

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

Spichtig, A., Pascoe, J., Ferrara, J., & Vorstius, C. (2017). A comparison of eye movement measures across reading efficiency quartile groups in elementary, middle, and high school students in the US. *Journal of Eye Movement Research, 10*(4), 5. https://doi.org/10.16910/jemr.10.4.5

Spinelli, D., De Luca, M., Di Filippo, G., Mancini, M., Martelli, M., & Zoccolotti, P. (2005). Length effect in word naming in reading: Role of reading experience and reading deficit in Italian readers. *Developmental Neuropsychology, 27*(2), 217–235. https://doi.org/10.1207/s15326942dn2702_2

Taylor, S. E. (1965). Eye movements while reading: Facts and fallacies. *American Educational Research Journal, 2*(4), 187–202. https://doi.org/10.3102/00028312002004187

Thompson, F. T., & Levine, D. U. (1997). Examples of easily explainable suppressor variables in multiple regression research. *Multiple Linear Regression Viewpoints, 24*(1), 11–13. ⟨http://home.ubalt.edu/tmitch/645/articles/Thompson%20&%20Levine%20Ex%20supressor%20vars.pdf⟩.

Tiffin-Richards, S. P., & Schroeder, S. (2015). Word length and frequency effects on children's eye movements during silent reading. *Vision Research, 113*, 33–43. https://doi.org/10.1016/j.visres.2015.05.008

Varga, V., Tóth, D., & Csépe, V. (2020). Orthographic-phonological mapping and the emergence of visual expertise for print: A developmental event-related potential study. *Child Development, 91*(1), e1–e13. https://doi.org/10.1111/cdev.13159

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

Vorstius, C., Radach, R., & Lonigan, C. J. (2014). Eye movements in developing readers: A comparison of silent and oral sentence reading. *Visual Cognition, 22*(3–4), 458–485. https://doi.org/10.1080/13506285.2014.881445

Yap, M. J., Balota, D. A., Sibley, D. E., & Ratcliff, R. (2012). Individual differences in visual word recognition: Insights from the English Lexicon Project. *Journal of Experimental Psychology: Human Perception and Performance, 38*(1), 53. https://doi.org/10.1037/a0024177

Ziegler, J. C., Perry, C., & Zorzi, M. (2019). Modeling the variability of developmental dyslexia. In L. Verhoeven, C. Perfetti, & K. Pugh (Eds.), *Developmental dyslexia across languages and writing systems* (pp. 350–371). Cambridge University Press.

Ziegler, J. C., Perry, C., & Zorzi, M. (2020). Learning to read and dyslexia: From theory to intervention through personalized computational models. *Current Directions in Psychological Science, 29*(3), 293–300. https://doi.org/10.1177/0963721420915873

Zoccolotti, P., De Luca, M., Di Filippo, G., Judica, A., & Martelli, M. (2009). Reading development in an orthographically regular language: Effects of length, frequency, lexicality and global processing ability. *Reading and Writing, 22*(9), 1053–1079. https://doi.org/10.1007/s11145-008-9144-8