CHAPTER FIVE: Reliability and Individual Differences in the Eye-voice Span and its Modulation

5.1. Abstract

Individual differences in children's reading have been studied using offline behavioural methods and, in recent years, online eye movement measures. However, less is known about individual differences in online eye-voice span (EVS) measures which can potentially provide more information about underlying reading processes in children who primarily read aloud. EVS data from 52 seven to ten-year-olds reading short passages were obtained. In addition, offline ability measures of reading, spelling, vocabulary and rapid automatized naming (RAN) speed were assessed. The reliability of the EVS was evaluated and estimated to be 0.77. Further individual difference analysis revealed that reading, vocabulary, and RAN predicted mean EVS, while spelling ability predicted variability in the EVS. In the eye movement data, spelling ability influenced saccade length and word reading ability influenced gaze duration. The result suggests that spelling ability influences early letter encoding during reading, and fixation times are modulated by reading ability confirming prior work in adults. More importantly, we conclude that the EVS may reflect reading-related offline ability measures more than gaze duration.

5.2. Introduction

Children's ability to read (for comprehension) is dependent on code-related skills such as phonological awareness, and decoding (Clarke et al., 2005; Georgiou et al., 2013; Landerl et al., 2021) and broader language skills such as vocabulary and listening comprehension (Gough & Tunmer, 1986; Nation & Snowling, 2004; Protopapas et al., 2013). These skills are different from the actual reading process and are usually categorized as reading-related skills. Children differ throughout their development by mastering some of these skills and continuing to develop in others as they grow (Paris, 2005). Similarly, inter-individual differences in reading are reflected in inter-individual differences in reading-related skills. More evidence exists on the relationship between reading and reading-related skills using offline measures such as comprehension scores or reading speed than on online measures such as eye-tracking (Kim et al., 2020; Long & Freed, 2020). While most eye-tracking studies focus on skilled adult readers, technological advancements have led to a spurt in developmental studies (Foster et al., 2018; Rayner, 1998). By deploying multiple measures such as fixation and saccadic measures, eye-tracking research can unveil moment-by-moment processes during reading and provide a greater understanding of reading development (Blythe, 2014; Miller & O'Donnell, 2013; Schroeder et al., 2015).

Another measure derived from eye-tracking methods when speech output is simultaneously recorded is the *eye-voice span* (EVS). The EVS is the distance between the eye and the voice during oral reading. This measure has received relatively less attention in recent times compared to eye movement measures. Sparse research on the EVS may be due to the methodological challenges of manually segmenting speech files to obtain precise temporal information for aligning eye and voice locations. However, for developing readers who primarily read aloud (Marx et al., 2016; Rasinski & Hoffman, 2003; Rayner et al., 2016), EVS research may have significant potential in tracking individual differences in reading and reading-related skills (Buswell, 1920; Easson et al., 2020; Levin & Turner, 1966). However, before such conclusions can be drawn, it is necessary to undertake a systematic investigation of the EVS. To achieve this, we had three goals in the current study. First, to explore the reliability of the EVS alongside other eye movement measures during oral reading. Second, to investigate how individual differences in offline measures of reading-related skills relate to a reader's EVS. Third, to examine individual differences in the way readers modulate the EVS. We begin by describing reliability and individual differences in eye movements and summarise what is known about these in relation to the EVS.

5.2.1. Reliability of eye movements in children's reading

Individual differences in correlational research require reliable measures. Reliability is a measure's capacity to estimate differences between individuals consistently (Hedge et al., 2018; Staub, 2021). Poor reliability signals that such differences reflect noise rather than stable individual differences (Staub, 2021). Additionally, the relationship deduced from two variables is limited without sufficient and consistent variability (Hedge et al., 2018). Therefore, estimating the relationship between two measures is futile if either measure is unreliable (Spearman, 1910). Fortunately, most eye movement measures during skilled reading are highly reliable (Carter & Luke, 2018; Kuperman & Van Dyke, 2011; Staub, 2021). Split-half reliability estimates for first fixation duration, gaze duration, regression and skipping range from 0.75 to 0.90 in adults (Staub, 2021). In developing readers (6 to 8 year olds), Foster et al. (2018) reported alternate form reliability ranging from 0.37 to 0.89 and test re-test reliability ranging from 0.15 to 0.82 for various eye movement measures. Interestingly, while the adult data showed the highest reliability for word skipping (0.90; Staub, 2021), the children's data showed the lowest for skipping (0.37 for alternate form and 0.15 for test re-test; Foster et al., 2018). Higher skipping probabilities in skilled readers are associated with the capacity to use information in the right parafoveal region, 5 degrees to the right of the fixation point (Drieghe et al., 2004; Schotter et al., 2012). However, children focus more on processing information in the foveal region, 2 degrees around the fixation point. Hence, a lack of reliability in skipping could be that children do not show consistent skipping patterns (Blythe et al., 2011; Häikiö et al., 2009; Lee et al., 2021). Poor reliability in children's word skipping behaviour may also reflect their engagement levels, which may inconsistently influence reading performance (Foster et al., 2018). Inconsistent levels of engagement results in considerable intra-individual variability in children's eye movements (Joseph et al., 2008; Joseph et al., 2013; Yang et al., 2002). However, despite such within-subject variability, it appears children's eye movements are highly reliable, with the exception of word skipping.

Silent reading studies have provided these reliability estimates, and there are hardly any studies examining the reliability of these measures during oral reading. Although oral and silent reading modes share similar cognitive processes, eye movements in oral reading are much more related to reading proficiency in beginning readers (Kim et al., 2019). This difference may be because oral reading ensures that full attentional capacity is given to word recognition processes due to the articulation requirement (Kim et al., 2019). To this end, some researchers use the oral reading paradigm to increase engagement or have direct control during children's reading (Molina et al., 2020; Trauzettel-Klosinski et al., 2010). Therefore, one possibility is that the reliability estimates in oral reading would be higher than reported for silent reading. However, since children's eye movement behaviour during oral and silent reading are more strongly correlated than adults (Anderson & Swanson, 1937), differences in reliability estimates between the two modalities might be non-existent.

5.2.2. Individual differences in children's eye movements

In recent years, there have been several individual difference studies on the eye movements of skilled readers (Andrews & Lo, 2013; Kuperman et al., 2018; Kuperman & Van Dyke, 2011; Long & Freed, 2021; Luke et al., 2018; Parker & Slattery, 2021; Veldre & Andrews, 2014, 2015) and developing readers (Barnes & Kim, 2016; de Leeuw et al., 2016; Foster et al., 2018; Kim et al., 2019, 2020; Lee et al., 2021). These studies have shown that eye movements are valid measures of a wide range of reading-related skills (see Kuperman & Van Dyke, 2011 for a comprehensive list). Due to testing time constraints, we considered only four important reading-related skills which have been widely studied. These are word reading efficiency, spelling ability, vocabulary, and rapid automatised naming (RAN) speed. These measures were the focus of this study for a number of reasons. As will be seen, word reading, spelling and vocabulary are indexes of lexical quality which have been studied in relation to eye movements but less so for the EVS. Word reading and RAN have also been found to be the most reliable predictors of a wide range of eye movement measures in adult readers (Kuperman & Van Dyke, 2011). While RAN has been studied in relation to the EVS where the EVS is measured during RAN tasks, no study has directly investigated how the RAN on an independent task contributes to the EVS measure during sentence reading.

Across children and adult data, there is ample evidence that word reading efficiency is perhaps the most significant predictor of temporal eye movement measures such as gaze durations (de Leeuw et al., 2016; Hessel et al., 2020; Kim et al., 2020; Kuperman & Van Dyke, 2011; Lee et al., 2021, but see Long & Freed, 2021). These findings are consistent and provide support for the E-Z reader model of eye movements, which opine that the rate of lexical (word) processing during reading is the engine that drives the eyes forward (Reichle et al., 2013; Reichle et al., 1998; Reichle et al., 2003). According to the E-Z reader model, two stages are involved in lexical processing, an initial familiarity check stage indicating that word identification is imminent and the second stage of complete lexical access. Completing the first and second stages triggers saccade programming and shifting of attention, respectively, to the next word. Therefore, readers with lower word reading proficiency will likely take more time in the familiarity check stage, thereby delaying saccade initiation to the next word. This would also have a knock-on effect on the time the reader shifts attention to the next word, thereby reducing parafoveal processing of the upcoming word.

The impact of word reading proficiency on children's spatial eye movement measures is less pronounced than on temporal measures (de Leeuw et al., 2016; Foster et al., 2018; Kim et al., 2020; Lee et al., 2021). For example, word reading ability is not associated with skipping probability or regression probability (but see de Leeuw et al., 2016; Foster et al., 2018; Lee et al., 2021) and is only moderately associated with saccade length (Kim et al., 2020). However, other measures such as oral reading fluency and comprehension appear to predict skipping (Lee et al., 2021). Again, children rarely skip words, which may account for such null or inconsistent effects (Foster et al., 2018).

Skipping a word is dependent on predictability from prior context, word length and the capacity to extract phonological or orthographic information acquired in the parafovea (Ashby & Rayner, 2004; Brysbaert & Vitu, 1998; Fitzsimmons & Drieghe, 2011; Parker & Slattery, 2021; Schotter et al., 2012). Therefore, skipping probability is considered an early index of lexical processing (Andrews & Veldre, 2019), as skipped words are assumed to be processed mainly in the parafovea during the fixation of the previous word.

While word reading predicts skipping probability and fixation times in adults (Drieghe et al., 2019; Kuperman et al., 2018), spelling ability has been uniquely associated with skipping probabilities and saccade length in skilled readers (Drieghe et al., 2019; Veldre & Andrews, 2014, 2015). This is because unlike word reading, spelling requires precise orthographic representations of letter and letter order in words (Perfetti, 1992). Therefore, spelling ability is considered the most reliable measure of the quality of a reader's lexicon (Perfetti, 2007; Perfetti, 1992; Veldre & Andrews, 2014). These ideas have been fully explored under the *lexical quality hypothesis* (Perfetti, 1992; Perfetti & Hart, 2002). This hypothesis contrasts the idea of lexical *quantity*, i.e., the number of words in one's vocabulary. Instead, it emphasizes quality, the extent to which each word possesses well-specified orthographic, phonological and semantic representations (Perfetti, 2007). Therefore, skilled readers with high spelling ability can extract precise orthographic representations in the parafovea, impacting their saccade targeting behaviour. However, less is known about whether a similar pattern is obtainable for children as there are few studies examining spelling ability and eye movements for developing readers. One study by Ashby et al. (2013) reported that receptive spelling for low frequency words predicted total sentence reading time during silent reading and oral reading fluency of children in grade two (7 years). This study provides some support for the impact of spelling on children's eye movement behaviour.

The lexical quality hypothesis also includes a flexible semantic constituent (Perfetti, 2007; Perfetti & Hart, 2002) where different word forms could represent the same or different meaning. This constituent can be measured with a reader's vocabulary knowledge. Perfetti and Stafura (2013) consider vocabulary knowledge as the output of word recognition processes and the input for reading comprehension. In line with the latter, a great deal of evidence shows a strong relationship between vocabulary knowledge and reading comprehension (Braze et al., 2016; Protopapas et al., 2013; Tunmer & Chapman, 2012). Although most studies have used a combination of reading comprehension and vocabulary (Ashby et al., 2005; Drieghe et al., 2019; Veldre & Andrews, 2016), and only reading comprehension (Veldre & Andrews, 2015) to assess the semantic constituent, others have used only vocabulary knowledge (Andrews & Lo, 2013, but see Long & Freed, 2021). Using a combination of reading comprehension and vocabulary as measured by the Nelson-Denny Reading Test (Brown, Fishco, & Hanna, 1993), better readers relied less on context for word recognition (Ashby et al., 2005) and benefitted more from semantically related previews (Veldre & Andrews, 2016) compared to less skilled readers. More broadly, shorter gaze and fixation durations were associated with higher reading skills (Ashby et al., 2005; Veldre & Andrews, 2016). These findings are consistent with developmental studies showing vocabulary knowledge as a predictor of temporal eye movement measures (de Leeuw et al., 2016; Hessel et al., 2020; but see: Lee et al., 2021).

Rapid automatized naming speed is strongly associated with eye movements, presumably because its serial presentation format resembles that of sentence reading (Kuperman & Van Dyke, 2011). Additionally, this measure is a significant predictor of word reading skills (Caravolas et al., 2019; Clayton et al., 2019; Snowling & Hulme, 2021). Although Kuperman and Van Dyke (2011) reported independent effects of RAN on skilled readers' eye movements, word identification was not controlled for in the model. In contrast, the impact of RAN on third graders' eye movements was primarily mediated by word reading (Kim et al., 2020). This raises the question of whether RAN speed independently contributes to the *where* and *when* of eye movements, above and beyond word reading or decoding skills.

5.2.3. The eye-voice span: reliability and individual differences

The oral reading paradigm provides access to two streams of measurements: eye movement measures and speech production measures. A combination of these streams through the synchronisation of eye movements and the voice yields the "*eye-voice span*", the spatial or temporal distance between the eye and the voice during reading aloud (Buswell, 1920; Fairbanks, 1937; Inhoff et al., 2011; Laubrock & Kliegl, 2015; Levin & Turner, 1966). This measure reflects an individual's proficiency with written language and coordinative capacities between visual and speech domains. Unlike eye movement measures which make a distinction between early word identification (e.g., first fixation duration) and late semantic integration measures (e.g., total viewing time), the EVS is associated with both lower-level processes such as visual processing, lexical identification, print to sound translation (Gordon & Hoedemaker, 2016; Pan et al., 2013) as well as higher-order processes such as semantic and syntactic integration (Balajthy Jr, 1978; Levin & Turner, 1966; Levin & Wanat, 1967). Additionally, the degree to which these processes have become automatised to free up attentional resources (Nayar et al., 2018) and working memory capacity (Laubrock & Kliegl, 2015) are related to the EVS.

Critically, there is a knowledge gap in the literature about the reliability of the EVS. To our knowledge, only one study has reported an EVS reliability estimate. Gordon and Hoedemaker (2016) reported an estimate of 0.88 during a RAN task among adult readers. For a long time, the RAN task has been associated with reading in serial processing and orthography to phonology mappings (Clarke et al., 2005; Georgiou et al., 2013; Georgiou et al., 2016). However, sentence reading differs from RAN due to the additional semantic and syntactic integration processes. Therefore, the need to make meaning and read with expression during oral reading suggests a potentially more variable EVS during sentence reading.

On average, readers' eye fixations tend to be between 2 and 3 words or 15-character spaces ahead of their voice (Buswell, 1920; Inhoff et al., 2011; Rayner, 1998). Several early researchers used this spatial EVS measure as it does not always require simultaneous eye-voice recording. However, other methods used to derive the spatial EVS estimate, such as lights-off or occlusion paradigms (Levin & Addis, 1979; Levin & Turner, 1966; Quantz, 1897; Resnick, 1970; Stuart-Hamilton & Rabbit, 1997), have been found to overestimate the EVS (Laubrock & Kliegl, 2015; Levin & Addis, 1979). Findings regarding the spatial EVS and how it relates to reading skills are usually clear-cut; better readers have larger spans (Buswell, 1920; De Luca et al., 2013; Halm et al., 2011). This was supported in Chapter 3 of this thesis. Most researchers have compared the EVS between groups, i.e., readers with dyslexia and typical readers, or used reading rate and reading comprehension measures to discriminate between readers. For example, participants classified as good readers based on reading comprehension (Buswell, 1920; Fairbanks, 1937; Holgerson, 1977) or reading rate (Morton, 1964; Quantz, 1897) have a larger EVS compared to poor or slow readers. Other researchers have used the EVS during RAN tasks to distinguish different reading ability levels (Jones et al., 2008; Jones et al., 2016; Pan et al., 2013). For instance, Pan et al. (2013) found larger EVS for typical readers than dyslexic readers during a RAN digit and dice naming task. However, to our knowledge, no research has examined the relationship between the EVS and other code-related linguistic skills such as word reading, spelling ability and language skills such as vocabulary knowledge. In the same vein, a larger EVS during RAN tasks predicts faster RAN completion time for skilled readers and typically developing fifth graders compared to less able peers (Gordon & Hoedemaker, 2016; Pan et al., 2013). However, the relationship between RAN and the EVS during a separate reading-aloud task is yet to be examined.

Researchers have also used a somewhat controversial EVS measure, the temporal EVS. This measure is sometimes called *fixation speech interval* (FSI; Inhoff et al., 2011; Järvilehto et al., 2009). It measures the time between looking and articulating a word and can be measured from the beginning (onset) [6] or the end of gaze (offset; see Figure 5.1). Unlike the spatial EVS, both temporal measures require simultaneous recording of the eye and voice to obtain the precise timings of eye fixations and voice onsets (Geyer, 1967; Inhoff et al., 2011; Järvilehto et al., 2009; Morton, 1964). The temporal EVS is used as an indicator of cognitive effort, and it varies from 500ms to 1000ms in skilled readers (Järvilehto et al., 2009; Morton, 1964; Silva et al., 2016; Zhou et al., 2021). Readers with dyslexia have a longer temporal EVS (1624ms) than typical readers (714ms; Halm et al., 2011). Recently, Zhou et al. (2021) compared the temporal EVS between reading aloud and sight translation tasks, where the latter was deemed more cognitively demanding. In line with their expectations, longer temporal EVS was found at the start and end of sentences during the sight translation task compared to the reading aloud task. Reading experience also accounts for variability in the temporal EVS, where less experienced readers have longer spans than more experienced readers (Järvilehto et al., 2009).

Interestingly, the temporal offset EVS has not always yielded a consistent pattern like the onset measure. For example, Morton (1964) reported that slow undergraduate English readers had longer offset EVSs than fast readers. This result was only significant when reading passages with statistical approximations to the eighth order (e.g., *recognize her abilities in music after he scolded him before*; see Miller & Selfridge, 1950 for more examples on statistical approximations). For normal texts, slow readers had shorter spans than fast readers, but this difference was not significant (Morton, 1964). However, a significant positive relationship between reading rate and temporal offset EVS was observed by Silva et al. (2016), where Portuguese readers with longer offset EVS had higher reading rates. This finding suggests that fast readers are efficiently allocating more time to recognize the next word before articulating the previous word. In contrast, articulation would proceed shortly after the end of gaze for slower readers whose eyes and voice are closely knit.

[6] Unless otherwise noted, temporal EVS is the onset measure

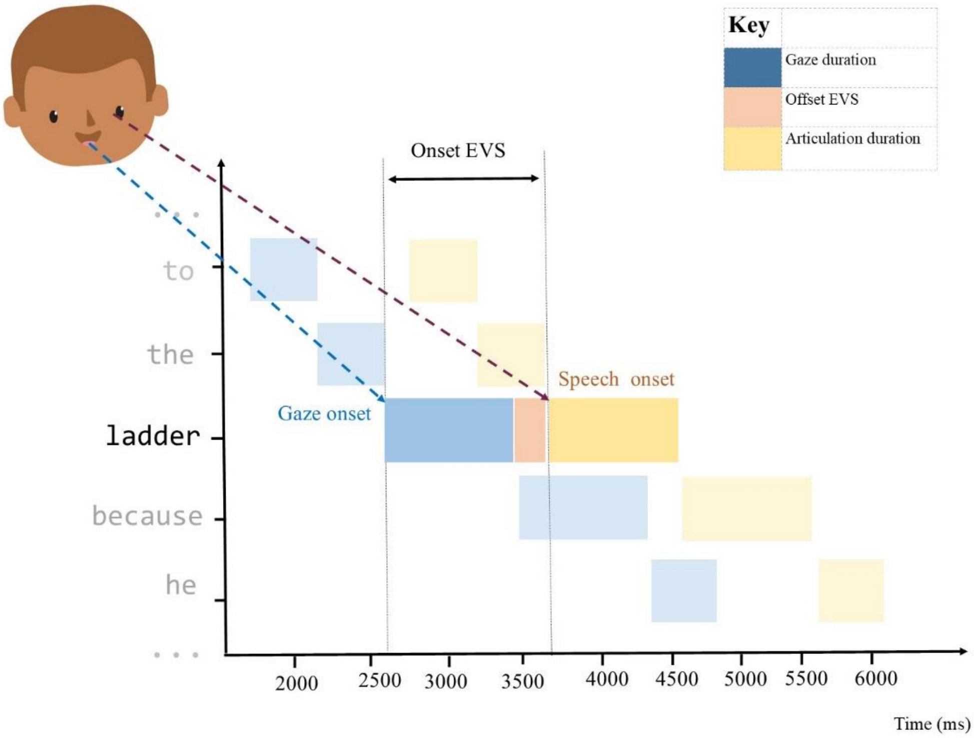

Figure 5.1. Illustration of temporal onset and offset EVS

Clearly, the onset and offset temporal measures seem to operate in opposite directions in relation to reading skills, where longer onset and shorter offset are related to poor and slow reading skills and vice versa. It is important to note that researchers have reported that the temporal onset EVS may be problematic because of its sensitivity to speech errors, pauses and articulation of previous words (Gordon & Hoedemaker, 2016; Järvilehto et al., 2009). Therefore, it may not purely reflect an individual's time to plan speech during reading aloud, like naming latencies. An additional complication with the temporal EVS measure is that it is undefined for words that are not fixated during reading (i.e. skipped words), as the measure for a word begins at the point of fixation and terminates at the point of pronunciation.

Taken together, all three measures of the EVS: spatial EVS, temporal onset and temporal offset EVS, seem to work in relatively different directions; whether they reflect similar cognitive processes (Zhou et al., 2021) is yet unexplored. Therefore, an individual differences approach using simultaneous eye-tracking and voice recording affords the precise measurement of both measures, which may allow us to unravel this puzzle in a single study.

5.2.4. Modulation of the EVS

Like eye-movement measures, intra-individual variability in the EVS exists. The EVS changes within and across, lines and sentences (Buswell, 1920; Fairbanks, 1937; Quantz, 1897). It also changes based on local processing difficulties such as word frequency (Laubrock & Kliegl, 2015; Silva et al., 2016), where larger spans are associated with high frequency words compared to low frequency words. This intra-individual variability could also reflect the assimilation rate of meaning (Buswell, 1920) and the moment-by-moment change in the amount of phonological information buffered in working memory for later articulation (Laubrock & Kliegl, 2015). Considering the latter, situations where the EVS falls outside the working memory capacity, are met with oculomotor adjustments (Inhoff et al., 2011; Laubrock & Kliegl, 2015). Oculomotor adjustments occur through increases in fixation durations, refixations and regressions to reduce the span or allow the voice to catch up before progressing through the text. The analysis in Chapter 4, section 4.4.6. further confirmed increases in articulation rates due to a large EVS. However, what remains unexplored is whether there are individual differences in the way readers modulate their span as they read. Perhaps, good readers are likely to use articulation rates to maintain the EVS rather than eye movements. This proposition is likely as their high-quality lexical representations (Perfetti, 2007) may result in faster and rapid eye movements that are less penetrable to the influence of higher-order cognitive processes and require fewer attentional resources (Herdman & LeFevre, 1992). Such a proposition would resemble the notion of encapsulated word recognition processes in skilled readers (Perfetti & McCutchen, 1986; Stanovich, 1990), where context effects are minimal (Ashby et al., 2005). On the other hand, less-skilled readers may use the oculomotor system more than the speech system to maintain the EVS due to their already slow eye movements.

5.2.5. The present study

Our primary goal was to document the reliability of children's EVS during oral reading and use this as a justification to further examine the sources of individual variability in the EVS. We also present these for eye movement measures due to the paucity of documented reliability estimates in oral reading. Various EVS measures have been examined within eye movement studies of oral reading. However, to our knowledge, there has been no direct examination of the consistency of these measures. Taken together, we aimed to address the following research questions:

1. How reliable are measures of the eye-voice span and eye movements during oral reading?
2. What is the relationship between offline measures of reading-related skills and eye-movement measures?
3. What is the relationship between offline measures of reading-related skills and eye-voice span measures?
4. Are there individual differences in how the EVS is modulated during reading?

5.3. Method

This section is partly taken from Adedeji et al. (under revision), as the data analysed is a subset of the dataset collected during a larger project designed to examine children's reading development during oral reading.

5.3.1. Participants

The initial sample comprised 64 children from two primary schools in Bournemouth. They participated in the research after school permission, parental consent and child assent were given. All participants were in grades three, four or five and aged between 7 and 10 years (29 female, *Mean*= 8.88, *SD*= 0.93). Except for seven children who spoke at least one other language, all were English monolinguals. Their vision was reported as normal or corrected to normal, and none had a prior diagnosis of reading disorders. The data from 12 participants were excluded due to: technical errors (10), school absence during offline measures data collection (1) and excessive head movements during the eye-tracking experiment (1). Therefore, the data from 52 participants between the ages of 7 to 10 years were included in the current analyses. All participants were naïve as to the purpose of the experiment. Bournemouth University's Research Ethics Committee approved the study (ID 28325).

5.3.2. Materials and Design

Two sets of 42 experimental passages were developed, and each participant read only one set. Each passage consisted of two sentences spanning two lines and between 70 and 101 characters (*M*= 87.18 characters, *SD*= 8.03). There were between 13 and 22 words in each passage (*M*= 17.56, *SD*= 2.10). The experimental manipulation included six-letter target words of either one or two syllables embedded in the first sentence of each passage. However, this manipulation was not considered for the current analysis. The 84 passages were assessed for readability using Flesch-Kincaid reading ease metrics (*M*= 2.33, *SD*=

0.92). As each participant only read 42 passages, the assignment of participants to the different sets was counterbalanced.

5.3.4. Offline ability measures

Word and Non-word reading

The children's reading skill was measured using the sight word efficiency and pseudo word decoding efficiency subtests of the Test of Word Reading Efficiency 2- Form A (TOWRE; Torgesen et al., 1999). Each subtest consisted of a list of words and nonwords respectively where children were asked to read aloud as many items as they could in 45 seconds in each list. The TOWRE-II tests confirmed that all the children read at age-appropriate levels (*Mean*= 105.51, *SD*=9.58).

Spelling

The spelling subtest of the Wechsler Individual Achievement Test II for Teachers (WIAT-II-T; Weschler, 2006) was administered to assess their spelling production ability. In this test, the child was asked to write down a list of dictated words of increasing difficulty.

Rapid automatised naming speed

Rapid Automatized Naming and Rapid Alternating Stimulus tests (RAN/RAS; Wolf & Denckla, 2005) assessed naming speed. Specifically, only the letters and numbers subtests were used as these are more predictive of eye movements in skilled reading (Kuperman & Van Dyke, 2011) and discriminate between good and poor readers (Wolf et al., 1986). A card with 50 letters or numbers randomly presented in a ten-by-five matrix was presented to the children, and they were asked to name them as quickly as they could while they were timed.

Vocabulary

The Wechsler Abbreviated Scale of Intelligence II (WASI-II; Weschler, 2011) was administered to ensure all children possessed the requisite language and reasoning skills to participate in the experiment. Only the test's matrix reasoning and vocabulary subtest were administered to derive a composite IQ measure. The WASI-II indicated that all the children were of average intelligence (Mean= 109.16; SD= 15.84). Additionally, the vocabulary subtest score was used to measure reading-related ability in the analysis.

5.3.5. Apparatus

An SR Research EyeLink 1000 Plus desktop-mounted eye-tracker was used to collect eye movement data with a sampling frequency of 1000Hz. Although reading was binocular, only recordings from one eye were taken from each participant (Left eye movements were recorded for five participants due to tracking problems). Participants sat 70cm from the screen, and the forehead rest was used to minimise head movements. However, the chin rest was not used to allow unhindered articulation while reading. The passage, formatted in a black 22-point monospaced Consolas font on a white background and each letter subtended ~ 0.34^{o} horizontally, was presented on a BenQ XL2410 T LCD monitor with a 1920 x 1080 screen resolution and 60 Hz refresh rate. The lines were doubled spaced, justified to the left and presented in the middle of the screen vertically and with a 550-pixel horizontally offset. A Fifine USB Microphone –K056 Model device was used to record participants voices (*lag range*= 3 to 24 ms). MATLAB R2018a (MathWorks, 2014), Psychtoolbox v.3.0.11 (Brainard, 1997; Pelli, 1997) and Eyelink (Cornelissen, Peters, & Palmer, 2002) libraries were used for experimental programming and run on a Windows 7 operating system.

5.3.6. Procedure

Testing was administered in quiet rooms within the children's schools, where they completed two sessions. The children completed an eye-tracking experiment in one session and a paper-and-pencil offline assessment of reading and cognitive ability in the second session. There was no order to completing both sessions as some completed the eye-tracking experiment first and others the offline assessment first.

The eye-tracking experiment began by instructing participants to read the passages aloud and say "done" once they finished so that the experimenter could terminate the trial. Three practice trials were included at the start to familiarise participants with these instructions. The participants completed a 9- point calibration and validation procedure. Validation accuracies were always < 0.40, and recalibration was done whenever the drift check fell below this level. Recalibration was also done following a 2-minute break after half of the passages were read. Participants' fixation on a 50-pixel black gaze box centred on the first letter of the passage triggered the passage presentation. Participants answered TRUE/FALSE comprehension questions that appeared after 14 passages (33%) by pressing one of two buttons on a keyboard. The offline testing and experimental sessions lasted approximately 1 hour, with opportunities for a break between sessions.

5.3.7. Data analysis

Eye movement data were manually pre-processed using Eye-doctor v.0.6.5 (Stracuzzi & Kinsey, 2009), and the EMreading R package (Vasilev, 2018) software was used to obtain the fixation and word-level data. Audio data were pre-processed manually using the PRAAT software (Boersma & Weenink, 2019). The eye and voice data were synchronised using an R script to compute the EVS. It was crucial to capture both the when and where of eye movement behaviour for the individual difference analysis. Therefore, gaze duration was included because it captures early lexical processing in children. Skipping probability and forward saccade length were chosen as spatial measures because these measures give an indication of the extent of parafoveal processing. Regression probability was included as they reflect comprehension processes. In line with Staub (2021), it is expected these four measures should provide a representative view of eye movement behaviour in children. Additionally, three average estimates of the EVS were analysed: spatial EVS, temporal onset EVS and temporal offset EVS (see Figure 5.2 for a graphical illustration of spatial and temporal onset EVS). We also considered the variability in spatial EVS using the standard deviation (SD) of each participant's EVS. Variability in the EVS has been associated with better reading skills where good readers have more variability in their spans to reflect the rhythmic and expressive component of fluent reading (Buswell, 1920; Percle et al., 2020). The unit of measurement of the EVS was a character space (a single grapheme or interword space).

Different measures of EVS are highlighted in Table 5.1, and each of these measures is subject to some amount of data loss. Words can be omitted and/or skipped during oral reading. Although omissions (failing to pronounce a word) are rare, skipping words with the eyes is quite common, especially for adults. Therefore, the different measures will vary in number of observations (See Table 5.1). In the current data, the spatial EVS (articulation onset measure) had 29,366 observations, the spatial EVS (fixation-onset measure) had 30,105 observations and the temporal onset and offset measure had 24,038 observations. The temporal measure had fewer observations likely due to the number of skipped words. However, the fixation-based measure is also affected by word skipping but has a higher number of observations due to refixations and regressive fixations. As will be seen spatial EVS (articulation onset measure) was used for the individual differences in EVS analysis in section 5.4.2 due to its higher reliability estimate compared to spatial EVS (fixation onset measure). However in section 5.4.3, the spatial EVS (fixation onset measure) was used because of the need to measure moment-by-moment changes in fixation durations due to the EVS. Using the articulation onset measure would be impossible as the EVS at this temporal and spatial location is not linked to the eye movement behaviour at that time. Similarly, fixation duration rather than gaze duration was used in section 5.4.3.

In this study, we adopted split-half reliability for convenience and also because reliability estimates within a single sitting have been shown to produce more robust estimates measures (Foster et al., 2018; Staub, 2021; Tinker, 1936). Our choice of analysis included Pearson’s product moment correlation, multiple linear regressions and linear mixed modelling for the reliability statistics, individual differences in eye movement and EVS, and modulation of the EVS, respectively. These analyses were executed in R software v 4.0.3 (R Core Team, 2020). For most tables, statistically and marginally significant values are formatted in bold and italics, respectively.

5.4. Results

5.4.1. Descriptive and Reliability Statistics

Reliability estimates and descriptive statistics are highlighted in Table 5.2. All local temporal eye movement measures showed excellent internal consistency with reliability estimates of at least 0.80. Similarly, other spatial measures (i.e., skipping, regression and forward saccade length) were within the excellent range except for refixation probability. Although both measures are highly correlated, there was a 2-character space difference between the two spatial EVS measures. We revisit this and a potential explanation in the discussion section. Although the measure based on articulation onset was slightly more reliable, the two spatial EVS measures showed similar reliabilities. Interestingly, the temporal onset EVS was as reliable as most temporal eye movement measures.

The relationship between eye movements, EVS and offline measures are shown in Table 5.3. Reading, spelling, and vocabulary scores were negatively and moderately related to gaze duration (-.47 to -.55) and positively and moderately related to spatial EVS (.53 to .61). Higher scores in spelling were also associated with longer saccades and larger variability in the spatial EVS. Additionally, higher vocabulary was related to greater variability in the EVS. RAN exhibited no significant relationship with any eye movement or EVS measures. Regression and skipping probabilities and temporal onset and offset EVS exhibited no significant relationship with the offline measures.

Larger average spatial EVS was significantly associated with shorter gaze durations and faster oral reading rate, while greater variability in the EVS was significantly associated with longer saccade lengths, higher regression probability and faster reading rate. Longer temporal onset EVS was associated with a higher regression probability, while offset EVS was associated with high regression probability and longer saccade lengths. Examining the relationship between all the EVS measures, only temporal offset EVS was moderately and positively related to all three other EVS measures.

Table 5.1. EVS measures and Examples of Authors who have used each Measure (the author list is not exhaustive)

Measure	Sub-measure	Definition	Data loss	Authors
Spatial EVS (Character spaces)	Articulation onset	The distance in character spaces between the first letter of the articulated word and the fixation location at the start of articulation.	Only verbally omitted words will not have an EVS	Buswell (1920); Gordon and Hoedemaker (2016)*
	Fixation onset	The distance in character spaces between the fixation location and approximate phoneme the voice is uttering either at the fixation's start (ONSET) or end (OFFSET).	Skipped words and some words at the onset of the trial will not have an EVS	Fairbanks (1937); De Luca et al. (2013); Laubrock and Kliegl (2015); Easson et al. (2020)[a]
Temporal EVS (Milliseconds)	-	The time difference between the start of the first fixation on a word and the start of its articulation (ONSET) or the end of the gaze and the start of its articulation (OFFSET)	Skipped words will not have an EVS	Inhoff et al. (2011); Järvilehto et al. (2009); Laubrock and Kliegl (2015); Silva et al. (2016)[b]; Geyer (1967); Halm et al. (2011); Morton (1964); Zhou et al. (2021); Easson et al. (2020)[a]

Note. [a] RAN studies. [b] Offset EVS measure used.

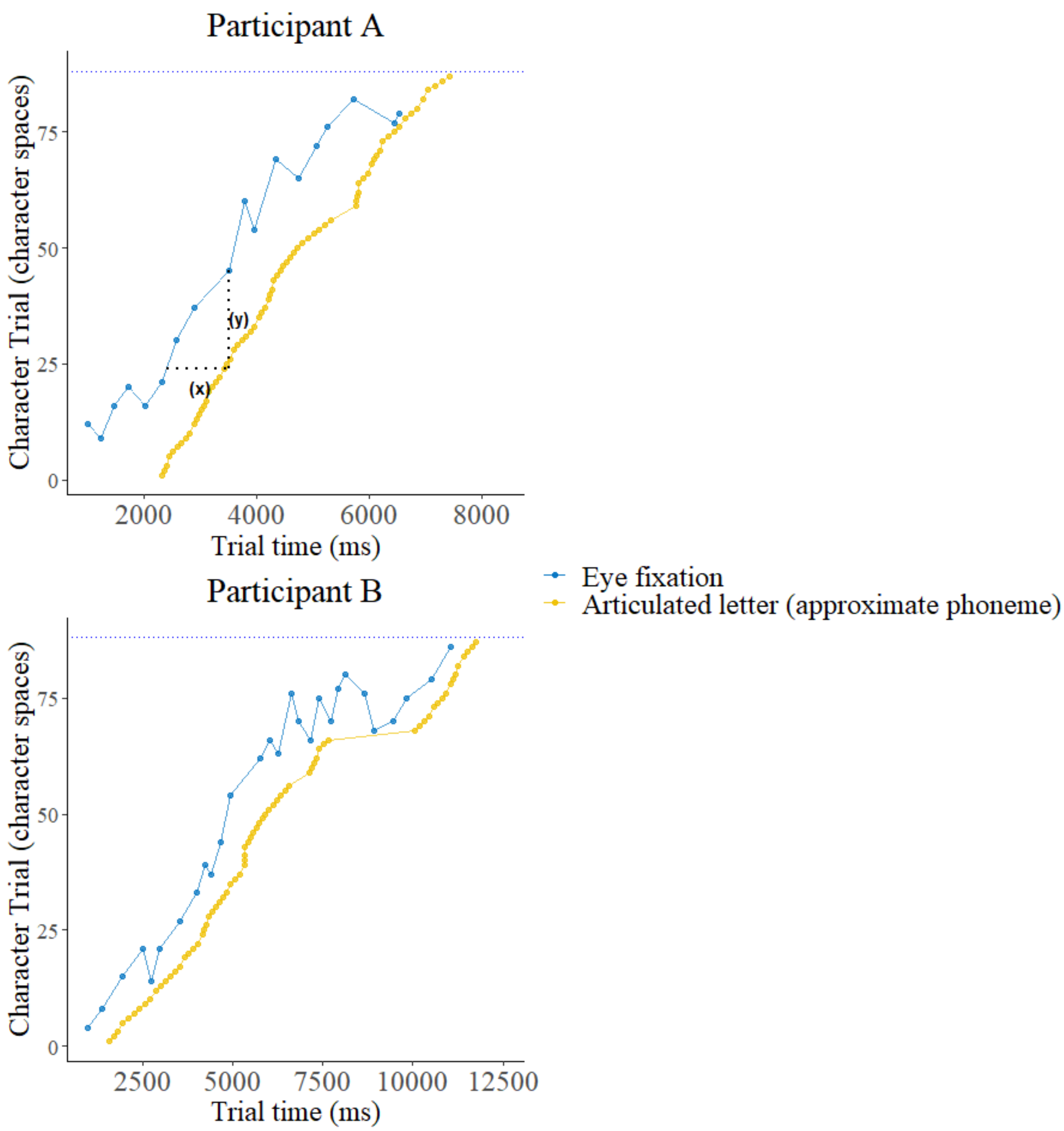

Figure 5.2. Graphical illustration of the spatial EVS (y) and temporal onset EVS (x) for participant A with an average spatial EVS of 10.20 character spaces and TOWRE standard score of 120 and participant B with an average spatial EVS of 6.34 character spaces and TOWRE standard score of 90.

Table 5.2. Reliabilities and Descriptive Statistics for Eye Movement, EVS and Offline Measures

Measure (N=52)	**Mean**	**SD**	**Minimum**	**Maximum**	**Reliability**
Eye movement measures					
First fixation duration	370	82	239	617	0.96
Gaze duration	434	94	277	649	0.96
Total viewing time	571	104	415	893	0.94
Refixation probability	0.17	0.05	0.14	0.65	0.79
Skipping probability	0.16	0.07	0.07	0.38	0.92
Regression probability	0.30	0.10	0.14	0.69	0.84
Forward saccade length	6.03	1.01	4.27	9.31	0.95
Eye-voice span measures					
Spatial EVS-fixation onset	10.96	1.82	7.35	15.05	0.72
Spatial EVS- articulation onset*	8.96	1.57	5.86	12.91	0.77
Variability In Spatial EVS	5.84	2.06	3.12	13.71	0.77
Temporal onset EVS	1077	175	822	1572	0.92
Temporal offset EVS	614	192	246	1151	0.92
Offline ability measures (unstandardised scores)					
TOWRE-II-Reading	51.66	8.56	36.5	69.0	0.94 -0.97
WIAT-II-T-Spelling	30.88	6.37	20	45	0.94
RAN/RAS speed	27.80	5.38	18.23	40.53	0.87-0.98
WASI-II Vocabulary	27.12	6.57	9	37	0.91

Note. Forward saccade length excludes return-sweep saccades. * Spatial EVS at articulation is used for subsequent analysis due to its higher reliability. TOWRE scores include word and non-word reading. Standardized scores for the offline measures are in Appendix D. Reliability statistics for the offline measures were obtained from the respective test manual

Table 5.3. Correlation Matrix for Eye Movement, EVS and Offline Measures

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**
1. GD	-													
2. TVT	.86*	-												
3. Sac Len.	-.33	-.18	-											
4. Skip Prob.	-.42	-.22	.70*	-										
5. Regress Prob.	-.44	.07	.40	.48*	-									
6 Spat. EVS _M	-.61*	-.58*	.33	.38	0.17	-								
7. Spat. EVS_V	-.42	-.20	.61*	.38	0.48*	.29	-							
8. Temp. Onset EVS	.09	.48*	.24	.35	0.66*	.24	.33	-						
9. Temp. Offset EVS	-.43	-.01	.40	.52*	0.82*	.53*	.52*	.86*	-					
10. ORR	-.73*	-.81*	.45*	.24	0.01	.68*	.51*	-.27	.16	-				
11. Reading	-.55*	-.57*	.42	.29	0.10	.61*	.43	-.10	.20	.77*	-			
12. Spelling	-.48*	-.48*	.53*	.34	0.13	.53*	.55*	.00	.26	.70*	.70*	-		
13. Vocab	-.47*	-.38	.42	.20	0.24	.54*	.49*	.15	.40	.57*	.51*	.57*	-	
14. RAN	.20	.20	-.27	-.18	-0.04	-.11	-.34	.15	.03	-.40	-.50*	-.25	-.28	-

Note. *p<0.05 (Holm's correction for multiple comparisons). FFD had similar correlations coefficients to GD (Gaze Duration), TVT (Total Viewing Time), ORR (Oral Reading Rate), M(Mean), V(Variability), RAN (Rapid Automatised Naming)

5.4.2. Contributions of reading-related skills to the eye movements and EVS

Multiple linear regression models predicting the eye movement and EVS measures from the offline measures are shown in Tables 5.4 and 5.5. Statistically and marginally significant values are formatted in bold and italics, respectively. Scatterplots showing the relationship between the EVS, and the offline measures are illustrated in Figure 5.3. The analysis reported here are global representing eye movement measures that were averaged across all words for each participant. With regard to eye movement measures, only the reading score significantly predicted gaze duration, and the model accounted for 31% of the variance. Spelling significantly predicted forward saccade length, with the model accounting for 26% of the variance. None of the offline measures predicted skipping or regression probability, with the models explaining 5% and negligible variance, respectively.

To reiterate, in this analysis, the spatial EVS used was the articulation onset measure due to its higher reliability. Reading, vocabulary and RAN speed significantly predicted average spatial EVS, with the model accounting for 46% of the variance. Participants with higher reading and vocabulary were more likely to have larger spatial EVS. However, participants with faster RAN speed were more likely to have smaller EVS when all other offline variables were controlled. The offline measures of reading-related skills accounted for 33% of EVS variability. Vocabulary and spelling marginally and significantly predicted EVS variability respectively. Vocabulary predicted temporal offset EVS with the model explaining 12% of the variance. The model explained negligible variance for temporal onset EVS.

Table 5.4. Linear Regression Models Showing Eye Movement Measures as a Function of the Offline Measures of Reading-related skills

	Gaze duration			Saccade amplitude		
	b	*SE*	*t*	*b*	*SE*	*t*
Intercept	438.64	10.85	40.43	6.02	0.12	49.74
Reading	-41.19	17.59	**-2.35**	-0.02	0.19	-0.10
Spelling	-6.32	16.74	-0.39	0.43	0.19	**2.29**
Vocabulary	-22.45	13.68	-1.64	0.16	0.15	1.07
RAN	-9.85	12.88	-0.77	-0.13	0.14	-0.939
	Skipping probability			Regression probability		
	b	*SE*	*z*	*b*	*SE*	*z*
Intercept	0.154	0.010	17.166	0.2990	0.0138	21.5560
Reading	0.003	0.014	0.187	-0.0026	0.0224	-0.1160
Spelling	0.019	0.013	1.394	0.0001	0.0214	0.0050
Vocabulary	-0.001	0.011	-0.081	0.0260	0.0175	1.4880
RAN	-0.006	0.011	-0.585	0.0023	0.0164	0.1410

Table 5.5. Linear Regression Models Showing EVS Measures as a Function of the Offline Measures of Reading-related Skills

	Mean Spatial EVS (character spaces)			Variability in EVS (character spaces)		
	b	*SE*	*t*	*b*	*SE*	*t*
Intercept	8.92	0.16	56.84	7.21	0.35	20.68
Reading	0.85	0.25	**3.34**	-0.32	0.57	-0.57
Spelling	0.04	0.24	0.15	1.31	0.54	**2.44**
Vocab.	0.48	0.20	**2.45**	*0.74*	*0.44*	*1.68*
RAN	0.40	0.19	**2.13**	-0.66	0.41	-1.60
	Mean Temporal Onset EVS (ms)			Mean Temporal Offset EVS (ms)		
	b	*SE*	*t*	*b*	*SE*	*t*
Intercept	1076.63	24.07	44.72	611.56	29.93	24.52
Reading	-29.45	39.04	-0.75	9.65	40.43	0.81
Spelling	-0.93	37.15	-0.03	8.44	38.48	0.22
Vocab.	48.23	30.37	1.59	**75.84**	**31.45**	**2.41**
RAN	23.96	28.59	0.84	33.69	29.61	1.14

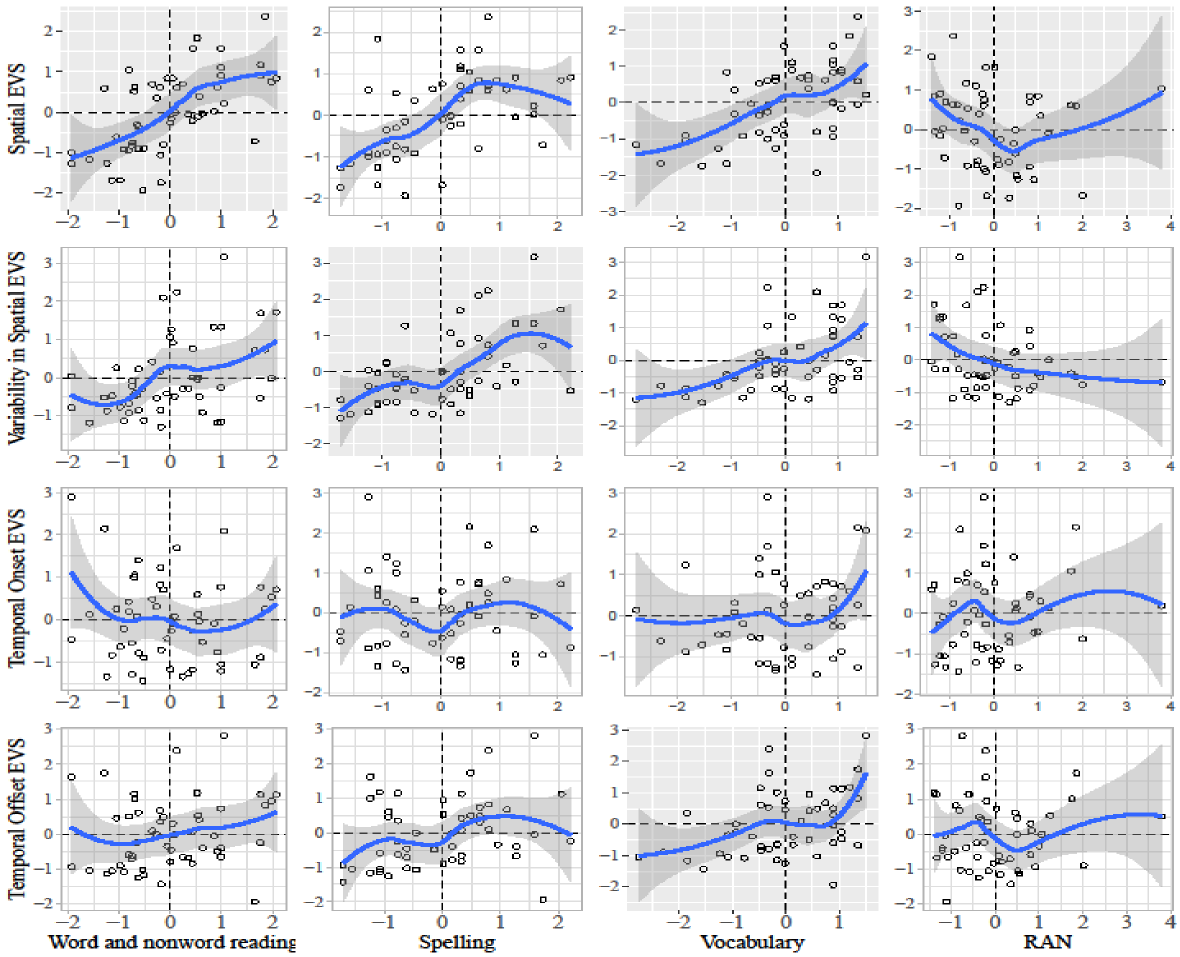

Figure 5.3. Scatterplots showing the relationship between EVS measures and offline measures. Gray themed plots highlight significant relationships from the linear model.

5.4.3. Individual differences in EVS modulation

As readers progress through the text, their EVS is constantly changing. With each forward saccade, the EVS widens. In this section, fixation durations on each word (rather than averaged across all words for each participants as in section 5.4.2) were analysed as dependent variable rather than gaze duration because during each fixation, the EVS narrows as the voice catches up with the eyes. The analysis of fixation durations is also compatible with the analysis by Laubrock and Kliegl (2015) in adults. In order to maintain an EVS within a reasonable range, readers must modulate their eye movements and articulation rates. We explored individual differences in how the EVS is modulated during oral reading as previous research in Chapter 4 showed that both voice and eye movements could modulate the eye and voice distance. As an initial step, we used the anova() function to confirm that including random slopes in the fixation duration, regression probability and articulation rate models improved the fit considerably. The loglikelihood for the fixation duration baseline model without random slopes was 3119.3. When random slopes for spatial EVS by subjects were included, the change in loglikelihood, $\Delta\chi 2$ (2 df) = 64.54, $p < 9.6e - 15$, shows a significant improvement in the goodness of fit. Additionally, the loglikelihood for the regression probability baseline model, -8567.2, was improved significantly by including random slopes, $\Delta\chi 2$ (2 df) = 26.59, $p < 1.7e - 06$. The loglikelihood for the articulation rate baseline model without random slopes was 9923.6. When random slopes for spatial EVS by subjects were included, the change in loglikelihood, $\Delta\chi 2$ (2 df) = 480.28, $p < 2.2e -16$, showed a significant improvement in the goodness of fit. Rather than extract conditional means to examine individual differences, which lead to exaggerated correlations, we followed recommendations by Kliegl et al. (2010) and examined the interaction between EVS and word reading ability on fixation durations, articulation rates and regression probability.

We fitted (generalised) linear mixed models similar to Chapter 4 including a similar set of covariates of word frequency and length for the fixation duration analysis, frequency for the articulation rate analysis (since phoneme length had been controlled for in the calculation of articulation rate). For the regression analysis, no lexical variables were included as we were interested in simple effect of the EVS and its interaction with word reading skills. This decision is also compatible with Laubrock and Kliegl (2015). Given the focus on examining individual differences in the EVS modulation, no consideration was given to interactive effects between reading skill and the lexical variables. The (G)LMMs and effects plots are shown in Tables 5.6, 5.7, 5.8 and Figures

5.4 and 5.5. The results show a main effect of spatial EVS on fixation durations where larger EVS leads to longer fixations. We also found an effect of reading ability in the expected direction. Better word and non-word readers had shorter fixation durations. Interestingly and in line with our individual differences approach, we found an interaction between EVS and reading skills where the EVS had a greater impact on readers of less ability compared to more proficient readers or decoders. We found a main effect of EVS for regression probability, and other effects were absent. Similar to the fixation duration model, the EVS impacted articulation rate. Reading ability also impacted articulation rates, and there was a two-way interaction between EVS and reading ability where the EVS had a greater impact on the articulation rate of lower ability readers.

Table 5.6. LMM Showing Log-transformed Fixation Durations as a Function of TOWRE Word and Non-word Reading Scores and Spatial EVS

Fixed effects	*b*	*SE*	*t*	*p*
(Intercept)	2.4655	0.0076	323.5916	<0.01
Spatial EVS	0.0090	0.0014	6.2653	**<0.01**
TOWRE-Reading	-0.0471	0.0107	-4.4154	**<0.01**
Fixated Word length	-0.0273	0.0020	-13.3654	**<0.01**
Fixated Word frequency	-0.0211	0.0020	-10.5698	**<0.01**
Spatial EVS * TOWRE-Reading	-0.0061	0.0019	-3.1539	**0.0016**
Fixated Word length* Fixated Word frequency	-0.0106	0.0016	-6.5559	**<0.01**
Random effects	Var.	SD		
Item intercept	0.00008	0.00900		
Participant intercept	0.00274	0.05237		
Residual	0.04635	0.21528		

Table 5.7. GLMM Showing Regression Probability as a Function of TOWRE Word and Non-word Reading Scores and Spatial EVS

Fixed effects	*b*	*SE*	*t*	*p*
(Intercept)	-1.3104	0.0647	-20.2401	<0.01
Spatial EVS	0.5864	0.024	24.4273	**<0.01**
TOWRE-Reading	-0.1271	0.0886	-1.4343	0.1515
Spatial EVS * TOWRE-Reading	-0.013	0.0327	-0.3971	0.6913
Random effects	Var.	SD		
Item intercept	0.0339	0.1842		
Participant intercept	0.1698	0.4120		

Table 5.8. LMM Showing Log-transformed Articulation Rate as a Function of TOWRE Word and Non-word Reading Scores and Spatial EVS

Fixed effects	*b*	*SE*	*t*	*p*
(Intercept)	0.9693	0.0073	133.229	**<0.01**
Spatial EVS	0.0407	0.0012	35.2299	**<0.01**
TOWRE-Reading	0.0713	0.0098	7.2875	**<0.01**
Word frequency	0.0294	0.001	28.3805	**<0.01**
Spatial EVS * TOWRE-Reading	-0.0182	0.0015	-12.2594	**<0.01**
Random effects	Var.	SD		
Item intercept	0.0050	0.0223		
Participant intercept	0.0023	0.0482		
Residual	0.0265	0.1628		

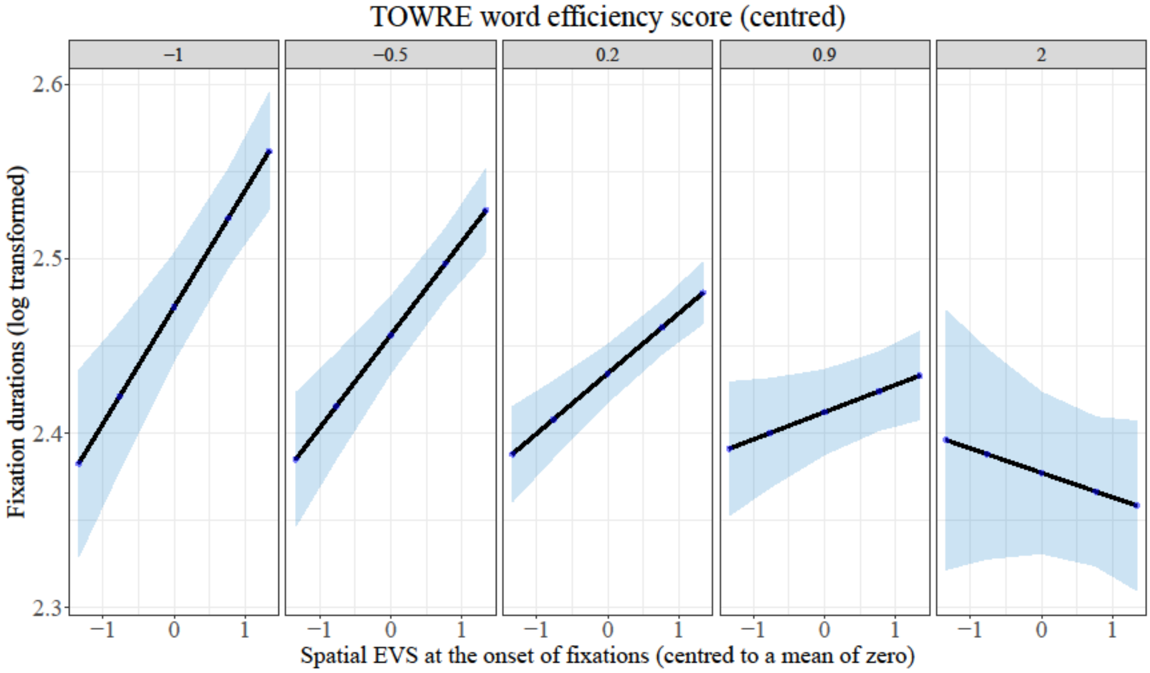

Figure 5.4. Line effect plots showing the interaction between EVS and TOWRE reading for log-transformed fixation durations. Shading shows ± 1 SE.

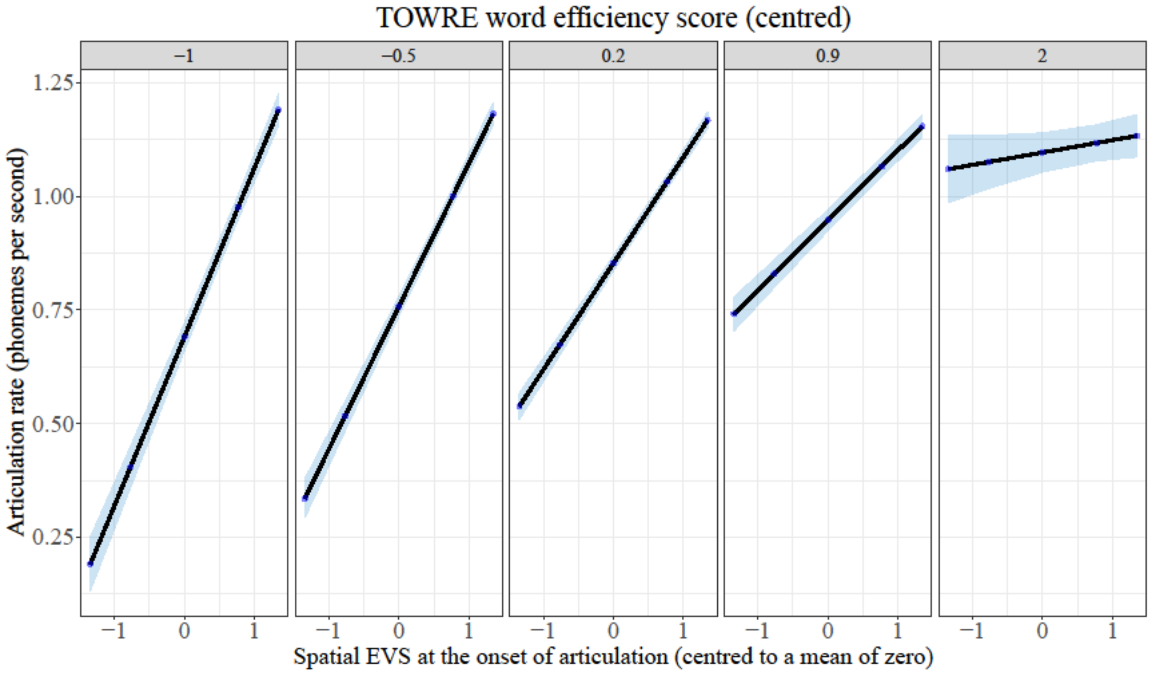

Figure 5.5. Line effect plots showing the interaction between EVS and reading for log-transformed articulation rates. Shading shows ± 1 SE.

5.5. Discussion

Eye-tracking studies exploring developmental reading processes have grown in the last decade. However, few studies have utilised the oral reading paradigm to track reading development and individual differences compared to silent reading. Similarly, studies on the EVS are scarce in developing readers. In this study, we employed simultaneous voice and eye movement recording to obtain the EVS and eye movement measures. This methodology provides a holistic understanding of individual differences in children's reading. For the first time, we report adequate reliability estimates of two broad measures of the EVS during children's sentence reading: spatial and temporal measures. Except for skipping probability, spatial measures of eye movements and the EVS showed lower but acceptable reliabilities compared to temporal measures of eye movements and EVS, which had high reliabilities. Additionally, we explored individual differences in eye movements and the EVS and how these individual differences may influence the modulation of the EVS during reading. Word reading efficiency and spelling skill predicted gaze duration and saccade length, respectively. Individual differences in the EVS were predicted by vocabulary knowledge, spelling and RAN speed. Only spelling and vocabulary marginally predicted intra-individual variability in the EVS, respectively.

Vocabulary knowledge was the only skill to predict temporal offset EVS. Finally, the fixation duration and articulation rate of readers with high scores on the TOWRE word reading efficiency test were less affected by the EVS compared to those of less skilled readers.

The temporal eye movement measures such as first fixation duration (0.96), gaze duration (0.96), total viewing time (0.94) showed excellent reliability above 0.8 (Hedge et al., 2018). These estimates are slightly higher than what has been found recently with adults and children. For example, gaze duration in adults has been reported as 0.87 or 0.91 (Staub, 2021) and 0.82 or 0.89 for children (Foster et al., 2018). Our estimate of 0.96 in children's oral reading could be due to using split-half reliability, which gives better reliable estimates (Staub, 2021) and using all words in each trial rather than target words (Foster et al., 2018; Tinker, 1946). Surprisingly, the reliability estimate for skipping is higher than has been previously reported. Foster et al. (2018) reported low alternative form reliability, while we report an estimate of 0.92. Differences in this estimate could be due to differences in reading modality and reliability method. Oral reading requires overt articulation, making it crucial that most words are fixated with little room for risky reading, i.e., skipping words (Adedeji et al., 2021). Although skipping rates are lower in oral reading, they appear to be more consistent within individuals over different trials compared to silent reading. This proposition was confirmed using a reliability analysis from Parker et al. (2020) where children read silently, and the skipping reliability was 0.67 (see Appendix D). Reliability estimates for gaze duration, first fixation and total viewing time were generally comparable with the current study. Therefore, skipping behaviour may be less consistent in children's silent reading providing support for studies using oral reading rather than silent reading (Molina et al., 2020; Trauzettel-Klosinski et al., 2010).

The spatial EVS using either articulation or fixation onset showed good reliability, at least 0.7 (Hedge et al., 2018). However, our reported estimates (0.72 and 0.77) differ from what has been previously reported in the RAN literature (0.88; Gordon & Hoedemaker, 2016). This difference could be due to the semantic and syntactic constraints of sentence reading which increases the variability of the EVS, where differences in structure and punctuation in each item may cause variability within an individual participant. Additionally, the difference between both spatial measures may reflect measurement errors. This is because locating the articulated phoneme at the start of fixation for the spatial EVS at fixation onset was based on the assumption that the time to articulate each letter/phoneme is dependent on the word duration and number of letters in the word (Laubrock & Kliegl, 2015). There was also a difference in the mean EVS for the two spatial EVSs. The EVS at fixation onset was 2-character spaces more than the EVS at articulation onset. This difference is likely because the EVS is typically highest when measured at the onset of forward fixations, and articulation onset is likely to occur not at the start but during the fixation. During the time between the start of fixation and the start of articulation for a word, the voice is catching up to the eyes. We found that, on average, articulation begins approximately 194ms after starting a fixation (range: 1-969ms), and approximately four phonemes are articulated within the average fixation of 400ms. Therefore, at the start of word articulation, the EVS would have decreased by approximately two phonemes or letters, which is the difference we found. Interestingly, the temporal EVS had a high reliability estimate like the temporal eye movement measures. This could be because the temporal estimate of the EVS is less variable within individuals. Geyer (1967) proposed that the temporal EVS was a constant value of about 1000ms and his later studies showed that the temporal EVS did not vary due to difficulty of text like the spatial EVS. This suggests that individuals consistently produce speech within a constant time frame once they begin to process it.

Having found reasonable reliability estimates, we explored individual differences in eye movements and the EVS using four measures of reading-related skills. Our regression analysis broadly supports previous findings. First, reading ability, measured by sight word reading and non-word decoding, predicted differences in gaze duration. These findings support silent and oral reading research where lexical processing triggers the eye’s forward movement (Foster et al., 2018; Hessel et al., 2020; Kim et al., 2019, 2020; Lee, 2001; Reichle et al., 2003). We extend this work by showing that spelling ability impacts saccade length in developing readers. Although spelling and reading were moderately correlated (0.70), a clear differentiation between the impact of word reading and spelling on gaze duration and saccade length, respectively, is quite interesting. Reading ability has been associated with foveal processing and spelling ability parafoveal processing (Slattery & Yates, 2018). Therefore, this finding supports previous work showing that precise orthographic lexical representations as indexed by spelling ability, influences early letter encoding allowing skilled readers to use parafoveal information more efficiently for saccade targeting (Parker & Slattery, 2021; Perfetti, 2007; Perfetti & Hart, 2002; Veldre & Andrews, 2014, 2015). Because parafoveal processing of a word may have occurred on the prior fixation, spelling ability may not predict fixation times. The absence of a RAN speed effect on gaze duration after controlling for word reading supports Kim et al. (2020), where RAN speed did not contribute to variability in eye movements when word reading was accounted for. None of the offline measures predicted skipping and regression probabilities in line with previous developmental research (Foster et al., 2018; Lee et al., 2021).

The EVS reflects lower-level linguistic processing and semantic and syntactic integration (Balajthy Jr, 1978; Gordon & Hoedemaker, 2016; Levin & Turner, 1966). In line with this evidence, word reading, and vocabulary knowledge significantly predicted spatial EVS. Therefore, the rate of access to orthographic, phonological, and semantic representations determines how far the eyes travel in relation to the voice during oral reading. Although the predictive power of vocabulary on gaze duration approached significance, the EVS results suggest that the EVS may be a more sensitive measure in tapping into multiple reading-related skills. Likely, accessing only orthographic and phonological representations may not sufficiently allow the eyes to travel further ahead of the voice. For instance, when word reading and RAN are controlled, a participant with a higher vocabulary is likely to have a larger span than a participant with a lower vocabulary. The ability to chunk information in meaningful units for later articulation seem necessary for maintaining an EVS that is important for fluent oral reading. This finding is in line with the lexical quality hypothesis (Perfetti & Hart, 2002), where rapid form and meaning recognition are necessary for reading comprehension (Perfetti, 2007; Protopapas et al., 2013). Potentially, the EVS may be a better reflection of online comprehension processes compared to eye movement measures (Cooley, 1981).

Surprisingly RAN had an opposite effect on the spatial EVS. Slower RAN performance predicted a larger span between the eye and the voice when all other variables such as reading, spelling and vocabulary were controlled. This pattern bears a semblance with suppressor variables where the direction of a variable (RAN) which has a weaker relationship with the dependent variable ($r=-0.11$; spatial EVS), suppresses the noise in another variable (TOWRE-Reading) which it is correlated with (Friedman & Wall, 2005; Rij et al., 2020). In our models, the predictive strength of reading increased with the inclusion of RAN speed. This also improved the model fit (~4% increase). Although caution must be exercised when interpreting suppressor variables, it could be that participants with similar reading skills may have a similar rate of eye movements. However, if they differ in RAN speed, then slow RAN readers would have their voice lagging behind their eyes compared to those with fast RAN speed leading to a larger span for such readers.

The association between a large span and low ability has been documented previously. Fairbanks (1937) noted that poor readers might have a large EVS due to their inability to identify errors. Rather than regressing to a previous point in the text where an oral reading error occurs, poor readers were likely to continue reading. On the other hand, good readers were more likely to regress after an error had been made to correct it. In fact, Fairbanks (1937) noted that the EVS of poor readers when oral reading errors were made remained similar to their average EVS. This idea is somewhat linked with Buswell's proposition that good readers have flexible and adaptive EVS (Buswell, 1920).

According to Buswell, good readers have more variability in their EVS. However, this is not primarily because they regress to correct oral reading errors as Fairbank (1937) suggested. Instead, a variability in the EVS of good readers indicates their capacity to read in meaning units with rhythm and emphasis rather than monotonously like less skilled readers. In the current study, word reading ability did not predict EVS variability, but vocabulary did marginally. This is in line with a meaning explanation of the EVS and provides support for the link between intra-individual variability in the EVS and reading comprehension skill (Li et al., 2009). Our measure of reading assessed decoding skills rather than reading comprehension. Consistent with this, De Luca et al. (2013) found no differences in EVS variability for typical and dyslexic 11-year-olds, who primarily differ in decoding skills. Instead, we observed that spelling ability also predicted this variability. One possible explanation for this may be linked to saccade targeting. Better spellers tend to make longer forward saccades. Such long saccades may widen the span, leading to regressions to bring EVS back within an acceptable range. This explanation is compatible with moderate correlations between variability in EVS and regressive saccades (0.48). In sum, the offline reading-related measures captured more of the variation in spatial EVS (45%) compared to gaze duration (31%). These findings align with the proposition that the EVS may reflect a multidimensional construct of the reading process compared to what eye movement measures alone can offer.

Although variations in temporal onset EVS measure have been reported between dyslexic and typical readers (Halm et al., 2011) and cognitive demanding and less demanding tasks (Zhou et al., 2021), none of the offline measures predicted the temporal onset EVS. Similar to our explanation regarding the reliability estimate of this EVS measure, Geyer (1967)'s proposition of less variable temporal EVS during skilled passage reading may account for the null effects. This value was hypothesized to be the storage capacity of the buffer between input and output systems during reading aloud such that words need to be articulated within 1000ms of being recognized to prevent fading (Geyer, 1967). Indeed, the average temporal EVS reported here and in Zhou et al. (2021) are consistent with this value but differ from what has been reported in skilled single-sentence reading (Inhoff et al., 2011). Therefore, passage length, passage difficulty and sample characteristics may account for this difference.

Interestingly, the temporal offset measure was predicted by vocabulary. The offset measure of the temporal EVS has been associated with reading rate (Silva et al., 2016), where faster readers exhibit more time between the end of gaze and onset of word articulation. This offset measure is calculated by deducting gaze duration from the temporal onset EVS. Essentially, the offset measure begins at the end of gaze and thus may reflect post-lexical integration of a word in context captured by vocabulary knowledge.

It is perhaps important to draw some attention to the direction of effects for the temporal onset (Järvilehto et al., 2009) and offset EVS (Silva et al., 2016). Recall that from the introduction, these two measures often move in opposite directions, where better reading skill is associated with shorter onset EVS (Järvilehto et al., 2009) but longer offset EVS (Silva et al., 2016). There seemed to be an indication of this opposite pattern for reading and spelling (see Table 5). However, this did not approach significance and more research is needed. Taken together, temporal onset EVS despite showing high reliability, may be insufficient in discriminating individual differences within age groups (see Easson et al., 2020 contrasting evidence in RAN tasks). However, it may be sensitive to group differences (Halm et al., 2011) or task differences (Zhou et al., 2021) during sentence reading tasks.

The relationship between the EVS and eye movement measures has been examined scarcely, but also in different ways. When making a comparison between participants, the average spatial EVS of 11-year-old children with and without dyslexia had no relationship with their average fixation duration (De Luca et al., 2013). In contrast, we report a moderate relationship of -0.61 between spatial EVS and gaze duration, indicating that both measures reflect shared and yet different processes. However, the final part of this study was not concerned with this kind of relationship. Instead, we examined individual differences in intra-individual modulation of the EVS. Laubrock and Kliegl's (2015) analyses suggested that participants increased their fixation duration in response to a large EVS. Our data extend this evidence by showing that there are individual differences in the way this modulation occurs. Children modulate their fixation durations and articulation rates to keep the EVS at a reasonable distance (Chapter 4). However, we found that the EVS effect on fixation durations and articulation rates was reduced for children with higher TOWRE word efficiency scores. This novel investigation of individual differences suggests that better readers may be more tolerant of a large span and do not require immediate strategic adjustments by the voice or eyes to lessen the span. This finding may seem to be at odds with the strong effect of EVS on fixation durations of skilled readers (Laubrock & Kliegl, 2015). However, adult oral reading rates reach a ceiling where speech processes cannot meet the demands of fast and rapid eye movements that come with increased lexical processing efficiency (Calabrese et al., 2016; Carver, 1983; van den Boer et al., 2022). Therefore, speech processes may place more constraints on the eye movements of skilled readers compared to developing readers whose speech rates and eye movements may align more closely. No such interactions between reading skill and EVS were observed for regression probability, indicating that all children modulate the span by regressive saccades regardless of ability.

Overall, the present findings corroborate and extend previous work on eye movements and the EVS in a sample of 52 developing readers. While this is considered a limited sample for an individual differences study, the study lays the groundwork for future EVS studies in showing that the EVS measure captures multiple aspects of the reading process and presents as a valuable tool for understanding reading processes. Further replications with larger samples and comparisons with adults are necessary. Including other cognitive measures such as working memory which has been linked to the EVS (Laubrock & Kliegl, 2015), and reading comprehension may help substantiate some of these findings in a broader and theoretically relevant context.

5.6. Chapter Overview

Several findings in this Chapter agree with linguistic control models of eye movements as well as the lexical quality hypothesis (LQH) documented in Chapter 1 of this thesis. In line with these models, reading-related skills predicted the eye movements as well as the EVS. Word reading and spelling impacted gaze durations and saccade length, respectively. This supports recent research in skilled reading, showing that spelling impacts early orthographic encoding. Our results suggest that children’s capacity to extract precise orthographic representations as indexed by spelling is an important determiner of saccade targeting. Importantly, we found that the spatial EVS was a reliable measure with more of the children’s offline reading abilities explaining a lot of its variance compared to the variance in gaze duration.

In the next chapter and final study of this thesis, we explore a phenomenon that naturally arises from having an EVS, i.e., spoken words are usually different from fixated words during oral reading at any given time. This was important to assess other factors that may impact upon eye movements during oral reading in addition to the EVS.