Vision Research 49 (2009) 2948–2959

Contents lists available at ScienceDirect

Vision Research

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

Oculomotor and linguistic determinants of reading development: A longitudinal study

Lynn Huestegge [a,*], Ralph Radach [b], Daniel Corbic [b], Sujata M. Huestegge [a]

[a] Institute of Psychology, RWTH Aachen University, Aachen, Germany
[b] Center for Reading Research, Florida State University, Tallahassee, USA

ARTICLE INFO

Article history:
Received 19 May 2009
Received in revised form 17 September 2009

Keywords:
Reading
Children
Development
Eye movements
Naming

ABSTRACT

We longitudinally assessed the development of oculomotor control in reading from second to fourth grade by having children read sentences with embedded target words of varying length and frequency. Additionally, participants completed oculomotor (pro-/anti-saccades) and linguistic tasks (word/picture naming), the latter containing the same item material as the reading task. Results revealed a 36% increase of reading efficiency. Younger readers utilized a global refixation strategy to gain more time for word decoding. Linguistic rather than oculomotor skills determined the development of reading abilities, although naming latencies of fourth graders did not reliably reflect word decoding processes in normal sentence reading.

1. Introduction

Reading and its development has been studied for well over a century and yielded a diverse array of methodological approaches. In this introduction, we will focus on three major research lines underlying the rationale of the present study. First, in the tradition of psychometric assessments participants are typically asked to read words, sentences, or coherent text passages. Speed and comprehension are assessed, for example via answering comprehension questions, filling in blanks, judging the meaningfulness of sentences, or assigning corresponding pictures (see Küspert & Schneider, 1998; Landerl, Wimmer, & Moser, 2001; Mayringer & Wimmer, 2003, for German examples). However, although this *psychometric assessment approach* has proven extremely useful in assessing literacy skills, it is principally limited in addressing a more fine-grained level of information processing.

A second approach consists of experimental tasks where participants name or categorize briefly presented single words that are manipulated with respect to features like word length, frequency, or orthographic properties. This *single-word approach* has led to considerable insight into language processing and its development (e.g., Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001; Grainger & Jacobs, 1996; Jacobs & Grainger, 1994). However, although it allows controlled experimentation, a restriction exists in that it remains difficult to know whether and how findings generalize to word processing in sentence reading (see Grainger, 2003; Radach & Kennedy, 2004).

Third, within the *dynamic reading approach* the measurement of eye movements has proven to be a valuable tool for inferring ongoing visual and linguistic processing, either for the purpose of testing specific hypotheses or as a base for models of oculomotor control during reading (e.g., Engbert, Nuthmann, Richter, & Kliegl, 2005; McDonald, Carpenter, & Shillcock, 2005; Radach, Reilly, & Inhoff, 2006; Rayner, 1998; Reichle, Rayner, & Pollatsek, 2003; Reilly & Radach, 2006; Yang, 2006). To date virtually no such modeling approach exists for developing readers (but see Feng, 2006), partly due to a lack of an empirical base regarding children's eye movements in reading. The main aim of the present study is twofold. First, we would like to provide a solid empirical database regarding the intraindividual development of reading that avoids limitations of earlier studies (see below) for future modeling purposes. Second, we asked to what extent the development of a complex skill like reading depends on basic oculomotor vs. linguistic abilities.

Early studies of children's eye movements in reading were conducted by Buswell (1922) and Taylor, Frackenpohl, and Petee (1960), who reported a reduction of mean fixation duration and number of fixations over the course of the first 6 years in school. Interestingly, there was no clear trend in the data for a reduction of regressions (eye movements back to previously inspected text), which ranged between 20% and 25% of all saccades until sixth grade.

Recent work using more advanced methods replicated these early results (e.g., Feng, Miller, Shu, & Zhang, 2009; Rayner, 1985).

* Corresponding author. Address: Institute of Psychology, RWTH Aachen University, Jägerstrasse 17-19, 52056 Aachen, Germany. Fax: +49 241 8092318.
E-mail address: Lynn.Huestegge@psych.rwth-aachen.de (L. Huestegge).

0042-6989/$ - see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.visres.2009.09.012

Additionally, these studies addressed more specific research questions. For example, McConkie et al. (1991) reported that children's eye movements exhibit far more variability than those of adults. Furthermore, they investigated saccade landing positions in words and found that first graders already show typical characteristics of adult saccade metrics: A Gaussian distribution of incoming saccade landing positions, commonly referred to as the preferred viewing position phenomenon (Rayner, 1979), strongly supported the word based nature of saccade control (see also Joseph, Liversedge, Blythe, White, & Rayner, 2009). Another key finding was a sharp reduction of the frequency with which a currently fixated word was immediately refixated from grade one to six. However, over the several measurements each student read age-appropriate (thus different) texts, so that changes in the eye movement records are not clearly attributable to developmental factors.

Another key study reported that the perceptual span (the region around the current fixation within which useful information can be processed) is smaller for developing readers than for adults (Rayner, 1986). Hyönä and Olson (1995) collected the first accurate oculomotor data while fourth graders were reading aloud and found that children spent more time on infrequent as compared to frequent and on long as compared to short words. However, this was no developmental study, and the findings were based on post hoc analyzes of children's reading patterns instead of an experimental manipulation of word length and frequency within controlled sentence environments. In sum, an experimentally well controlled comprehensive longitudinal study of the development of eye movements in reading is still missing.

Whereas each of the three approaches to reading development increased our knowledge in specific domains, little effort has been invested into combining these alternatives. For example, only a few studies combined single-word paradigms and eye tracking during normal reading, and none of these deal with developing readers (e.g., Folk & Morris, 1995; Inhoff, Briihl, & Schwarz, 1996; Inhoff & Topolski, 1994; Schilling, Rayner, & Chumbley, 1998). In the present work, we integrated key features of these approaches by using one common item pool for the study of both word/picture naming and eye movements in sentence reading.

On a general level, reading includes two streams of processing, one being the linguistic processing of words and the other consisting of visuomotor control aimed at providing optimal visual input. While word and picture naming can be utilized as a method to assess comparatively 'pure' linguistic processing skills with the aim to pronounce words (resembling oral reading) without the need to move the eyes, it is also possible to assess oculomotor skills without the need for simultaneous linguistic processing. For example, the prosaccade task requires moving the gaze to a peripheral (meaningless) stimulus and can be used to measure the ability to execute exogeneously triggered (automatic) saccades. In contrast, the antisaccade task requires saccades into the opposite direction of a peripheral stimulus, demanding voluntary oculomotor responses (Findlay & Walker, 1999).

It has previously been suggested that apart from linguistic skills such basic oculomotor skills are driving forces behind successful reading development, including both the execution of bottom-up generated automatic saccades (e.g., based on physical characteristics of the text) and the execution of top-down generated deliberate saccades (e.g., based on current linguistic processing demands). Previous studies reported a clear developmental trajectory of oculomotor performance during childhood (Klein, 2001; Kramer, Gonzales de Sather, & Cassavaugh, 2005) and a substantial reliability of the respective measures (Klein & Fischer, 2005). More importantly, it has been argued that an inadequate development of reflexive and/or voluntary basic oculomotor skills might cause reading disabilities (e.g., Biscaldi, Fischer, & Hartnegg, 2000; Fischer & Weber, 1990). These considerations suggest that basic aspects of oculomotor control as reflected in visuomotor tasks may play a crucial role in the development of reading skills, which will be evaluated in the present study.

More specifically, we implemented a longitudinal design to assess the development of reading and reading-related processes in primary school children, with a focus on the relative role of oculomotor and linguistic determinants of reading skills. We collected children's eye movements during oral reading in second and fourth grade. Oral reading was utilized to closely match the demands in the naming task. Oral reading fluency is at the focus of early reading curricula and serves as a benchmark in numerous reading assessments. Although most adult reading studies utilize silent reading, there is no evidence that the fundamental principles of oculomotor control in reading (e.g., as evidenced by word length and frequency effects on spatial/temporal parameters) substantially differ between silent and oral reading (Hyönä & Olson, 1995), even though some phenomena (e.g., the typical eye-voice span, see Levin & Buckler-Addis, 1979) are unique in oral reading. We additionally examined word and picture naming (utilizing the same item pool as in the sentence reading task) as well as pro- and anti-saccade performance. Using a regression approach, these parameters can be combined as predictors for performance in the reading task, allowing to estimate the relative contribution of oculomotor and linguistic abilities to reading development.

In addition, we included psychometric tests assessing general cognitive abilities (CPM; Raven, Raven, & Court, 1998a, 1998b) and several reading ability tests. In fourth grade, we added short assessments of rapid automatized naming (R.A.N.) performance (as an alternative to the standard naming task) and phonological awareness to draw a more detailed picture regarding determinants and/or correlates of literacy development.

2. Method

2.1. Participants

Twenty-one primary school children with normal or corrected-to-normal vision were tested within 2 months at the end of second (t_1) and fourth grade (t_2), 15 female and 6 male. Their mean age was 8;0 and 10;0 years, respectively ($SD = 0.45$). All attended a local primary school in Aachen, Germany, and enjoyed the same amount of prior reading instruction. They exhibited average to slightly above average reading skills (SLS score: $M = 114.95$, $SD = 17.51$). At school, we additionally tested reading abilities (SLS & WLLP, see below) of all second graders at t_1 ($N = 103$, SLS score: $M = 109$, $SD = 19$), ensuring the representativity of our sample. Since we were interested in normal reading development, we did not test subjects with low reading skills (below 15th percentile). All children participated voluntarily and parents gave their informed consent.

2.2. Material

The selection of items in the *word and picture naming task* was based on 260 color pictures (Rossion & Pourtois, 2004; based on Snodgrass and Vanderwart (1980)). One hundred items were excluded, including homonyms or ambiguous pictures (indicated by less than perfect response correspondence in an ad hoc naming study with two children and three adults). The remainder was partitioned with respect to corresponding word length and frequency. Since we relied on a predetermined item pool, our degrees of freedom in this process were somewhat limited. We classified 4- to 5-letter words ($M = 4.55$) as short and 6- to 9-letter words ($M = 7.05$) as long, and words with < 1 ln words per million as being of low frequency ($M = 0.51$) and the remainder as being of high frequency

($M = 1.59$) (Celex, 1995). Equal item distribution across the four categories was achieved with 30 items in each cell of the design (120 total). However, in this item pool word length and frequency were still statistically associated. For the purpose of examining word frequency and word length effects as part of a factorial design in the reading and naming tasks, we defined a subset of 4×20 items that were fully orthogonal: frequency did not significantly differ between short vs. long words, neither for items of low mean frequency ($M = 0.32$ for short words and $M = 0.36$ for long words, $t(18) = 1.64$, $p > .10$), nor for items with high mean frequency ($M = 1.66$ for short words and $M = 1.55$ for long words, $t(18) = 0.87$, $p > .10$). Note that the frequency variation was limited, since we only included words known to second graders.

For the *reading task*, all 120 items were embedded in active declarative sentences, each containing one item as a target word. They contained child-appropriate statements about animals, family, and school with reappearing protagonists, but without resulting in a coherent story. Sentence length varied from 50 to 70 characters. Target words were never positioned at the first or last two word positions in the sentence. The mean position of the first letter of the target word in the sentence did not differ significantly between all four target word conditions, $F(3, 116) = 1.84$, $p > .05$. Target words were preceded by an adjective of 5–8 letters in length, with a mean frequency of 2.75 ln words per million and frequency not differing significantly across conditions, $F(3, 116) = 0.38$, $p > .05$. The post-target word had a fixed length of 3–5 letters.

2.3. Apparatus

For the reading and oculomotor tasks we used an infrared head mounted eye tracker (Eyelink 2, SR Research Ltd., Canada). Eye movements of both eyes were measured simultaneously with 500 Hz temporal and 0.01° spatial resolution. Saccades were defined by using a combined threshold of minimum saccade velocity (>30°/s), acceleration (>8000°/s) and amplitude (>0.15°). Fixations and saccades interrupted by blinks were excluded from further analysis.

Participants were seated at a distance of 67 cm in front of a 21″ monitor (1024 × 768 pixels, 100 Hz refresh rate). Sentences in the reading task were presented on one central horizontal line on the screen, with each letter comprising a visual angle of 0.33° horizontally.

For the word and picture naming task, we used "E-Prime" software (Psychology Software Tools (PST) Inc., Pittsburgh, USA) and a 15″ notebook equipped with a voice key. Pictures were presented centrally with a resolution of 72 dpi and a size of 281 × 197 pixels, comprising an angle of 10° at a viewing distance of about 50 cm.

2.4. Procedure

All 21 children were tested at t_1 and t_2. At t_1, two short reading tests (group versions of SLS & WLLP) were administered at a primary school in Aachen, Germany. The SLS (Mayringer & Wimmer, 2003; parallel-test reliability: $r > .90$) is a speeded (3 min) *sentence comprehension test* that consists of a list of short simple statements (e.g., "Bananas are blue"). Subjects indicate whether statements (ordered by increasing reading difficulty) are right or false by marking corresponding answer tags. Dependent variable is the amount of correctly classified sentences. The WLLP (Küspert & Schneider, 1998; parallel-test reliability: $.82 < r < .92$) is a speeded (5 min) *word comprehension test* and consists of a list of words each combined with four picture alternatives, including semantically or phonologically similar items. Subjects mark the picture that corresponds to a given word.

Further testing was executed in the laboratory and consisted of the pro-/anti-saccade task and the reading task (see below). A second session conducted about one week later consisted of the word/picture naming task, a more detailed test of reading skills (SLRT) and a short (∼8 min) test for general cognitive skills (CPM; Raven, Raven, & Court, 1998a, 1998b). The SLRT (Landerl, Wimmer & Moser, 1997; parallel-test reliability: $.83 < r < .99$, duration: 10 min) involves speeded assessments of reading on word and sentence level, but also includes *pseudoword reading*. More specifically, it consisted of reading lists of simple nouns, pseudowords (either similar or dissimilar to existing words), and text reading. Dependent variables were list/text reading times.

At fourth grade (t_2), the same children were tested again with exactly the same material in the same order. During the second session we additionally included R.A.N. tasks, a working memory task, and a phonological awareness test (see below).

2.4.1. First measurement (t_1)

At the beginning of the first session (2004), children were seated in front of the presentation monitor and the eye tracker was calibrated using a three-point routine. In the *prosaccade task* students were asked to initially look at a green central fixation cross on a black background (1000–1500 ms) and then direct their gaze to a small green smiley appearing in the periphery. Targets appeared either to the left or right for 1000 ms at an eccentricity of 8°. The smileys as well as the central fixation cross comprised a visual angle of 0.33°. During each trial, the fixation cross either remained visible throughout the trial (overlap condition) or was blanked 200 ms before target onset and reappeared at the beginning of the next trial (gap condition). Gap/overlap trials were mixed. Usually, gap conditions substantially reduce saccade latencies, and previous research linked the ability to produce short latency saccades in a prosaccade/gap task to successful reading development (e.g., Fischer & Weber, 1990). The prosaccade task consisted of 120 randomly arranged trials (plus 10 practice trials). Recalibrations were conducted after each 30 trials.

The *antisaccade task* differed from the prosaccade task only with respect to the instruction, according to which children were asked to look away from the stimulus (opposite direction with comparable amplitude). The ability of executing antisaccades was previously discussed as a crucial factor in successful reading development (Biscaldi et al., 2000). In the oculomotor tasks (30 min) we measured saccade latencies, amplitudes, and the percentage of erroneous prosaccades in the antisaccade task.

After a short break, the *reading task* (50 min) was administered. 120 sentences were presented one after another. A calibration of the eyetracker was executed prior to each sentence. Children were asked to read each sentence orally at normal pace so that they would understand its meaning. Each trial ended with a key press. At unpredictable intervals, a sentence was followed by a comprehension question (17 altogether) which had to be responded to as precisely as possible. The questions were inserted to serve as a comprehension measure. For a sentence like "Lina has again thrown her little shoe out of the car" a question could be "What has Lina thrown out of the car?" Demonstrations of sentences and feedback including the complete adequate answers were given. Each correctly reproduced adjective and noun counted as one score point ("little shoe" resulting in two score points).

Seven practice sentences (plus two questions) were presented at the beginning. Dependent variables in the reading task included reading rate, percentage of regressions (saccades from right to left against normal reading direction), and saccade amplitudes. On target words, we determined initial fixation durations (duration of the first fixation on a word), gaze durations (time spent on a word until it is left for the first time), total reading times (including all reinspections of the word after it was left for the first time), word skipping rates, and initial landing positions of saccades (see Radach & Kennedy, 2004).

The second session (max. 1 week later) started with the *word and picture naming task* (40 min). Items were displayed centrally until a response was given (max. 5 s) and were preceded by a central fixation cross (2 s). Subjects were asked to pronounce the words and pictures as fast and precisely as possible. Words and pictures were mixed in quasirandomized sequence (plus 13 practice trials), with half of the items occurring as pictures first and the other half occurring as words first, so that any potential memory effects were equally distributed across both item categories. Switches between word and picture trials were equally distributed. Dependent variables were response latencies (i.e., the interval between item onset and verbal response initiation) and the amount of errors. Synonyms, generic terms and responses that were too quiet or slightly mispronounced were not counted as errors, but excluded from mean latency calculations.

2.4.2. Second measurement (t_2)

At t_2, all tests described above were administered again. The CPM was supplemented with (more difficult) "set C" items from the Standard Progressive Matrices (SPM; Raven, Raven, & Court, 2000). SLS and WLLP were included at the end of the second session. The faster completion of the tests at fourth grade level allowed including three further tasks, namely rapid automatized naming (R.A.N.), a working memory task, and a phonological awareness task.

For the *rapid automatized naming* (R.A.N.) task, four tables were designed, each consisting of five different digits, pictures, letters, or words. Items in each category were arranged in randomized sequence on one page of five columns and ten rows. Word and picture tables consisted of identical concepts (ball, house, fish, car, chair). All four tables were presented successively. Prior to each administration, the instructor ensured that the children knew how to pronounce all items. Children were asked to name all 50 items per table line-by-line as quickly as possible. List reading duration was measured. Errors (usually triggering self-corrections affecting overall speed) were not separately analyzed.

The *phonological awareness assessment* included two subtests. In the initial sound replacement task, children were asked to replace the initial sound of an orally presented given word with another orally presented sound (e.g. "time" and "/l/" resulted in "lime"). In the initial component change task, children had to change the initial sounds of composite word components (e.g., "beefsteak"-"steefbeak"). For each task, examples were presented to ensure comprehension of the task. Each subtest consisted of 16 items. Every new item was given immediately after the child had responded to the previous item. The total time until completion of the task was recorded as the main variable of interest. Errors (usually triggering self-corrections affecting overall speed) were not separately analyzed. We also implemented a backward digit span task (from HAWIK-R, Tewes, 1983) to assess working memory capacity, but since it did not correlate with any of the reading-related measures, corresponding results will not be further reported.

2.5. Design

The main independent variable (IV) was "grade" (second/fourth). Some tests were analyzed by using matched samples *t*-tests. For variability analyzes we used a dependent-samples *t*-test for equality of variance (Kirk, 1990, p. 414). For target word analyzes in the reading task the IV word frequency and word length were included, requiring a three-way repeated measurement ANOVA. In the naming task, item type (word vs. picture) was added, resulting in a four-way ANOVA. Both oculomotor tasks consisted of the IV condition (overlap vs. gap) and direction (left vs. right) and were each analyzed using three-way ANOVAs. Alpha level was 5% throughout. Due to the overall amount of data, correlation analyzes were restricted only to the most informative tests with respect to the central research questions, and interactions were in some cases only reported when significant.

3. Results

3.1. Reading task

The mean reading rate (based on oculomotor data) increased from 66 words per minute (SD = 28.08) in the second grade to 103 words per minute (SD = 16.92) in the fourth grade, $t(20) = 7.02$, $p < .001$, representing a 36% increase of speed. At the same time, text comprehension (in% relative to maximum scores) increased from 79.95% (SD = 11.24) in the second to 84.84% (SD = 6.29) in the fourth grade, $t(20) = 2.35$, $p = .029$.

Interindividual variability in reading rate decreased, $t(20) = 8.11$, $p < .001$. The mean number of fixations per sentence decreased from 35.92 (SD = 12.55) to 27.17 (SD = 8.76), $t(20) = 2.84$, $p = .01$. Mean fixation durations also decreased substantially from 358 ms (SD = 56) to 297 ms (SD = 58), $t(20) = 4.27$, $p < .001$. The overall percentage of regressive saccades did not change significantly, with a mean of 36% (SD = 13) in second and 42% (SD = 12) in fourth grade, $t < 1$. Mean saccade amplitude in reading direction (progressions) increased from 5.27 letter units (SD = 1.29) in grade 2 to 6.31 letter units (SD = 1.51) in grade 4, $t(20) = 3.58$, $p = .002$. Mean regression amplitudes remained unchanged (overall M = 5-letter units), $t < 1$.

Initial fixation durations on target words were not significantly affected by word frequency, $F(1, 20) = 2.92$, $p = .103$, or word length, $F(1, 20) = 1.95$, $p = .178$, but by grade, $F(1, 20) = 15.57$, $p = .001$. Table 1 depicts all means and standard errors, showing that initial fixation durations were reduced by 77% from second to fourth grade. Grade did not interact with either word frequency or word length, all $F < 1$.

Gaze durations on target words were significantly affected by word frequency, $F(1, 20) = 8.89$, $p = .007$, word length, $F(1, 20) = 112.88$, $p < .001$, and grade, $F(1, 20) = 48.65$, $p < .001$ (see Table 1). Frequent as well as short words were read with shorter gaze durations than infrequent and long words, respectively. Gaze durations were reduced by 37% from second to fourth grade. The word length effect was significantly reduced from second to fourth grade, $F(1, 20) = 21.72$, $p < .001$, whereas the effect of word frequency

Table 1
Mean initial fixation durations, gaze durations and total reading times (in ms) in the reading task as a function of word length (short vs. long), word frequency (high vs. low), and grade (second vs. first). Standard errors are given in parantheses.

Length	Frequency	Parameter	Second grade	Fourth grade
Short words	High frequency	Initial fixation duration	360 (18)	274 (16)
		Gaze duration	458 (33)	307 (13)
		Total reading time	640 (57)	404 (20)
	Low frequency	Initial fixation duration	365 (20)	279 (20)
		Gaze duration	479 (25)	336 (20)
		Total reading time	743 (63)	428 (21)
Long words	High frequency	Initial fixation duration	357 (20)	280 (19)
		Gaze duration	695 (40)	383 (21)
		Total reading time	927 (102)	548 (34)
	Low frequency	Initial fixation duration	384 (22)	294 (20)
		Gaze duration	614 (49)	395 (23)
		Total reading time	1150 (150)	543 (34)

only tended to be smaller, without resulting in a statistically significant difference, $F(1, 20) = 1.77$, $p > .05$. The interindividual variability of gaze durations decreased significantly, $t(20) = 4.20$, $p < .001$.

Total reading times on target words also include revisiting fixations (either following regressive or progressive re-reading saccades) after the word has been left for the first time. They were significantly affected by word frequency, $F(1, 20) = 12.32$, $p = .002$, word length, $F(1, 20) = 31.07$, $p < .001$, and grade, $F(1, 20) = 24.13$, $p < .001$ (see Table 1), with a data pattern comparable to that of the gaze durations. Overall, total reading time was reduced by 44% from second to fourth grade. The word length effect on total reading times also significantly decreased from second to fourth grade, $F(1, 20) = 10.38$, $p = .004$, as did the effect of word frequency, $F(1, 20) = 18.88$, $p < .001$. The interindividual variability of the total reading times was also significantly reduced, $t(20) = 8.04$, $p < .001$.

Refixations of target words were more frequent in the second grade than in the fourth grade, $F(1, 20) = 6.19$, $p = .022$. An analysis of refixation time, defined as gaze durations minus initial fixation durations, revealed a corresponding effect of grade, $F(1, 20) = 46.44$, $p < .001$, and of word length, $F(1, 20) = 165.64$, $p < .001$, indicating that refixation time was reduced from second to fourth grade and for shorter compared to longer words. The interaction of grade and word length was significant, too, $F(1, 20) = 48.19$, $p < .001$, indicating that the effect of word length on refixation time was greater in the second compared to the fourth grade. All remaining effects and interactions were not significant, all $p > .10$. An analysis of re-reading time, defined as the total reading time minus gaze duration, revealed a marginally significant effect of word frequency, $F(1, 20) = 3.34$, $p = .083$, and significant effects of word length, $F(1, 20) = 8.72$, $p = .006$, and grade, $F(1, 20) = 9.43$, $p = .006$, indicating that re-reading was reduced from second to fourth grade and tended to be longer for long and infrequent words. The interaction of grade and word frequency was significant, too, $F(1, 20) = 6.45$, $p = .019$, indicating that the effect of word frequency on re-reading was greater in the second compared to the fourth grade. All remaining interactions were not significant, all $p > .05$.

Overall, initial saccade landing positions on target words were located 0.28 letter units further into the word from second to fourth grade, $F(1, 20) = 20.61$, $p < .001$. Separate analyzes were conducted to rule out the possibility that this difference is due to differences in saccade launch site. Fig. 1 depicts the mean initial landing position on target words of either 4–5 or 6–7 letter length for different launch sites, computed as the distance (in letter units) from the beginning of the fixated word (e.g., Radach & McConkie, 1998). Note that irrespective of launch site distance, second graders fixated target words more towards their beginning as compared to fourth graders. As expected, saccades were directed further into longer compared to shorter words, $F(1, 20) = 87.04$, $p < .001$, and no significant differences with respect to word frequency were observed, $F(1, 20) = 2.91$, $p = .104$.

Initial fixations closer to the word beginning were typically of shorter duration (i.e., the inverted optimal viewing position effect) and more likely followed by a refixation of the same word (i.e., the optimal viewing position effect) compared with initial fixations located closer to the middle of the word (see Vitu, McConkie, Kerr, & O'Regan, 2001). Fig. 2 indicates that these effects were comparable across grade levels.

Word skipping was more frequent for short (18%) as compared to long target words (5%), $F(1, 20) = 47.67$, $p < .001$, but there were no significant main effects of grade or frequency and no significant interactions, all $p > .05$. Mean absolute fixation disparity ($M = 2.1$ letters, 57% crossed fixations) did not differ between grade levels, $t < 1$.

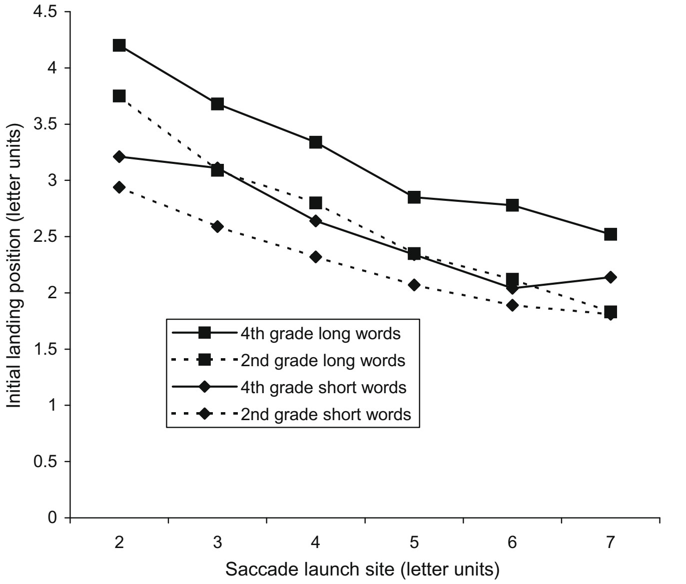

Fig. 1. Mean initial landing position (in letter units) in the target words of the reading task for short (4–5) and long (6–7) target words as a function of the saccade launch site distance (in letter units) and grade (second vs. fourth). For example, a launch site of 2 means that the saccade into the target word started two character spaces in front of the word beginning.

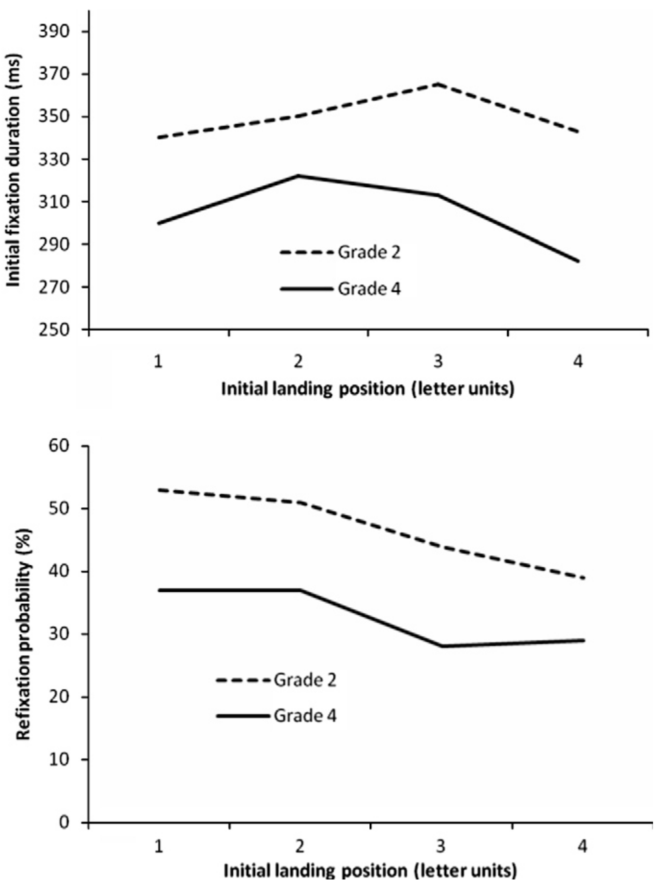

Fig. 2. Refixation probability (lower panel) and mean fixation duration (upper panel) as a function of initial landing position in 6–7-letter words. Note that due to a lack of sufficient data initial landing positions 5–7 are not displayed in the graphs.

3.2. Pro- and anti-saccades

In prosaccades we found a significant reduction of latencies from second to fourth grade, $F(1, 20) = 6.41$, $p = .020$ (see Table 2). Gap latencies were lower than overlap latencies, $F(1, 20) = 33.74$, $p < .001$, representing a typical gap effect. No significant effects of gaze direction were observed, $F < 1$. The gap effect did not significantly differ between grade levels, $F(1, 20) = 2.57$, $p = .124$.

An analysis of the percentage of express saccades, defined as saccade latencies between 70–120 ms (Fischer & Weber, 1993), revealed a marginally significant tendency towards an increase at grade 4 (43.8%) compared with grade 2 (34.5%), $t(20) = 1.84$, $p = .086$. Fig. 3 depicts the distribution of saccade latencies in the prosaccade gap condition across grade levels and does not indicate a clear bimodal distribution typically associated with the notion of express saccades (Fischer & Weber, 1993).

A comparison of saccade amplitudes in the prosaccade task revealed no significant effects of grade, $F < 1$, and condition (gap vs. overlap), $F(1, 20) = 2.71$, $p = .116$, but of direction, $F(1, 20) = 7.21$, $p = .014$ (6.13° rightwards vs. 6.35° leftwards).

In antisaccades we found a significant reduction of saccade latencies from second to fourth grade, too, $F(1, 20) = 8.69$, $p = .010$ (see Table 2). The gap effect was significant, $F(1, 20) = 49.43$, $p < .001$, but did not differ between grade levels, $F < 1$. No significant effects of direction were observed, $F(1, 20) = 2.24$, $p = .16$.

Students made more erroneous prosaccades in second grade ($M = 52.86\%$, $SD = 32.31$) than in fourth grade ($M = 38.71\%$, $SD = 15.89$), $t(20) = 3.31$, $p < .01$. Due to high error rates in second grade, a comparison of saccade amplitudes in correct trials was not conducted.

Table 2
Mean saccade latencies (in ms) in the pro- and anti-saccade tasks as a function of condition (gap vs. overlap) and grade (second vs. fourth). Standard errors are given in parantheses.

			Second grade	Fourth grade
Prosaccades	Gap condition	Leftwards	162 (39)	147 (31)
		Rightwards	159 (36)	148 (26)
	Overlap condition	Leftwards	230 (89)	200 (53)
		Rightwards	221 (72)	196 (52)
Antisaccades	Gap condition	Leftwards	346 (176)	280 (78)
		Rightwards	332 (193)	250 (75)
	Overlap condition	Leftwards	398 (112)	350 (65)
		Rightwards	394 (111)	339 (80)

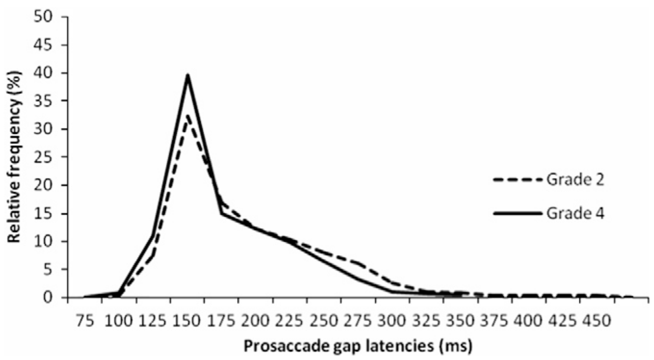

Fig. 3. Frequency distribution of prosaccade gap latencies across grade levels.

3.3. Naming task

Table 3 summarizes latencies in the naming task. Overall, they significantly decreased from second to fourth grade, $F(1, 20) = 14.41$, $p = .001$. Latencies were shorter for high compared to low frequency words, $F(1, 20) = 53.96$, $p < .001$, for short compared to long words, $F(1, 20) = 19.11$, $p < .001$ and for words compared to pictures, $F(1, 20) = 14.53$, $p = .001$. The frequency effect was less pronounced in the fourth compared to the second grade, $F(1, 20) = 22.56$, $p < .001$. The same holds for the word length effect, $F(1, 20) = 18.02$, $p < .001$. Word naming latency did not decrease more pronounced from second to fourth grade than picture naming latency, $F(1, 20) = 1.60$, $p = .22$. Word frequency effects were more pronounced for pictures than for words, $F(1, 20) = 17.16$, $p = .001$. All remaining interactions were non-significant. Overall, mean naming latencies for all words decreased from second to fourth grade from 1058 ms to 900 ms (15%), and for pictures from 1166 ms to 1064 ms (9%). Interindividual variability in the naming task significantly decreased, $t(20) = 2.72$, $p < .05$.

Picture naming errors amounted to 1.23% ($SD = 0.87$) in the second and 0.88% ($SD = 0.65$) in the fourth grade, $t(20) = 1.68$, $p = .11$. The error rate for word naming was 0.57% ($SD = 0.80$) in the second and 0.06% ($SD = 0.15$) in the fourth grade, $t(20) = 2.98$, $p < .01$.

In the second grade, word naming latencies and total reading times of the same words in the reading task were highly correlated in the respective word category, $r(20) = .61$ for short low frequency words, $r(20) = .57$ for short high frequency words, $r(20) = .80$ for long low frequency words and $r(20) = .90$ for long high frequency words, all $p < .01$. In fourth grade, these correlations were lower, with $r(20) = .46$, $r(20) = .26$, $r(20) = .60$, and $r(20) = .58$, respectively, all $p < .06$. A similar pattern was found for gaze durations, with slightly lower correlations for the longer words.

To address the question whether word naming latencies reflect the processing time of a word in normal reading, we computed mean word naming latencies along with gaze durations and total reading times from the reading task for each item, averaged across subjects. When correlating these item-based values, we found that for the second graders word naming latency was significantly correlated with gaze durations ($r(79) = .544$) and total reading times ($r(79) = .619$), both $p < .001$. However, in the fourth grade no significant correlations were observed, $r(79) = .147$, $p = .19$, and $r(79) = .118$, $p = .30$, respectively.

3.4. Reading and cognitive ability tests

Table 4 summarizes the results from the standardized reading and cognitive ability tests (SLS, WLLP, SLRT, CPM). No child scored below an IQ of 85 in either of the two assessments (CPM), and IQ

Table 3
Mean naming latencies (in ms) in the word and picture naming tasks as a function of (corresponding) word length (short vs. long), word frequency (high vs. low), and grade (second vs. fourth).

			Second grade	Fourth grade
Word naming	Short words	High frequency	940 (45)	858 (25)
		Low frequency	1068 (72)	930 (38)
	Long words	High frequency	1061 (72)	895 (33)
		Low frequency	1161 (96)	917 (37)
Picture naming	Short words	High frequency	995 (35)	975 (29)
		Low frequency	1234 (42)	1114 (36)
	Long words	High frequency	1137 (47)	1048 (36)
		Low frequency	1295 (46)	1115 (48)

Table 4
Results of the reading and cognitive ability tests across grade levels.

Test	Parameter	Grade 2 *M* (SD)	Grade 4 *M* (SD)	Statistics *t*(20); *p*
SLS	Correctly solved items	38 (10)	53 (11)	13.0; <.001
	Standardized reading quotient (*M* = 100)	115 (18)	110 (18)	2.4; <.05
WLLP	Correctly solved items	73 (19)	118 (16)	12.5; <.001
	Standardized percentiles	50 (26)	57 (30)	1.5; >.05
SLRT	Noun list reading time (s)	23 (6)	15 (3)	7.5; <.001
	Dissimilar pseudoword list reading (s)	51 (15)	37 (11)	7.3; <.001
	Similar pseudoword list reading (s)	47 (16)	30 (9)	6.3; <.001
	Text reading (Standardized percentiles)	73 (25)	73 (26)	<1; >.05
CPM	Correctly solved items	32.5 (3.7)	33.9 (2.2)	2.6; .018
	Standardized IQ	120 (19)	112 (14)	2.8; <.01

was not correlated with any measure from the reading task, the naming task, and the oculomotor tasks, all $p > .05$.

Retest reliability of the reading tests amounted to $r(20) = .87$, $p < .001$ for the SLS and $r(20) = .56$, $p = .008$ for the WLLP. Interindividual variability in the SLS did not differ significantly between grade levels, $t(20) < 1$. Raw values of the WLLP and SLS correlated with $r(20) = .50$ in the second grade, $p = .02$, and with $r(20) = .87$ in the fourth grade, $p < .001$. In the fourth grade, the SLS correlated higher with sentence reading times in the reading task, $r(20) = -.77$, $p < .001$ than did the WLLP, $r(20) = -.56$, $p = .008$. This pattern also holds for the second grade, and is also present with respect to other parameters in the reading task, suggesting an advantage of the SLS over the WLLP as a reading test.

3.5. Further cognitive tests at fourth grade level

An analysis of the two measures of *phonological awareness* revealed that only the initial component change task significantly correlated with measures of reading performance, such as reading of nouns ($r(20) = .61$, $p < .01$) and similar pseudowords ($r(20) = .45$, $p < .05$) in the SLRT as well as word naming latencies ($r(20) = .71$, $p < .001$). Note that these measures are all based on reading of single, unconnected words. In contrast, phonological awareness did not significantly correlate with gaze durations ($r(20) = .36$, $p > .05$) or total reading times ($r(20) = .24$, $p > .05$) on target words in the reading task or with SLS scores ($r(20) = -.38$, $p > .05$), with all these measures involving reading of complete sentences.

From the *R.A.N. measures* (as a potential alternative to the naming task), only the digit ($r(20) = -.49$, $p < .05$) and the letter task ($r(20) = -.54$, $p < .05$) significantly correlated with reading abilities (SLS scores), whereas pictures ($r(20) = -.36$, $p > .05$) and words ($r(20) = -.17$, $p > .05$) yielded no significant correlations. Especially the letter version yielded the highest correlations throughout all measures assessed by the reading tests, and also correlated with gaze durations ($r(20) = .48$, $p < .05$) and total reading times ($r(20) = .66$, $p < .01$) on the target words in the reading task. The letter version of the R.A.N. also correlated significantly with the initial component change task for the assessment of phonological awareness ($r(20) = .81$, $p < .05$). All versions of the R.A.N. were correlated among themselves, $.57 < r(20) < .76$, all $p < .01$. Interestingly, the picture version of the R.A.N. did not correlate with latencies in the picture naming task, $r(20) = .14$, $p = .53$, and the word R.A.N. did also not correlate with latencies in the word naming task, $r(20) = .13$, $p = .56$.

3.6. Predicting reading performance using oculomotor and linguistic measures

For the assessment of the importance of oculomotor and linguistic components for the development of reading skills, we selected total reading times on the target words (reading task) in the fourth grade as the criterion in a linear regression analysis. As predictors, we used variables from the second grade assessment. Two oculomotor predictors were chosen: first, a composite measure of oculomotor speed, based on mean (equally weighted) pro- and anti-saccade latencies, and second, the number of erroneous prosaccades in the antisaccade task. As an overall linguistic predictor, we used the mean word and picture naming latencies. We restricted the regression analysis to the test of the main research question, that is, how oculomotor and linguistic predictors contribute to the development of efficient dynamic reading, to avoid an accumulation of pseudo-significant results. As a result, we found no significant influence of the oculomotor variables ($\beta(20) = -.23$, $p = .29$ and $\beta(20) = .04$, $p = .84$, respectively), but a substantial effect of the linguistic predictor ($\beta(20) = .68$, $p = .004$), overall $R = .63$ (see Fig. 4 for individual scatterplots). Further post hoc analyzes revealed that substituting word naming performance for overall naming performance as a predictor did not change the pattern of results. The same was true for substituting overlap prosaccade latencies or the percentage of express saccades in prosaccade gap trials for oculomotor speed as a predictor. Finally, using initial fixation durations as an alternative criterion resulted in the same data pattern, too. A closer look at the raw data revealed that none of the oculomotor variables correlated significantly with any measure of reading skill, including mean and initial fixation durations, regardless of the grade level, all $p > .10$.

4. Discussion

The present longitudinal study examined the normal development of primary school children's eye movements during sentence reading. We added oculomotor and picture/word naming tasks to assess the contribution of oculomotor and linguistic skills to reading development. We utilized a common item pool for both the reading and the naming task to determine whether single item processing generalizes to normal reading of words in context. Additional assessments included tests of phonological awareness, rapid automatized naming, and intelligence. Standardized reading tests indicated that the present sample is representative of normally developing school children.

Children were tested at second and fourth grade level using identical material, allowing for meaningful comparisons between data sets. This benefit was traded against possible sequence effects due to a second encounter with the material, but this drawback was estimated as being a comparatively minor issue given the long inter-test interval, even though word repetition effects might well be very long lasting. It should be noted that prior longitudinal studies of children's eye movements in reading (e.g., McConkie et al., 1991) utilized age-appropriate reading materials which probably differed in a mixture of variables on the letter, word, and sentence level of processing, making changes in eye movement patterns difficult to interpret.

4.1. Reading task

In the oral reading task, the mean reading rate, which can be regarded as a measure of the overall efficiency gain from second to fourth grade, increased by about 36%. This did not go hand in hand with adverse effects on comprehension assessed within the same task. This efficiency increase is backed by the data from the reading

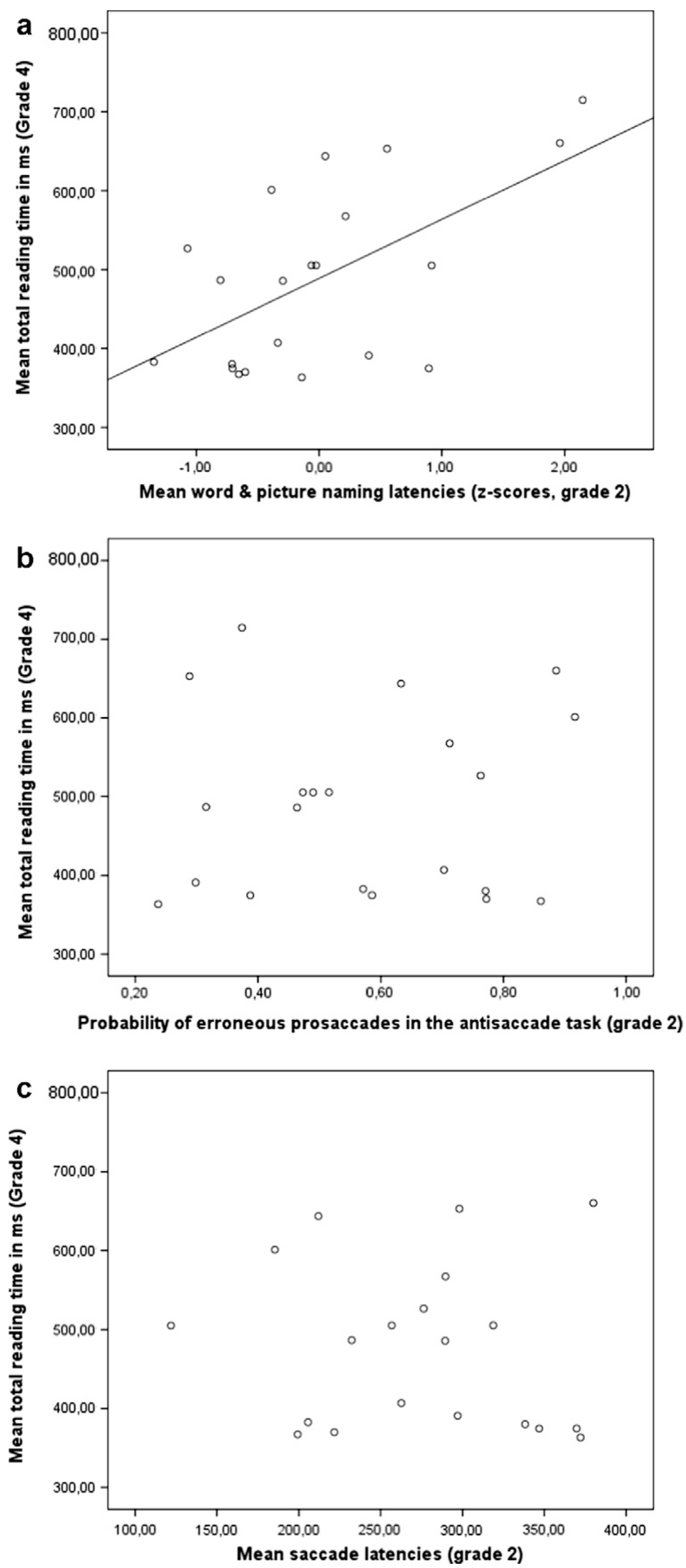

Fig. 4. Scatterplots of the correlation of the three predictors oculomotor speed, erroneous prosaccades in the antisaccade task, and naming latencies (grade 2) with the criterion total reading time (grade 4).

screening (SLS), where children completed 39.8% more sentences in the fourth compared to the second grade. Interestingly, the mean number of fixations for reading a sentence only decreased by about 25%. This decrease is smaller than the decrease in reading rate, probably due to the concurrent decrease of the mean fixation durations from second to fourth grade of about 50 ms per fixation. These data are remarkably similar to those from previous studies in English language (Buswell, 1922; Taylor et al., 1960; Rayner, 1985), despite any specificities of the German language (regular orthography, less lexical ambiguity, productive compounding etc.) or possible educational differences.

Gaze durations and total reading times on the target words suggested an even more pronounced efficiency gain (up to 44%), but these measures only refer to specific nouns, and therefore seem not suited for an assessment of overall reading efficiency. On the other hand, word naming latencies of the same items *under*estimated the magnitude of the development, with an improvement of only 15%. This underestimation might result from a task-specific lack of many cognitive processes (e.g., semantic processing, syntactic integration, and working memory processing) important for reading and understanding whole sentences.

In line with previous findings (McConkie et al., 1991), an analysis of interindividual variability in the reading task revealed a substantial decrease of variability from second to fourth grade with respect to reading rate, gaze durations, and total reading times. This likely reflects substantial individual differences in skills around and after school enrollment.

The oculomotor reading data provide further clues regarding underlying cognitive mechanisms of the efficiency gain. The refixation rates revealed that second graders tended to fixate each word more often before leaving to the next word. This goes along with shorter saccade amplitudes. Interestingly, the overall percentage of regressions was not increased for the younger readers (although the overall level was quite high, probably due to specifics of the German language). This indicates that their greater difficulty with decoding text is compensated for by a strategic increase of the amount of refixations to avoid an increase of regressive saccades back to previously inspected text. This also explains why previous studies did not find markedly reduced regression rates during the first years in school (Buswell, 1922; Taylor et al., 1960; McConkie et al., 1991; Rayner, 1985). The idea of such a "careful" reading strategy associated with a higher refixation probability was already introduced by O'Regan (1992). In his "strategy-tactics" theory of oculomotor control, he distinguished between careful and risky reading, depending upon the reader's goals. Careful reading implies that whenever an initial fixation in a word is not located at an optimal position for word decoding, the probability of triggering a refixation is higher. The present data suggest that second graders are more careful readers than fourth graders. A similar tendency of higher refixation rates was also found for word list reading in dyslexic children compared to a group of healthy individuals (De Luca, Borrelli, Judica, Spinelli, & Zoccolotti, 2002; Hutzler & Wimmer, 2004), suggesting that high refixation rates can be regarded as a general characteristic for less developed reading skills.

Additionally, we found that second graders initially directed their eyes more to the beginning of a word as compared to fourth graders, which might be a byproduct of the refixation strategy. This observation also renders it unlikely that the difference in refixation rates only reflects online processing difficulties with the refixated word (Engbert et al., 2005; Reichle et al., 2003), since initial landing positions usually remain largely unaffected by current linguistic processing demands. Instead, it is in line with previous findings showing that careful reading strategies can be reflected in shifts of initial fixation positions towards the word beginning (e.g., Radach, Huestegge, & Reilly, 2008). If a reader plans to refixate a word in advance, it is efficient to initially fixate the word beginning, since the word end will be decoded with another fixation. Previous research demonstrated that refixations are indeed less likely when the initial fixation is located near the word center, while fixation durations are longest towards the word center (e.g., Vitu, McConkie, Kerr, & O'Regan, 2001). Fourth graders tended to decode words more often with only one fixation, which in turn was located near the middle to maximize letter visibility. Although this strategic shift towards the word center slightly increases fixation durations (see Fig. 2), this drawback is negligible compared with the overall increase of gaze durations caused by additional refixations. One might argue that the difference in initial landing positions between grade levels could result from the fact that second graders were probably less likely to skip shorter words, and hence the proportion of longer words (within the short word category) was greater at second grade level. However, Fig. 1 clearly shows that the same results were obtained for long words, which are rarely skipped. A shift of initial landing positions was not present in the study of McConkie et al. (1991), probably due to the use of different text material and individuals in their landing position analyzes across grade levels, which might have concealed important effects.

Overall, we observed classic word length and frequency effects on total reading times and, except for an only marginally significant frequency effect, on gaze durations. Both effects were reduced in fourth grade. A closer inspection revealed that especially total reading times on infrequent target words were reduced from second to fourth grade, which might be explained in terms of a more automatic lexical access in fourth graders specifically for infrequent words. It is important to note that all infrequent words were known to the second graders, as suggested by the low error rates of <1.5% in the picture naming task, ruling out the possibility that second graders only looked at some of these words for a long time simply because they did not know them.

Note that we observed a decrease of word processing time from second to fourth grade in initial fixation durations, refixation time, and re-reading time, indicating that performance is not selectively enhanced for early or late word processing stages, but rather extends across the whole process of word decoding and integration. At first sight it appears puzzling why the decrease of re-reading times did not go hand in hand with a marked decrease in the proportion of regressive saccades. However, a post hoc visualization of the eye movement data revealed that some second graders tended to reread whole parts of sentences after they had been read for the first time, probably resulting from more serious comprehension problems on sentence level. Thus a more targeted and effective mode of re-inspecting previously read text at grade 4 may well explain why the decrease in re-reading times did not propagate onto regression rates.

4.2. Oculomotor tasks

Based on the assumption that dynamic text reading relies on both linguistic processing and effective eye guidance, the oculomotor tasks were administered to assess the development of basic oculomotor skills. In line with previous research (Klein, 2001), saccades were initiated faster in fourth as compared to second grade (between 8% and 17%, depending on task and condition). Saccade amplitudes as well as attentional and/or fixational disengagement skills (as measured by the gap effect) remained unaffected. However, the amount of erroneous prosaccades in the antisaccade task decreased, indicating a more effective voluntary control of saccades.

4.3. Naming task

In the word/picture naming task, participants named the same items as in the reading task. Since it was always conducted after the reading task, it is likely that items were overall somewhat easier to process, but this general effect should not compromise the interpretation of the results. The error rate for pictures revealed no significant differences between grade levels. In word naming, however, fewer errors occurred at fourth grade level despite the low overall error rate (<1%), suggesting a stronger developmental trajectory for word naming. This was to be expected since object naming is already highly trained in pre-school age. In line with this observation, mean latencies decreased by about 9% for pictures and about 15% for words, although this difference in decrease was not statistically significant. Similar to the reading task, we found significant word length and frequency effects, which decreased from second to fourth grade, indicating speeded lexical access in fourth grade (especially for infrequent words).

Overall, picture naming was slower than word naming, since the former necessarily involves lexical-semantic encoding, during which lexical units are selected from a set of candidates (Levelt, Roelofs, & Meyer, 1999). In contrast, word naming requires phonological retrieval without lemma selection (Coltheart et al., 2001).

A comparison of word naming and target word reading in the reading task revealed that naming times were generally longer than total reading times on the same items. This might be due to the more dynamic cognitive processing in sentence reading, where the eyes do not remain on a word until it is finally processed (see, e.g., Kliegl, Nuthmann, & Engbert, 2006; Reichle et al., 2003), and where the context provides word decoding cues. Intraindividual correlations between word naming latencies and total reading times were only moderate, with $.57 < r < .90$ for second graders, and $.26 < r < .60$ for fourth graders. On the one hand, the correlations show that single-word naming is to some extent able to predict word reading times in oral reading (see Schilling et al., 1998), especially for long words. On the other hand, this correlation is far from being perfect. A possible explanation for this observation might be that since in word naming a stimulus may activate semantic knowledge, this is not a necessary prerequisite, and no syntactic integration processes are involved. It is interesting to note that correlations are substantially lower in fourth compared to second grade. Possibly, the driving force behind the correlation is the grapheme–phoneme conversion that to some extent is common to both task demands, and in fourth grade this process consumes less time, leading to a greater relative contribution of other linguistic processes while reading sentences.

This line of reasoning is further supported by the item-based correlations of word naming latencies with gaze durations and total reading times. Results indicated that in second grade naming latencies significantly correlated with processing time on words during normal reading, whereas this was no longer the case in fourth grade. Obviously, for a more developed reader single-word naming latencies do not reliably reflect word processing in normal reading. Since previous reports of substantial correlations solely relied on intraindividual correlation analyzes, they likely overestimated the similarities of naming and reading (e.g., Schilling et al., 1998). As outlined above, the common variance in second grade probably reflects grapheme–phoneme conversion processes, whereas in fourth grade this ability is already highly developed, so that both tasks share fewer common processes, in turn leading to a substantial reduction of correlations. Since oral reading shares pronunciation processes with naming, the link between naming and silent reading should be even weaker. In sum, this corroborates previous claims that reading cannot sufficiently be studied in terms of context-free word decoding (e.g., Gough & Tunmer, 1986).

4.4. Prediction of reading performance

A final important aim of the present study was to estimate to what extent oculomotor and linguistic skills in second grade are relevant for the prediction of reading performance in fourth grade. Two predictors of oculomotor skills were chosen that significantly differed between grade levels. First, oculomotor speed was represented by the mean latencies averaged across pro- and anti-saccades and over gap and overlap trials in the second grade. Second, the number of erroneous prosaccades in the antisaccade task served as a measure of voluntary eye movement control (Findlay & Walker, 1999). Previous research suggested that both bottom-up and top-down control processes should be relevant for eye guidance in reading (Biscaldi et al., 2000; Fischer & Weber, 1990). Since naming errors were rather infrequent, we only used average naming speed as a predictor, representing linguistic skills in second grade. Naming speed was averaged over pictures and words because both item types require important processes involved in oral reading, such as lexical access, semantic processing, and pronunciation (see above). Since semantic processing is not necessary in word naming, but can nevertheless occur, we did not treat word and picture naming as separate predictors, which would have suggested that the underlying processes can be clearly separated. As a criterion, we chose to predict total reading times in the fourth grade, which should reflect all important reading-related processes. Although the underlying model of two separate sources of influence (linguistic vs. oculomotor) is certainly oversimplified in that it does not account for the complex interaction of both factors in a complex task like reading, it should be useful for at least broadly assessing the relative contribution of these factors. The results clearly indicated that only linguistic, not oculomotor skills were the driving force behind the acquisition of normal oral reading skills. Most likely, the oculomotor system is already well prepared for the task of guiding the eyes through the text at second grade level. This conclusion is in harmony with the finding by McConkie et al. (1991) that the basic oculomotor metrics of saccade landing positions in reading are already well developed at the end of first grade. Although oculomotor skills clearly develop, the limiting factor for developing reading skills seems to be linguistic word decoding, involving orthographic processing, grapheme–phoneme conversion, lexical access, and semantic analysis, among others (see, e.g., Cunningham & Stanovich, 1997; Storch & Whitehurst, 2002; Wagner & Torgesen, 1987; Whitehurst & Lonigan, 1998, for more in-depth approaches to the prediction of linguistic skills). Note that this conclusion need not necessarily transfer to dyslexic readers, although the lack of any significant correlation between oculomotor parameters and reading performance appears discouraging for the view that oculomotor skills may hold the key to severe reading problems, especially since the specific task demands in basic oculomotor tasks and reading differ quite substantially.

One might argue that the present sample size is rather low for reliable correlation analyzes. For example, the statistical power to detect a positive correlation between grade 2 naming latencies and grade 4 mean total reading times of $r = .59$ (which we observed) amounts to $1 - \beta = 90\%$. While this value appears reasonable, effects of oculomotor performance might be more subtle, requiring more subjects for comparable statistical power. However, given that none of the various independent basic oculomotor measures correlated with any of the reading-related measures (at any grade level), and that the regression analysis remained unaffected despite the use of different specific predictors and criteria, it appears unlikely that a lack of statistical power is a reasonable alternative explanation of the data.

Despite the irrelevance of basic oculomotor skills for normal reading development, it remains possible that more complex (reading-related) oculomotor routines play an important role. For example, word frequency effects on gaze durations did not vary between grades, but the effect of frequency on naming latencies was significantly reduced, suggesting that eye movement control in reading does not exactly follow the same developmental trajectory as linguistic performance alone. Additionally, the word length effect on fixation times (usually attributed to low-level visuomotor variables and not linguistic processing) differed between grades, suggesting a development of reading-related low-level oculomotor routines.

4.5. Additional psychometric assessments

Phonological awareness is discussed to be among the best predictors for reading ability (see Scarborough, 1998, for a discussion). One of the two tasks implemented here indeed correlated highly with reading performance, but only in reading tasks where lists of unrelated single words had to be read, like noun and pseudoword reading in the SLRT as well as word naming. However, phonological awareness did not correlate with measures that imply sentence reading, such as gaze durations and total reading times in the reading task or the SLS scores. It is therefore possible that in normal children (especially those reading in a regular orthography) good phonological awareness only facilitates performance in tasks where unpredictable words are pronounced as fast as possible, whereas in sentence reading, contextual information may be used to preactivate entries in the lexicon. This might facilitate phonological processing of words before they are entered, leading to a levelling of individual differences. In dyslexics (and probably also in the more unexperienced second graders), however, phonological awareness might pose a bottleneck so that sentence reading performance is affected (e.g., Snowling, 2000).

Unlike the word and picture naming task, *R.A.N.* consists of lists of only few pre-specified, but constantly reappearing items like letters, pictures, words, or digits (Denckla & Rudel, 1976). Especially the letter version correlated highly with measures related to sentence reading performance, including SLS scores as well as gaze durations and total reading times in the reading task (see Holland, McIntosh, & Huffman, 2004, for similar results with respect to word reading). Using structural equation modeling techniques, Neuhaus and Swank (2002) convincingly demonstrated that the letter version of R.A.N. forms a basic reading test, comprising fundamental prerequisites of normal oral reading such as phonological encoding, orthographic recognition and articulation. Interestingly, R.A.N. was not correlated with naming latencies, implying that the underlying task demands are actually quite different.

5. Conclusions

Adequately developed literacy skills are of great importance, and considerable scientific effort has been put into the search for causes and remedies for impaired reading development, which has a prevalence of 5–17% (Plume & Warnke, 2007; Shaywitz & Shaywitz, 2001). Although recent research advances have been impressive, no consistent general explanation of developmental dyslexia has emerged so far (see, e.g., Ramus et al., 2003; Vellutino, Fletcher, Snowling, & Scanlon, 2004, for discussions of competing accounts). One reason for this unsatisfactory state of affairs may be that the critically important investment into research on reading disabilities has not been backed by a thorough study of the development of normal reading abilities in some research areas, specifically regarding the study of the developmental dynamics of eye movements in reading. In our view, the present study (along with other recent publications, see, e.g., Feng et al., 2009; Joseph et al., 2009) provides a solid data base demonstrating how successful reading evolves and to what extent healthy development may vary, which is a prerequisite to understanding and compensating developmental delays and disabilities. More specifically, our results render it unlikely that treatments of reading problems solely based on basic oculomotor training or on single-word reading (instead of sentence reading) will substantially enhance reading skills.

In sum, the present study suggests that slower reading at second grade level was due to a global refixation strategy, with initial saccade landing positions located closer to word beginnings. In fourth grade, children exhibited improved lexical access. Despite a clear development of basic oculomotor skills, these played no role as a driving force behind reading development. Instead, linguistic skills (as indicated by naming latencies without the need for eye movements) in second grade significantly predicted oral reading performance in fourth grade. However, naming latencies predicted reading times of the same items in the reading task only at second grade level, but no longer at fourth grade level, indicating that naming latencies in more developed readers do not reliably reflect processes relevant for word processing during sentence reading.

Acknowledgments

We thank Francoise Vitu and two anonymous reviewers for comments on earlier drafts of this article.

References

Biscaldi, M., Fischer, B., & Hartnegg, K. (2000). Voluntary saccade control in dyslexia. *Perception, 29*, 509–521.

Buswell, G. T. (1922). *Fundamental reading habits, a study of their development*. Chicago: Chicago University Press.

Coltheart, M., Rastle, K., Perry, P., Langdon, R., & Ziegler, J. (2001). DRC: A dual route cascaded model of visual word recognition and reading aloud. *Psychological Review, 108*, 204–256.

Cunningham, A. E., & Stanovich, K. E. (1997). Early reading acquisition and its relation to reading experience and ability 10 years later. *Developmental Psychology, 33*, 934–945.

De Luca, M., Borrelli, M., Judica, A., Spinelli, D., & Zoccolotti, P. (2002). Reading words and pseudowords: An eye movement study of developmental dyslexia. *Brain and Language, 80*(3), 617–626.

Denckla, M. B., & Rudel, G. R. (1976). Rapid automatized naming (R.A.N.): Dyslexia differentiated from other language disabilities. *Neuropsychologia, 14*, 471–479.

Engbert, R., Nuthmann, A., Richter, E. M., & Kliegl, R. (2005). SWIFT: A dynamical model of saccade generation during reading. *Psychological Review, 112*, 777–813.

EX, C. E. L. (1995). *German database. Release D25. Computer software*. Nijmegen: Centre for Lexical Information.

Feng, G. (2006). Eye movements as time-series random variables: A stochastic model of eye movement control in reading. *Cognitive Systems Research, 7*, 70–95.

Feng, G., Miller, K., Shu, H., & Zhang, H. (2009). Orthography and the development of reading processes: An eye-movement study of Chinese and English. *Child Development, 80*, 720–735.

Findlay, J. M., & Walker, R. (1999). A model of saccade generation based on parallel processing and competitive inhibition. *Behavioral and Brain Sciences, 22*, 661–674.

Fischer, B., & Weber, H. (1990). Saccadic reaction times of dyslexic and age-mached normal subjects. *Perception, 19*, 805–818.

Fischer, B., & Weber, H. (1993). Express saccades and visual attention. *Behavioral and Brain Sciences, 16*, 553–567.

Folk, J. R., & Morris, R. (1995). Multiple lexical codes in reading: Evidence from eye movements, naming time, and oral reading. *Journal of Experimental Psychology: Learning, Memory and Cognition, 21*, 1412–1429.

Gough, P. B., & Tunmer, W. E. (1986). Decoding, reading and reading disability. *Remedial and Special Education, 7*, 6–10.

Grainger, J. (2003). Moving eyes and reading words: How can a computational model combine the two? In J. Hyönä, R. Radach, & H. Deubel (Eds.), *The mind's eye: Cognitive and applied aspects of eye movement research* (pp. 457–470). Oxford: Elsevier.

Grainger, J., & Jacobs, A. M. (1996). Orthographic processing in visual word recognition: A multiple read-out model. *Psychological Review, 103*, 518–565.

Holland, J., McIntosh, D. E., & Huffman, L. (2004). The role of phonological awareness, rapid automatized naming, and orthographic processing in word reading. *Journal of Psycho-educational Assessment, 22*, 233–260.

Hutzler, F., & Wimmer, H. (2004). Eye movements of dyslexic children when reading in a regular orthography. *Brain and Language, 89*(1), 235–242.

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), 1–11.

Inhoff, A. W., Briihl, D., & Schwarz, J. (1996). Compound words effects differ in reading, on-line naming, and delayed naming tasks. *Memory & Cognition, 24*, 466–476.

Inhoff, A. W., & Topolski, R. (1994). Use of phonological codes during eye fixations in reading and in on-line and delayed naming tasks. *Journal of Memory and Language, 33*, 689–713.

Jacobs, A. M., & Grainger, J. (1994). Models of visual word recognition: Sampling the state of the art. *Journal of Experimental Psychology: Human Perception and Performance, 20*, 1311–1334.

Joseph, H. S. S. L., 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*, 2078–2086.

Kirk, R. E. (1990). *Statistics: An Introduction*. Chicago: Holt, Rinehart and Winston.

Klein, C. (2001). Development of prosaccade and antisaccade task performance in participants aged 6 to 26 years. *Psychophysiology, 38*, 179–189.

Klein, C., & Fischer, B. (2005). Instrumental and test–retest reliability of saccadic measures. *Biological Psychology, 68*, 201–213.

Kliegl, R., Nuthmann, A., & Engbert, R. (2006). Tracking the mind during reading: The influence of past, present, and future words on fixation durations. *Journal of Experimental Psychology: General, 135*, 12–35.

Kramer, A. F., Gonzales de Sather, J. C. M., & Cassavaugh, N. D. (2005). Development of attentional and oculomotor control. *Developmental Psychology, 41*, 760–772.

Küspert, P., & Schneider, W. (1998). *Würzburger Leise Leseprobe (WLLP). Handanweisung*. Göttingen: Hogrefe.

Landerl, K., Wimmer, H., & Moser, E. (2001). *SLRT. Salzburger Lese- und Rechtschreibtest. Verfahren zur Differentialdiagnose von Störungen des Lesens und Schreibens für die 1. bis 4. Schulstufe. Handbuch*. Bern: Hans Huber.

Levelt, W. J. M., Roelofs, A. P. A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences, 22*(1), 1–37.

Levin, H., & Buckler-Addis, A. (1979). *The eye-voice span*. Cambridge, MA: MIT Press.

Mayringer, H., & Wimmer, H. (2003). *SLS 1–4. Salzburger Lese-Screening für die Klassenstufen 1–4. Manual*. Bern: Hans Huber.

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

McDonald, S. A., Carpenter, R. H., & Shillcock, R. C. (2005). An anatomically constrained, stochastic model of eye movement control in reading. *Psychological Review, 112*, 814–840.

Neuhaus, G. F., & Swank, P. R. (2002). Understanding the relations between RAN letter subtest components and word reading in first-grade students. *Journal of Learning Disabilities, 35*, 158–174.

O'Regan, J. K. (1992). Optimal viewing position in words and the strategy-tactics theory of eye movements in reading. In K. Rayner (Ed.), *Eye movements and visual cognition: Scene perception and reading* (pp. 333–354). New York: Springer.

Plume, E., & Warnke, A. (2007). Definition, symptoms, prevalence and diagnosis of dyslexia. *Monatsschrift Kinderheilkunde, 155*, 322–327.

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.

Radach, R., & Kennedy, A. (2004). Theoretical perspectives on eye movements in reading: Past controversies, current deficits and an agenda for future research. *European Journal of Cognitive Psychology, 16*, 3–26.

Radach, R., & McConkie, G. W. (1998). Determinants of fixation positions in words during reading. In G. Underwood (Ed.), *Eye guidance in reading and scene perception* (pp. 77–101). Oxford, England: Elsevier.

Radach, R., Reilly, R., & Inhoff, A. W. (2006). Models of oculomotor control in reading: Towards a theoretical foundation of current debates. In R. van Gompel, M. Fischer, W. Murray, & R. Hill (Eds.), *Eye movements: A window on mind and brain*. Elsevier: Oxford.

Ramus, F., Rosen, S., Dakin, S. C., Day, B. L., Castellote, J. M., White, S., et al. (2003). Theories of developmental dyslexia: Insights from a multiple case study of dyslexic adults. *Brain, 126*, 841–865.

Raven, J., Raven, J. C., & Court, J. H. (1998a). *Manual for Raven's progressive matrices and vocabulary scales. Section 2: The coloured progressive matrices*. San Antonio, TX: Harcourt Assessment.

Raven, J. C., Raven, J., & Court, J. H. (1998b). *Raven's progressive matrices und vocabulary scales*. Frankfurt: Swets & Zeitlinger.

Raven, J., Raven, J. C., & Court, J. H. (2000). *Manual for Raven's progressive matrices and vocabulary scales. Section 3: The standard progressive matrices*. San Antonio, TX: Harcourt Assessment.

Rayner, K. (1979). Eye guidance in reading: Fixation locations within words. *Perception, 8*, 21–30.

Rayner, K. (1985). Visual selection in reading, picture perception, and visual search: A tutorial review. In H. Bouma & D. G. Bouwhuis (Eds.), *Attention and performance X: Control of language processes* (pp. 67–96). Hillsdale, NJ: Erlbaum.

Rayner, K. (1986). Eye movements and the perceptual span in beginning and skilled readers. *Journal of Experimental Child Psychology, 41*, S. 211–S. 236.

Rayner, K. (1998). Eye movements in reading and information processing: 20 Years of research. *Psychological Bulletin, 124*, 372–422.

Reichle, E. D., Rayner, K., & Pollatsek, A. (2003). The E–Z reader model of eye movement control in reading: Comparisons to other models. *Behavioral and Brain Sciences, 26*, 445–476.

Reilly, R., & Radach, R. (2006). Some empirical tests of an interactive activation model of eye movement control in reading. *Cognitive Systems Research, 7*, 34–55.

Rossion, B., & Pourtois, G. (2004). Revisiting Snodgrass and Vanderwart's object set: The role of surface detail in basic-level object recognition. *Perception, 33*, 217–236.

Scarborough, H. S. (1998). Predicting the future achievement of second graders with reading disabilities: Contributions of phonemic awareness, verbal memory, rapid serial naming, and IQ. *Annals of Dyslexia, 48*, 115–136.

Schilling, H. E. H., Rayner, K., & Chumbley, J. I. (1998). Comparing naming, lexical decision, and eye fixation times: Word frequency effects and individual differences. *Memory & Cognition, 26*, 1270–1281.

Shaywitz, S. E., & Shaywitz, B. A. (2001). The neurobiology of reading and dyslexia. *Focus on Basics, 5*, 11–15.

Snodgrass, J. G., & Vanderwart, M. (1980). A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology: Human Learning and Memory, 6*, 174–215.

Snowling, M. J. (2000). *Dyslexia*. Oxford: Blackwell.

Storch, S. A., & Whitehurst, G. J. (2002). Oral language and code-related precursors to reading: Evidence from a longitudinal structural model. *Developmental Psychology, 38*, 934–947.

Taylor, S.E., Frackenpohl, H., & Petee, J.L. (1960). Grade level norms for the components of the fundamental reading skill. E.D.L. *Research and Information Bulletin* (Vol. 3). Huntington, NY: Educational Developmental Labs, Inc.

Tewes, U. (1983). *HAWIK-R. Handbuch und Testanweisung. [Handbook and test instructions.]* Bern: Huber.

Vellutino, F. R., Fletcher, J. M., Snowling, M. J., & Scanlon, D. M. (2004). Specific reading disability (dyslexia): What have we learned in the past four decades? *Journal of Child Psychology and Psychiatry, 45*, 2–40.

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.

Wagner, R., & Torgesen, J. K. (1987). The nature of phonological processing and its causal role in the acquisition of reading skills. *Psychological Bulletin, 101*, 192–212.

Whitehurst, G. J., & Lonigan, C. J. (1998). Child development and emergent literacy. *Child development, 69*, 848–872.

Yang, S.-N. (2006). A oculomotor-based model of eye movements in reading: The competition/interaction model. *Cognitive Systems Research, 7*, 56–69.