

Eye Movements of Children and Adults Reading in Three Different Orthographies

Sascha Schroeder

Tuomo Häikiö

Ascensión Pagán

Jonathan H. Dickins

Jukka Hyönä

Simon P. Liversedge

Author Note

Sascha Schroeder, University of Göttingen, Institute of Psychologie, Germany; Tuomo Häikiö, Department of Psychology and Speech-Language Pathology, University of Turku, Finland; Ascensión Pagán, Department of Neuroscience, Psychology and Behaviour, University of Leicester, UK; Jonathan H. Dickins, University of Southampton; Jukka Hyönä, Department of Psychology and Speech-Language Pathology, University of Turku, Finland; Simon P. Liversedge, School of Psychology and Computer Science, University of Central Lancashire, UK.

Correspondence concerning this article should be addressed to Sascha Schroeder, University of Göttingen, Department of Educational Psychology, Waldweg 26, 37073 Göttingen, Germany. Email: [sascha.schroeder@psych.uni-goettingen.de](mailto:sascha.schroeder@psych.uni-goettingen.de)

**©American Psychological Association, 2021. This paper is not the copy of record and may not exactly replicate the authoritative document published in the APA journal. The final article is available, upon publication, via its DOI: 10.1037/xlm0001099**

### **Abstract**

In this study, we investigated developmental aspects of eye movements during reading of three languages (English, German and Finnish) that vary widely in their orthographic complexity and predictability. Grapheme-phoneme correspondence rules are rather complex in English and German but relatively simple in Finnish. Despite their differences in complexity, the rules in German and Finnish are highly predictable, whereas English has many exceptions. Comparing eye movement development in these three languages, thus, allows us to investigate whether orthographic complexity and predictability have separate effects on eye movement development. Three groups of children, matched on years of reading instruction, along with a group of proficient adult readers in each language were tested. All participants read stimulus materials that were carefully translated and back-translated across all three languages. The length and frequency of 48 target words were manipulated experimentally within the stimulus set. For children, word length effects were stronger in Finnish and German than in English. In addition, in English effects of word frequency were weaker and only present for short words. Generally, English children showed a qualitatively different reading pattern, while German and Finnish children's reading behavior was rather similar. These results indicate that the predictability of an orthographic system is more important than its complexity for children's reading development. Adults' reading behavior, in contrast, was remarkably similar across languages. Our results, thus, demonstrate that eye movements are sensitive to language-specific features in children's reading, but become more homogenous as reading skill matures.

*Keywords:* eye movements, reading development, cross-linguistic comparison

### **Eye Movements of Children and Adults Reading in Three Different Orthographies**

An accumulating body of studies investigating children's eye movement control during reading development (see Schroeder et al., 2015, for a review) has provided valuable insights that have been replicated in different languages (Blythe & Joseph, 2011). However, a question that is impossible to answer by studying readers of a single language alone is the extent to which eye movements during reading are influenced by the orthography of the language that children are learning to read. Children's reading development and consequent eye movement behavior during reading is potentially affected by two broad classes of factors: universal and language-specific. Universal factors such as the maturation of the visual system and children's increasing efficiency in lexical processing underlie reading development independent of the writing system. In contrast, language-specific factors such as alphabetic status, orthographic consistency, and morphological complexity, might affect the way children learn to read in different writing systems (see also Feng et al., 2009).

In particular, one may envision two competing views on the impact of orthographic factors during eye movement development. One hypothesis is that children's eye movements are relatively similar at first, but become increasingly sensitive to the features of their orthography during reading development; that is, readers continuously adapt their eye movements to their writing system. Another possibility is that children's reading behavior initially depends on specific features of their writing system, but becomes more automatized and, as a consequence, more homogenous as the reading skill develops.

At present, the relative influence of universal and language-specific factors is largely unknown. Only few studies have investigated children's cross-linguistic reading development through examination of eye movements (Feng et al. 2009; Rau et al. 2016), and none has compared reading in more than two languages. In this study, we aimed at disentangling effects of these two classes of factors on reading development by investigating reading development in three alphabetic orthographies: English, German, and Finnish. These languages vary

in their orthographic consistency; that is, in how easy it is to derive the pronunciation of a word from its spelling. In each language, children, matched on years of reading instruction, as well as native adult speakers read sentences that had been carefully translated and back-translated to make them closely comparable across languages. The length and frequency of specific target words within the experimental sentences were manipulated in order to examine cross-linguistic differences in children's and adults' lexical processing.

### **Cross-Linguistic Differences in Reading Development**

During reading, visual information has to be transformed from an abstract orthographic code into word meanings that are then processed in relation to other linguistic information and real world knowledge (see Liversedge et al., 2011, for an overview). This is a complex psychological process that becomes continuously more automatized with extensive practice and accumulating print exposure. A growing number of studies has used eye movement recordings to investigate children's reading development (see Blythe, 2014; Blythe & Joseph, 2011; Milledge & Blythe, 2019; Schroeder et al., 2015, for reviews). The overall pattern emerging from these studies is that children read more slowly than adults, tend to make shorter saccades, fixate words more frequently and for a longer time, and show higher regression probabilities, but lower word skipping rates (see Reichle et al., 2013, for a summary). These effects are consistently found and have been observed in languages as diverse as English (Blythe et al., 2006; Blythe et al., 2015; Milledge et al. 2021), Finnish (Häikiö et al., 2009), German (Huestegge et al., 2009; Rau et al., 2014; Tiffin-Richards & Schroeder, 2015a), French (Mancheva et al., 2015), Chinese (Blythe et al., 2012; Chen & Ko, 2011; Zang et al., 2013; Liang et al., 2015, 2017, 2021), and Japanese (Jincho et al., 2014).

Although the same developmental pattern has been observed in several languages, this does not imply that there are no differences in developmental trajectories between languages. In order to investigate such differences, specifically designed cross-linguistic studies are needed that directly compare children's and adults' reading behavior using identical materials

in all languages. Unfortunately, few studies to date have employed this approach using eye-tracking. For skilled adult readers, previous research has shown more similarities than differences between readers' eye movements in different writing systems, even if the orthographies differ substantially from each other, as is the case for English and Chinese. For example, Sun et al. (1985) showed that when native Chinese and English speakers read what were argued to be comparable scientific articles, most eye movement variables, such as mean fixation duration, saccade length, and reading rate, were remarkably similar. In addition, Li et al. (2014) demonstrated that the effects of the properties of the currently fixated, previous, and next words on eye movements are comparable in Chinese and alphabetic languages such as English or German.

More recently and more relevant for the present study, Liversedge et al. (2016) compared adult readers' eye movements in English, Finnish, and Chinese using identical texts that have been translated into all three languages. Although they found consistent cross-linguistic differences, these were mainly related to differences in the visual density between the three orthographies. By contrast, participants in all languages showed comparable sentence reading times and their processing was affected by the same linguistic variables. This suggests that adults are similarly efficient in extracting the meaning of a sentence and that there are strong similarities in their eye movements in relation to linguistic processing, even if they are reading in very different orthographies.

By contrast, children's eye movements seem to be sensitive to the characteristics of the writing system in which they are learning to read. For example, Feng et al. (2009) compared text reading among native English and Chinese speakers in grade 3, grade 5, and young adulthood. In both languages, the usual developmental pattern (decreasing reading times and number of fixations, increasing saccade length) was observed. However, English-speaking children showed more pronounced developmental effects; that is, younger children showed longer

word reading times relative to Chinese children. By contrast, adults' eye movements were remarkably similar in both languages. These findings suggest that language-specific factors particularly affect beginning readers. It is difficult, however, to derive more detailed conclusions from this study, because the materials were matched only on overall semantic content, but there was no correspondence at the sentence or word level.

Similar findings have been reported by Rau et al. (2016), who investigated the eye movements of German- and English-speaking primary school children and adults who read single sentences aloud. In particular, they found that German children showed longer gaze durations and were more likely to refixate a word during first-pass reading. English children, by contrast, showed longer re-reading times and made more inter-word regressions. Adults' reading behavior in the two languages was very similar. Note, though, that Rau et al. (2016) employed an oral reading paradigm and did not include comprehension questions to ensure readers were understanding the sentences. Given that considerable differences have been found in eye movements during silent versus oral reading (e.g., Vorstius et al., 2014), it is unclear whether these findings generalize to reading situations without concurrent articulation. In addition, they used a reading-level matched design: English and German children were matched on gaze duration spent in fixating short, high-frequency words. As a consequence, the English children were older and had received considerably more reading instruction than the German children (31 vs. 48 months). It is, thus, unclear whether the same differences between English- and German-speaking children can also be observed for children with the same amount of reading instruction. Therefore, the results of Rau et al. (2016) should be replicated using a design in which children are matched on the amount on reading instruction they have received.

In order to account for their findings, Rau et al. (2016) applied the distinction between a "plodder" and an "explorer" reading strategy introduced by Olson et al. (1985) in the context of dyslexic reading. According to Olson et al., plodders make relatively few regressions, skip

over words only rarely, and move steadily forward in the text using many (re)fixations. Explorers, by contrast, skip words more often and make less refixations, but they also regress more frequently to words in case comprehension problems occur downstream in the reading processes. Rayner et al. (2006) have called this reading pattern “risky reading” in the context of comparing younger and older adults. It is important to realize that the plodders and explorers do not form distinct categories, but rather a continuum and it is likely that the relative position on the plodder-explorer spectrum is influenced by both person- (age, reading skill) and language-specific factors. Specifically, Rau et al. (2016) argue that German children, reading in an orthographically shallow orthography, are more likely to employ a localized, plodder-like reading style because they can use a smaller grain size in recognizing words. By contrast, English children, learning to read in an orthographically deep orthography, presumably use a more global, explorer-like reading style because they use a larger grain size. At present, however, it is unclear whether this also applies to other languages next to English and German, and how orthographic consistency is linked to the position on the plodder-explorer continuum.

### **Differences in Orthographic Consistency Between English, German, and Finnish**

Most cross-linguistic eye movement studies have compared either very different orthographies, such as English and Chinese, or rather similar orthographies, such as English and German. In the present study, by contrast, we aimed at an optimal trade-off between comparability and diversity. We investigated only alphabetic writing systems, which are easier to compare than, for example, alphabetic and logographic orthographies. At the same time, the three languages that we selected differ widely on several dimensions (see Table 1; see Verhoeven & Perfetti, 2017, 2021, for an extensive description of the characteristics of the three languages and their impact on reading development).

First, words in Finnish, German and English differ in their spatial density. In particular, the three languages differ with regard to the average number of characters that comprise the words in each language (Finnish > German > English). In addition, the three languages vary

considerably in their morphological complexity. In particular, English conveys very little morphological information. German, in contrast, has a rich derivational system and makes extensive use of compounding. Even more morphological information is provided in Finnish, which is an agglutinating language (i.e. much of the semantic content of a sentence is provided by inflectional morphemes at the end of content words).

Second, on the phonological dimension, the basic inventory of vowels and consonants is rather similar in the three languages. However, syllable structure in English and German, which are both Germanic languages, is more complex than in Finnish, which mostly comprises short, open syllables similar to Romance languages (see Seymour et al., 2003). As a consequence, the number of syllables is much larger in English and German than in Finnish.

Finally, on the orthographic dimension, the letter inventories of English, German, and Finnish are again very similar. However, the three languages differ substantially in the consistency of the mapping between phonology and orthography at the grapheme level. Finnish has an extremely consistent orthography and approximates a 1:1 mapping between letters and sounds; for example, the letter “a” is always pronounced as /a/. English, by contrast, is rather inconsistent in this respect and the letter–sound mapping is more complicated due to complex multi-letter graphemes, context-dependent rules, and irregularities. For example, the letter “a” is pronounced differently in the English words “ball,” “bake,” and “hand”. German has an intermediate level of orthographic consistency. For example, the letter “a” always represents the phoneme /a/ which is, however, either short or long depending on its orthographic context.

Orthographic consistency is not a unitary concept, however. As Schmalz et al. (2015) have argued, it is important to distinguish between two different aspects of orthographic consistency: complexity and predictability. The complexity of a writing system refers to the structure of its underlying grapheme-phoneme-correspondence (GPC) rules. Writing systems that have more numerous and more complex (i.e., multi-letter, context-sensitive, position-specific) GPC rules are more complex than writing systems that have fewer and simpler rules.



English, for example, makes abundant use of complex GPC rules such as the “magic e” rule that specifies that a silent “e” at the end of a word modifies the pronunciation of the preceding vowel (e.g., from /i/ to /ai/ in “rid” vs. “ride”). Most GPC rules in Finnish, by contrast, are not position- or context-sensitive and usually only involve single letters (with the notable exception for the velar nasal “ng”). In contrast to its complexity, the predictability of a writing system refers to the number of irregular words, that is, words in which the application of the GPC rules provides an incorrect pronunciation. For example, although the “magic e” rule mentioned above successfully predicts the correct pronunciation for many English words, there are also many exceptions to it (e.g., “give”). Finnish does not have any irregular words and its GPC rules are applied very consistently. German represents an interesting intermediate case here, because it has many complex GPC rules, which are, however, applied very consistently (Ziegler et al., 2000). For example, the length of a vowel in German depends on the number of the subsequent letters: if it is followed by a single consonant, its pronunciation is long (e.g., /a:/ in “Schal”; in English “scarf”), but if it is followed by two identical consonants, its pronunciation is short (e.g., /a/ in “Schall”; in English “sound”). Although this rule is rather complex, there are hardly any exceptions to it (except for very few function words).

Table 1.

*Characteristic differences between the written language systems of Finnish, English and German.*

Language	Length		Phonology			Orthography		
	Mean word length <sup>a</sup>	Morphological complexity	No. phonemes (vowels vs. consonants)	No. syllables	Maximal syllable structure	No. letters (vowel vs. consonants)	Regularity <sup>b</sup>	Number of GPC rules <sup>c</sup> (simple vs. complex)
Finnish	13.3	High	16, 22 = 38 phonemes	ca. 3,000	CVVC	8, 12 = 20 letters	0.00	(38, 2)
German	11.1	Medium	19, 24 = 41 phonemes	ca. 10,000	CCCVVCCC	9, 20 = 29 letters	0.45	(44, 86)
English	8.6	Low	22, 24 = 46 phonemes	ca. 12,000	CCCVCCCC	6, 20 = 26 letters	0.83	(38, 188)

*Note.* <sup>a</sup> English/German: CELEX corpus (all word forms), Finnish: Finnish Newspaper corpus. <sup>b</sup> The higher the entropy value, the more ortho-phonologically inconsistent the writing system (Borgwaldt et al., 2005). <sup>c</sup> English/German: Number of single-letter vs. multi-letter rules in the DRC model (see Schmalz et al., 2015), Finnish: Number of GPC rules according to Karlsson (1999).

To summarize, although English, German, and Finnish all adhere to the alphabetic principle, they differ widely in their orthographic consistency. Finnish is a language with a spelling system that is rather simple and extremely predictable. English, by contrast, has an orthography that contains many complex spelling rules but also has a high level of unpredictability. Although German is generally considered a consistent orthography, its status is somewhat ambiguous: similar to English, many spelling rules are rather complex, but similar to Finnish, the pronunciation of most words is highly predictable. Thus, comparing children's eye movements in English (- predictable, + complex), German (+ predictable, + complex), and Finnish (+ predictable, - complex) allows us to investigate the effects of orthographic predictability and orthographic complexity. Specifically, the contrast between English and German allows us to replicate the effect of orthographic predictability on children's eye movements (Rau et al., 2016). In addition, the contrast between German and Finnish allows us to test whether orthographic complexity has additional effects on children's eye movement development over and above orthographic predictability.

### **Word Length and Frequency Effects on Children's Eye Movements**

Dual-route models of visual word recognition assume two functionally different ways of accessing word meaning (e.g., Coltheart et al., 2001; Grainger & Ziegler, 2011; Perry et al., 2007): the lexical route through which the lexical entry is directly retrieved via (frequency-weighted) connections between the orthographic and the phonological or semantic lexicon, and the sublexical route through which the pronunciation is assembled via GPC rules and is then used to access the meaning of a word.

A marker effect for the influence of the lexical route is the effect of word frequency on word recognition, as the strength of representations in the orthographic and phonological **lexicons** are assumed to be weighted by how often a word has been encountered. Words that have been read or heard more often are thus easier and faster to recognize because they use the lexical route during lexical processing, while novel words are identified through the sublexical

route. There is ample empirical evidence in the developmental eye-tracking literature that children's eye movements are sensitive to word frequency (Blythe & Joseph, 2011; Schroeder et al., 2015) with longer fixation times and more fixations spent on infrequent than frequent words. These effects tend to be stronger in children than in adults, especially if word frequency norms appropriate for children are used (Joseph et al., 2013), and they have been demonstrated in both German (Huestegge et al., 2008; Tiffin-Richards & Schroeder, 2015b) and English (e.g., Blythe et al., 2009; Hyönä & Olson, 1995), although they are not reliably found in all studies (e.g., Blythe et al., 2006). As far as we know, frequency effects have not been investigated in Finnish children, partly due to the fact that no specific child frequency norms are available for that language.

Because the sublexical route is assumed to operate serially from left to right (Coltheart & Rastle, 1994), a marker effect for the influence of the sublexical route is the word length effect; that is, the extent to which reading times increase as a function of stimulus length. Previous eye-tracking studies have shown that the length of a word strongly affects children's eye movements. Children fixate longer words more often and for longer duration; these effects have been found in English (Joseph et al., 2009), German (Huestegge et al., 2009; Tiffin-Richards & Schroeder, 2015b), and Finnish (Blythe et al., 2011). Word length effects are usually larger in children than in adults (Joseph et al., 2009), decrease in magnitude from grade 2 to 4 (Huestegge et al., 2009), and are stronger for less skilled than for skilled readers (Hyönä & Olson, 1995). These findings are generally in line with the notion that younger and less skilled readers are more likely to adopt a sublexical decoding strategy.

In addition, because the indirect sublexical route is assumed to be slower than the direct lexical route, dual route models predict that sublexical assembly should interact with lexical activation (Coltheart et al., 2001). When there is little or no lexical activation, as in reading low-frequency words, length effects should be relatively strong. In contrast, in the presence of strong lexical activation, as in reading high-frequency words, small or no length effects

should be found. The size of the frequency by length interaction may thus be considered as a marker effect for the amount of lexical differentiation between high- and low-frequency words.

Although there is consistent evidence for the frequency by length interaction from single word recognition studies in English (e.g., Weekes, 1997), few eye movement studies have yet investigated it. The results from the child studies are mixed, but there is some evidence that word length effects on children's eye movements are indeed larger for infrequent than for frequent words. In English, Hyönä and Olson (1995) reported an interaction between word length and frequency for 10-year-old typically reading children as well as for a group of dyslexic children. Word length effects were larger for infrequent than for frequent words in terms of both first fixation and gaze duration. By contrast, Huestegge et al. (2009) found no interaction between word length and frequency in German 2<sup>nd</sup> and 4<sup>th</sup> graders. In both studies, however, word frequency norms for adults were used. When age-appropriate materials are employed, Tiffin-Richards and Schroeder (2015b) showed that length by frequency interactions can be observed even in German 2<sup>nd</sup> graders (see also Joseph et al., 2013). For Finnish, length by frequency interactions have not been investigated yet.

It is assumed that readers adapt their processing to the demands of the orthography they are reading (e.g., Katz & Feldman, 1983). In consistent orthographies, readers are believed to rely more strongly on the sublexical route, because the mapping between letters and sounds is relatively unambiguous. In inconsistent orthographies, on the other hand, readers rely more strongly on larger grain size and the lexical route. In line with the notion that orthographic consistency affects the specific mix of lexical and sublexical processing, cross-linguistic studies have shown that length effects are stronger in consistent than in inconsistent orthographies in adults (e.g., Ziegler et al., 2001) and children (Ziegler et al., 2003). Conversely, it has also been demonstrated that large-unit effects such as the effect of body neighborhood or fre-

quency are more noticeable in inconsistent orthographies (e.g., Ziegler et al., 2001). Computationally, differential length and frequency effects can be modelled by varying the relative weight of the sublexical and lexical route as a function of the consistency of a language (Perry & Ziegler, 2002).

At present, however, only one study has explicitly [investigated the relationship between orthographic differences and length and frequency effects](#) among children during reading. In the study by Rau et al. (2015), they recorded eye movements of German- and English-speaking primary school children and adults who were required to read single sentences aloud. Each sentence included a target word manipulation for word length (3–5 vs. 6–8 letters) and familiarity (high-frequency vs. low-frequency vs. nonwords) in order to investigate the effects of small (letter) and large (word) processing units. The authors found that German children showed larger length effects than English children during first-pass reading of words. In contrast, English children showed larger length effects than German children in re-reading time and evidently needed to regress to difficult words more often than German children. The authors explain this pattern of effects by arguing that English children use a larger grain size and more lexically based reading strategy during first-pass reading than German children. By contrast, although the main effects of familiarity and the familiarity by length interaction were significant for all eye movement measures, these effects did not differ substantially between English and German children except in re-reading time. However, as the familiarity manipulation also involved nonwords, it is not clear whether the effect can also be observed when only word frequency is manipulated. More importantly, it is not clear whether Finnish children will show similar effects as German children or a qualitatively different pattern because the Finnish orthography is less complex.

### **Rationale of the Present Study**

In the present study, we investigated reading development in English, German, and Finnish. Specifically, we examined how readers' eye movements are affected by orthographic

consistency among children and adults reading identical single sentences. We were particularly interested in word length and frequency effects as markers of sublexical and lexical processing. Therefore, target words that varied in length (short vs. long) and frequency (high vs. low) were embedded in sentences to investigate whether length effects are moderated by frequency in a similar way in all three languages.

We expected to replicate two important eye movement patterns obtained in previous studies on reading development. On one hand, we should observe the typical developmental pattern in all the languages, that is, children will read more slowly, make shorter saccades, make more and longer fixations on words, make more regressions and less word skipping than adults (e.g., Blythe et al., 2015; Häikiö et al., 2009; Tiffin-Richards & Schroeder, 2015a). On the other hand, according to Liversedge et al. (2016), adults should show similar sentence reading times and similar effects of linguistic factors in all languages. This predicted pattern is consistent with the idea that the effects of orthographic differences in the writing system are more [homogeneous](#) as reading skill develops. In addition, given the differences in word length between the three languages, we predicted adult Finnish readers to make more and shorter fixations and longer saccades than adult English readers (as per Liversedge et al., 2016). However, the pattern of effects for German adult readers is unclear, given that there are no studies comparing German with either Finnish or English adult readers.

Concerning cross-linguistic differences in reading among children, we should observe a different reading style for German and English children. According to Rau et al. (2016), German children will adopt a plodder-like reading style (more refixations, few regressions and skips), while English children will show an explorer-like reading style (more skips and regressions, less refixations). In addition, we expected German children to show larger word length effects than English children in early measures of word processing (first-pass reading or gaze duration), while an opposite pattern will be observed in later measures of word processing (total reading time) (as per Rau et al., 2014; 2015).

The unique contribution of the present study is that we investigated developing readers in three languages at the same time. This allows us to investigate the impact of the two dimensions of orthographic consistency: predictability and complexity. As elaborated above, German and English mainly differ in the amount of orthographic predictability while the differences in orthographic complexity are less pronounced. By contrast, Finnish is both more predictable and less complex than English. We already know that orthographic predictability affects children's eye movements (e.g., in the studies by Rau et al., 2015, 2016), but the status of orthographic complexity has not yet been explored. Therefore, a key question addressed in the present study is whether Finnish children's eye movements will show a qualitatively different pattern or cluster together with the German children. If differences in children's eye movement behavior are mainly related to the predictability of an orthographic system, children should read similarly in German and Finnish. If, by contrast, the complexity of an orthographic system has an independent effect on children's eye movements, we would expect to see qualitative differences between the two languages.

Finally, according to dual route models of word recognition (e.g., Coltheart et al., 2001), larger word length effects are predicted for low-frequency words compared to high-frequency words, suggesting that there are two different mechanisms to identify words depending on their frequency. High-frequency words are accessed via the lexical route while low-frequency words are accessed via the sublexical route. At the same time, this interaction implies that word frequency effects should be stronger for long than for short words. Therefore, if orthographic consistency affects the use of these mechanisms, the size of the frequency by length interaction will vary depending on orthographic predictability and/or complexity.

A particular methodological challenge for any cross-linguistic study is to ensure comparability of linguistic materials as well as that of participants. The three languages under investigation differ not only in their orthographic consistency, but also in other linguistic properties. In the present study, we used translation equivalents of target words that were matched



on number of letters and frequency norms for children. We additionally collected age-of-acquisition norms for all languages in order to ensure that there were no differences between languages in children's familiarity with the target words. Because English, German, and Finnish differ in their syntactic constraints, it is not always possible to achieve an exact word-by-word correspondence in the target sentences without compromising their naturalness. In the present study, we were able to minimize any such differences and to ensure semantic equivalence. Following Liversedge et al. (2016), sentences were matched on structural and lexical aspects as far as possible and were carefully translated and back-translated across all three languages. In addition, each sentence comprised exactly the same propositions in the same order up to the target word. In order to check that the sentences sounded natural in each language, we additionally collected naturalness ratings. We were thus able to investigate the impact of orthographic transparency while controlling for important semantic and pragmatic aspects of the target sentences.

Regarding comparability of participant samples, an important difference concerns the age at which the children enter school. In the present study, we matched children in terms of years of formal reading instruction and investigated children who had been learning to read for approximately 3 years. This grade-matched design necessarily implies differences in chronological age between our samples, because English children start school at the age of 5 years, whereas German and Finnish children at the age of 6 and 7 years, respectively. To check that children in all three languages showed reading behavior appropriate for the 4<sup>th</sup> grade, we administered a normed reading test that was similar in all languages.

## Methods

All materials, eye tracking data, and analysis scripts of this study can be found in the Open Science Framework, at the following URL: <https://osf.io/ay6xk/>.

### Participants

The final sample comprised 30 children and 25 adults in each language. All participants reported to be native speakers of English, German, and Finnish, respectively, with normal or corrected-to-normal vision and no known reading difficulties (see Table 3).

**English sample.** Children in grades 3 or 4 were recruited from schools in the Southampton area through the Psychology Research in Partnership Scheme of the University of Southampton (schools, rather than individual children, were compensated through this scheme by way of dissemination and outreach activities). Three additional child participants were recruited through a similar arrangement with schools in the Bournemouth area. Two children were replaced because they showed poor performance on the comprehension questions (i.e., < 70% correct). Adults were all students at the University of Southampton and received course credit for their participation. Four adults were replaced (two due to calibration problems and two because they scored below the 10<sup>th</sup> percentile on a standardized reading test, see below).

**German sample.** Children in grades 3 or 4 were recruited from one school in Berlin or through the participant database of the Max Planck Institute for Human Development, Berlin. They received candy or €5 financial reimbursement for their participation. One child was replaced due to calibration problems and a further three because they scored below the 10<sup>th</sup> percentile on a standardized reading test. Adults were recruited using the participant database of the Max Planck Institute for Human Development and received €10 for their participation. Twenty participants reported to be students while 5 had a high-school diploma, but were currently not studying, but employed. Two adults were replaced because of equipment failure and one because he scored below the 10<sup>th</sup> percentile on a standardized reading test.

**Finnish sample.** Children in grade 4 were recruited from one school in the Turku area. They received candy or stickers for their participation. No child had to be replaced. Adults were all students at the University of Turku and received course credit or a movie ticket for their participation. Two adults were replaced due to calibration problems or equipment failure.

### **Measures**

**Nonverbal intelligence.** In all three languages, the Matrix subtest from the Culture Fair Intelligence Test (CFT 2; Cattell, 1963) was used to assess children's and adults' nonverbal intelligence. In this test, participants are shown 12 geometric patterns, each with one part missing, and are asked to identify which of five possible continuations completes the pattern. As only one subtest was administered, no norm values were available.

**Reading skill.** Participants' reading skill was assessed using comparable reading fluency tests including both words and nonwords. In English, we used the TOWRE-2 test (Torgesen et al., 1999). In this test, participants are shown two lists of words and nonwords and instructed to read aloud as many items as possible from each list in 45 seconds. In German, we used the SLRT-II test (Moll & Landerl, 2009). The procedure is identical to that of TOWRE-2, but participants have 60 seconds to complete each list. In Finnish, we used a comparable word reading test, the Lukilasse reading test (Häyrynen et al., 1999), and an unpublished nonword list developed at the University of Jyväskylä. Again, participants read aloud a list of words and nonwords and had 120 and 45 seconds, respectively, to read as many items as possible.

The test score for each list was the number of words read correctly (i.e., total words minus reading errors). Because the items and the exact procedure differed across the three tests, differences in the raw scores cannot be interpreted between languages. Rather, we used norm values (percentile ranks). For the English and German tests, norm values for words and nonwords are available for adults and children (stratified by grade). For the Finnish test, norm

values are currently available only for children's reading of the word list (stratified by grade). In addition, because the Finnish test was developed to assess the reading skill of children, adults showed ceiling effects on the word subtest. Raw scores for words and nonwords correlated highly in each language (English:  $r = .55$ , German:  $r = .78$ , Finnish:  $r = .70$ ). In the English and German sample, we thus used a combined score (according to the 4<sup>th</sup> grade norms) for screening and replaced all participants who scored below the 10<sup>th</sup> percentile. In the Finnish sample, the cut-off score for children was the 10<sup>th</sup> percentile on the word list, whereas no objective exclusion criteria could be applied to adults.

### Materials

All materials were carefully matched and translated and back-translated multiple times between languages to ensure the highest possible level of correspondence. Items were generated in English and then translated into both German and Finnish. If a target word or syntactic construction provided difficulties during the translation process, it was discarded and a new item was generated. At the end of this process, all items were back-translated from German and Finnish into English by a different person. If the back-translation did not match the original item, it was discarded. The stimulus characteristics are provided in Table 2.

**Target Words.** We selected 48 target nouns that could easily be translated into all three languages without changing their meaning (i.e., cognates or translation equivalents, e.g. English: “goat”, German: “Ziege”, Finnish: “vuohi”; see Appendix A for a complete list of target words). Word length (short vs. long) and word frequency (high vs. low) of the target words were manipulated orthogonally: there were 12 target words for each combination of length and frequency (e.g., short/frequent: “goat”/“Ziege”/“vuohi”; short/infrequent: “panda”/“Panda”/“panda”; long/frequent: “elephant”/“Elefant”/“elefantti”; long/infrequent: “giraffe”/“Giraffe”/“kirahvi”).

Short words were 3–5 letters long and long words were 7–9 letters long. Some flexibility was needed in order to accommodate differences between the three languages. That is, all

short words were in the 3–5 letter range, but the exact number of letters was allowed to differ between languages (e.g., the translation of the English word “goat”, which is 4 letters long, have 5 letters in German and Finnish, see above). Overall, however, we ensured that short and long words were of similar average length in the three languages (see Table 2, upper section). For each language, there was a robust difference in word length between short and long words. However, short and long words did not differ in word frequency, and the strength of the word length manipulation was similar for high- and low-frequency words.

In manipulating word frequency, we were able to draw on word frequency norms for children in English (taken from the *Educator’s Word Frequency Guide*; WFG, Zeno et al., 1995) and German (taken from the *childLex* corpus; Schroeder et al., 2015a). In Finnish, however, there are currently no frequency norms available for children.

The English WFG norms (type frequencies, grades 1–6) were used to select high- and low-frequency words. Words with normalized frequencies below 11/million were defined as low-frequency words and those with normalized frequencies of at least 11/million as high-frequency words. This cut-off was rather liberal, but necessary to accommodate differences between languages. Overall, low-frequency words had a mean frequency of  $M = 4.3/\text{million}$  ( $SD = 3.2$ , range: 0–10), while high-frequency words had a mean frequency of  $M = 45.1/\text{million}$  ( $SD = 35.1$ ; range: 12–154). Thus, the frequency manipulation was comparable to that of other studies investigating frequency effects in children (Huestegge et al., 2009; Joseph et al., 2013). As Table 2 (second section) shows, low- and high-frequency words differed strongly in their average WFG frequency. However, there were no differences in word length between high- and low-frequency words, and the size of the word frequency manipulation was similar for short and long words. Similar results were obtained using the German *childLex* norms (version 0.15, lemma frequencies, see Table 2).

Table 2  
*Word and Sentence Characteristics*

	Frequency				<i>F</i> (1,44)		
	High		Low		Length	Frequency	Length × frequency
	Short	Long	Short	Long			
	<i>Length<sup>a</sup></i>						
English	4.2 (0.7)	8.0 (1.0)	4.3 (0.8)	7.3 (0.7)	228.3***	1.2	3.4
German	4.5 (0.7)	7.5 (0.9)	4.5 (0.7)	7.8 (0.8)	204.2***	<1	<1
Finnish	4.8 (0.5)	7.8 (0.9)	4.8 (0.5)	7.9 (0.8)	244.6***	<1	<1
	<i>Child Frequency<sup>b</sup></i>						
English	1.6 (0.3)	1.5 (0.3)	0.8 (0.2)	0.5 (0.3)	3.1	133.5***	1.3
German	1.4 (0.4)	1.4 (0.2)	0.7 (0.5)	0.6 (0.4)	<1	46.8***	<1
Finnish	-	-	-	-	-	-	-
	<i>Age-of-Acquisition<sup>c</sup></i>						
English	5.4 (1.6)	5.7 (1.4)	8.1 (2.1)	6.8 (1.7)	1.2	13.9**	2.3
German	4.2 (1.1)	4.5 (1.0)	6.1 (1.5)	5.9 (1.4)	<1	19.6**	<1
Finnish	4.4 (1.6)	4.6 (1.3)	6.5 (2.5)	6.3 (1.6)	<1	13.1**	<1
	<i>Adult Frequency<sup>d</sup></i>						
English	1.4 (0.3)	1.5 (0.4)	1.0 (0.3)	0.8 (0.3)	<1	32.9***	1.5
German	1.1 (0.4)	1.5 (0.4)	0.8 (0.4)	0.5 (0.3)	<1	36.0***	10.2**
Finnish	1.3 (0.5)	1.5 (0.6)	1.0 (0.4)	0.6 (0.3)	<1	21.8***	4.6*
	<i>OLD<sup>d</sup></i>						
English	1.5 (0.4)	1.6 (0.4)	3.1 (0.6)	2.7 (0.6)	109.9***	1.1	3.2
German	1.7 (0.3)	1.9 (0.3)	2.9 (0.6)	2.9 (0.4)	101.1***	<1	<1
Finnish	1.8 (0.1)	1.8 (0.2)	2.5 (0.4)	2.8 (0.6)	61.5***	3.3	2.4
	<i>Naturalness<sup>e</sup></i>						
English	4.5 (0.4)	4.6 (0.3)	4.5 (0.3)	4.6 (0.4)	<1	<1	<1
German	4.1 (0.4)	4.2 (0.3)	4.0 (0.3)	4.2 (0.4)	<1	<1	<1
Finnish	4.4 (0.2)	4.5 (0.3)	4.3 (0.4)	4.4 (0.4)	<1	<1	<1

*Note.* \*  $p < .05$ , \*\*  $p < .01$ , \*\*\*  $p < .001$ . *SDs* in parentheses. All frequencies incremented by 1 and log transformed to the base of 10. <sup>a</sup>*n* letters. <sup>b</sup>English frequencies: WFG corpus (Zeno et al., 1995), German frequencies: childLex (Schroeder et al., 2015a). <sup>c</sup>In years. English: Kuperman et al. (2012). German/Finnish: Student ratings. <sup>d</sup>English and German: CELEX (Baayen et al., 1995), Finnish: Turku Newspaper Corpus (Laine & Virtanen, 1999). <sup>e</sup>Student rating (0 = very unnatural to 5 = very natural).

In order to validate the child frequency norms and to ensure that they generalized to Finnish, we also checked age-of-acquisition (AoA) norms for the target words. For English, we used the AoA ratings provided by Kuperman et al. (2012). Here, participants were asked to specify the age (in years) at which they thought they learned the word. For German and Finnish, similar ratings were collected from 38 German and 20 Finnish university students. The reliability of the ratings was very high (German:  $ICC = .98$ , Finnish:  $ICC = .97$ ) and they were averaged across participants. In English and German, AoA ratings correlated highly with the child word frequency norms (English:  $r = -.53$ , German:  $r = -.61$ ), but not with the corresponding adult frequencies (see below; English:  $r = -.19$ , German:  $r = -.27$ ; in Finnish, the correlation between AoA and adult word frequency was also low,  $r = -.18$ ). This indicates that AoA ratings are better suited for approximating children's familiarity with words than they are for approximating adult word frequency counts. The average AoA ratings for the four target word conditions are provided in Table 2 (third section). AoAs differed strongly between high- and low-frequency words in all three languages, all  $F > 13$ . Overall, high-frequency words were acquired 2 years earlier than low-frequency words. By contrast, AoA ratings did not vary with word length, and the interactions between word length and frequency were not significant, all  $Fs < 1$ . The English ratings were higher than the German or the Finnish ratings, but this difference is presumably caused by differences in the data collection procedures.

We additionally obtained adult frequency norms for the target words in all three languages. For English and German, we used the CELEX norms (Baayen et al., 1995). For Finnish, we used the frequency norms of a newspaper corpus containing 22.7 million word forms (Laine & Virtanen, 1999). Average adult frequencies for the four target word conditions are provided in Table 2. There was a strong main effect of word frequency in each language. In contrast to the child frequency and the AoA analyses, however, the word length by word frequency interaction was significant in German and in Finnish. In all languages, the difference

between high- and low-frequency words was larger for long words, English:  $t = 4.9$ , German:  $t = 6.5$ , Finnish:  $t = 4.8$ , than for short words, English:  $t = 3.2$ , German:  $t = 2.0$ , Finnish:  $t = 1.8$ . The difference was mainly driven by short, high-frequency words, which were less frequent for adults than for children. That is, words like *ball* or *doll* are more frequently found in child than in adult corpora (see Schroeder et al., 2015b, for a detailed comparison of child and adult frequencies). This indicates that the frequency manipulation was similar in both word length conditions for children, but not for adults.

Finally, in order to ensure that the target words were equally representative for the three languages, we computed OLD20 (Yarkoni et al. 2008) values for each target word. OLD20 is the average Levenshtein distance of a word to its 20 closest orthographic neighbors. This measure quantifies how similar a word is to other words in a language. In order to compute the values, we again used the CELEX corpus for both English and German, and the Turku newspaper corpus for Finnish. Results show that in all languages short words had more orthographic neighbors and, as a consequence, lower OLD20 values than long words. The inverse relationship between word length and orthographic similarity is well established, as decreasing length increases the probability that fewer orthographic changes will produce another word (Yarkoni et al. 2008). This effect thus reflects the structure of written language. OLD20 values were very similar across all three languages demonstrating that words were representative. In addition, the interactions with word length and frequency were not significant, all  $F_s < 1$ .

**Sentences.** Each target word was embedded in a carrier sentence. In order to minimize any differences and to ensure semantic equivalence, we made sure that each sentence comprised exactly the same propositions in each language in the same order up to word  $n + 1$  following the target word. However, the same content was expressed slightly differently according to the syntactic constraints of each language. In Finnish, the only position in which nouns appear without any inflection is the subject position. Because Finnish does not use articles,



however, the subject is usually the first word of a sentence, which is suboptimal for eye-tracking studies. In our sentences, the target word was thus always the subject of an embedded subclause that was initiated by a connective. The syntactic structure of subclauses differs between the three languages. In English and Finnish, subclauses have a SVO word order; in German, most subclauses are SOV. There are, however, some connectives in German (e.g., *aber*, *denn*, *oder*, *und*, etc.) that select coordinated structures with a SVO word order. Thus, these connectives were used in order to ensure that the verb always appeared after the target word.

Examples 1a-c illustrate the structure of the carrier sentences and shows that there was both consistency and variability between the three translations (correspondence between constituents is represented by use of the same index):

1a. The trip<sub>1</sub> | to the zoo<sub>2</sub> | was<sub>3</sub> | funny<sub>4</sub> | because<sub>5</sub> | the **goat**<sub>6</sub> | had<sub>7</sub> | run<sub>8</sub> | into the teacher<sub>9</sub>.

1b. Der Ausflug<sub>1</sub> | in den Zoo<sub>2</sub> | war<sub>3</sub> | lustig<sub>4</sub> | denn<sub>5</sub> | die **Ziege**<sub>6</sub> | hatte<sub>7</sub> | den Lehrer<sub>9</sub> | umgerannt<sub>8</sub>.

1c. Retki<sub>1</sub> | eläintarhaan<sub>2</sub> | oli<sub>3</sub> | hauska<sub>4</sub> | koska<sub>5</sub> | **vuohi**<sub>6</sub> | oli<sub>7</sub> | juossut<sub>8</sub> | opettajaa päin<sub>9</sub>.

In the main clause, there was an almost word-to-word correspondence between all languages. However, prepositions are rarely used in Finnish and the corresponding information is part of the noun inflection (e.g., the Finnish locative “eläintarhaan” corresponds to the phrase “to the zoo” in English). Moreover, because sub-clauses in German are verb-final, the position of the object and the main verb are reversed here. Similarly, the phrase “run into the teacher” is expressed slightly differently in the three languages (e.g., in German, the literal translation is “the teacher over-run” and in Finnish it translates into “run teacher against”). Crucially, such differences did not occur before word  $n + 2$ .

Based on their meanings, four target words from each length  $\times$  frequency combination were grouped together and used in the same sentence frame that was identical up to word  $n + 1$  (see Examples 2a-d):

2a. The trip to the zoo was funny because the **goat** had *run into the teacher*.

2b. The trip to the zoo was funny because the **panda** had *spit at the teacher*.

2c. The trip to the zoo was funny because the **elephant** had *sprayed water on the teacher*.

2d. The trip to the zoo was funny because the **giraffe** had *spit on the teacher's head*.

Different continuations of the sentence were allowed to avoid semantic or pragmatic anomalies (e.g., if an elephant runs into a teacher, s/he will presumably be hurt, which is not funny). Because each participant was supposed to read every target word without reading the same carrier sentence twice, we created four different carrier sentences for each set of target words (see Examples 3a-d; similar sentences with different continuations after word  $n + 1$  were also created for the other three target words, see Appendix B for example sentences; all stimuli are available in the online supplemental materials).

3a. *The trip to the zoo was funny because the* **goat** *had run into the teacher.*

3b. *The children were happy because the* **goat** *had given birth to three small kids.*

3c. *Anna saw the animals in the zoo and the* **goat** *was really cute.*

3d. *The bear was angry because the* **goat** *had bitten him.*

There were 16 different sentences for each of the 12 sets of target words and thus 192 sentences in total. As every target word appeared in each of the four carrier sentences, words and sentences were crossed. Four presentation lists were created in which each target word was assigned to a carrier sentence according to a Latin square design. Participants were randomly assigned to one of the presentation lists.

There were only minimal differences between languages for the total number of letters within each sentence (English:  $M = 57.18$ ,  $SD = 12.15$ , German:  $M = 63.79$ ,  $SD = 14.81$ , Finnish:  $M = 58.23$ ,  $SD = 13.55$ ). However, due to syntactic differences between the languages (no articles, only few prepositions, etc. in Finnish), letters were distributed differently within the sentences. In particular, Finnish sentences were approximately 3 words shorter ( $M = 9.41$ ,  $SD = 2.11$ ) than either English ( $M = 12.58$ ,  $SD = 3.05$ ) or German sentences ( $M = 12.49$ ,  $SD = 2.90$ ), both  $t > 20$ , which did not differ from each other in the number of words,  $t < 1$ . By contrast, average word length increased with orthographic consistency (English:  $M = 4.61$ ,  $SD = 0.54$ , German:  $M = 5.13$ ,  $SD = 0.54$ , Finnish:  $M = 6.21$ ,  $SD = 0.67$ ), English–German:  $t > 13$ , German–Finnish:  $t > 20$ . Thus, there was a trade-off relationship between word and sentence length between languages: English sentences comprised many short words, while Finnish sentences comprised fewer, but longer words with German occupying an intermediate position. Crucially, within each language, there were no differences between the four target word conditions for either measure in each language, all  $F < 1.1$ .

In order to ensure that the sentences sounded natural in each language, we collected naturalness ratings from adults. In German, 82 university students rated one of the four presentation lists (i.e., approximately  $n = 20$  for each list) on a Likert scale (0 = “very unnatural” to 5 = “very natural”). The reliability of the ratings was similarly high for each list,  $ICC = .83$ – $.93$ . For English and Finnish, three employees of the University of Southampton and three employees of the University of Turku rated all sentences. The reliability of the ratings was acceptable, English:  $ICC = .75$ ; Finnish:  $ICC = .52$ . Averaged naturalness ratings are presented in Table 2 (fifth section). Ratings were generally high (between 4 and 5). More importantly, there were no significant effects in any of the languages, all  $F < 1$ .

Because the structure of the experimental sentences was highly standardized, 48 additional filler sentences were included in the materials. Sentences with varying structures (e.g., “Ella took the microphone and started to sing”) were drawn from the childLex corpus and

translated into all three languages using a similar translation/back-translation procedure as for the target sentences. However, no restrictions were imposed on word order or correspondence between phrases.

### **Apparatus**

The experimental setup was nearly identical at all three sites. Eye movements were recorded using a desktop-mounted Eyelink 1000 (SR Research, Canada), with a sampling rate of 1000 Hz and a spatial resolution of  $0.5^\circ$  visual angle. Viewing was binocular, but only the movements of one eye were recorded—usually the right eye, but the left eye was used if this improved the quality of the calibration. A chin rest was used to minimize head movements. Stimuli were presented on a 21-inch CRT (Southampton, Turku) or TFT monitor (Berlin), with a resolution of  $1,024 \times 768$  pixels at a viewing distance of 60cm. Sentences were presented in the middle of the screen in black on a white background in 20pt, double-spaced Courier New font. One character thus subtended approximately 0.5 degrees of visual angle. The experimental sentences usually covered one or two lines of text. The target words were not presented as the last or the first word on a line. Responses were collected using a Gamepad (Southampton, Berlin) or a Cedrus button box (Turku).

### **Procedure**

Children were assessed in autumn/winter in the first half of the school year. Testing took place in schools in an unused classroom or at the Max Planck Institute for Human Development. The study was formally approved by the ethics committee of the Max Planck Institute for Human Development and by the Berlin local education authority. Written consent was obtained from children's parents or guardians prior to testing.

Each child first completed the eye movement experiment. This began with a 9-dot calibration procedure, which was repeated until a precision of  $0.5^\circ$  of visual angle was achieved. Sentences were presented individually on the screen and children were instructed to read them silently for comprehension at their own pace. They were told that they would be asked

Yes/No comprehension questions on some of the sentences (see Appendix B for the questions). Children first read four practice sentences, each accompanied by a comprehension question, and they received feedback on their performance. After that, the calibration procedure was repeated, and children read the 48 experimental and 48 filler sentences in a randomized order. Before presentation of each sentence, children fixated a calibration point at the location where the first word in the sentence would appear. Children were recalibrated every 12 sentences or as necessary when the precision of the calibration check exceeded  $1^\circ$ . Children terminated each trial by pressing a response button. Twenty-five percent of the sentences were followed by a Yes/No question; these were rotated over all test sentences using a Latin square design. After the eye movement experiment, children completed the nonverbal intelligence test and the reading test.

Adults were assessed in eye movement laboratories at the three sites. They were told that they would read sentences that had been written for children, but that they should read them at their own pace for comprehension. All other aspects of the procedure were identical to the children's test session. A testing session lasted on average 40 minutes for children and 30 minutes for adults.

### **Data Analysis**

Data were pre-processed using the popEye package (version 0.6.3) implemented in R (version 4.0). popEye is an integrated environment to process and analyze eye-tracking data from reading experiments (Schroeder, 2019). During preprocessing, trials receiving less than 3 fixations or with misaligned fixations were discarded (16 of 7920 trials, corresponding to 0.2% of the overall data). In addition, fixations less than 80 ms were combined with the next fixation if they were within 1 character from each other and fixations of less than 40 ms were combined with the next fixation if they were within 3 characters from each other. Also, we removed four sentences in the English version because they contained spelling errors or the target word was the last or first word on a line. Moreover, we deleted all sentences comprising

one German target word, because it comprised only six letters but was mistakenly included as a long word. During this step, 2.5% of the data were excluded (199 trials). In order not to unnecessarily reduce the available sample size, we excluded only the individual problematic trials, not the corresponding sentences in the other languages or other sentences of the quadruple.

Data were analyzed separately for children and adults using (generalized) linear mixed effects models as implemented in the *lmer* package (Bates et al., 2014) in *R*. For the sentence-level analysis, participants and sentence were modeled as crossed random intercepts and Language (English vs. German vs. Finnish) served as a between-participants fixed factor. For the target word analysis, participants and words (nested in sentences) were modeled as crossed random intercepts. Language (English vs. German vs. Finnish) served as a between-participants fixed factor and Word Length (short vs. long) and Word Frequency (high vs. low) as within-participants factors. The effects of each factor were evaluated using Type II sum of squares as implemented in the *Anova* function of the *car* package (Fox & Weisberg, 2011). If necessary, post hoc comparisons were conducted using cell means coding and defining contrasts using the *glth* function in the *multcomp* package (Bretz et al., 2011). All reading time measures were log-transformed prior to the analyses. Effects were back-transformed in order to ease interpretation. Prior to each analysis, we removed values for the outcome variable deviating more than 2.5 *SDs* from either the person or item mean (removing between 1-2% of the data; see Baayen & Milin, 2010). We also conducted additional analyses using raw reading times or mixed-effect models using Gamma-distributed errors (Lo & Andrews, 2015). We also ran additional models without excluding any data. All these models showed the same qualitative pattern of effects and support the same theoretical conclusions.

## Results

### Sample Characteristics

Sample characteristics are provided in Table 3 for children and adults in all three languages. As can be seen, the three child samples were matched on the years of reading instruction received at the time of testing. As a consequence, there were differences in children's chronological age between the three languages. The English children were approximately 7 months younger than the German children, who were in turn approximately 13 months younger than the Finnish children. The English adults were on average 4 years younger than either the German or the Finnish adults, who did not significantly differ from each other. This disparity is due to differences between the structure of the academic system in the UK relative to Germany and Finland (where university starts later and student samples are usually more heterogeneous in age).

There were no significant differences across the three countries in participants' nonverbal intelligence scores for either children or adults. In addition, both children and adults showed age-appropriate reading performance and their percentile scores did not differ from the average score in any of the samples or between languages. Comprehension accuracy was high for both children,  $M=.907$ ,  $SD=.291$ , and adults,  $M=.972$ ,  $SD=.165$ , without any differences between languages. In sum, the analyses showed that our samples were well matched in each age group in terms years of reading instruction, gender composition, nonverbal intelligence, reading skill, and comprehension performance.

Table 3

*Sample Characteristics for Children and Adults*

Measure	Children				Adults			
	English	German	Finnish	<i>F</i> (2,87)	English	German	Finnish	<i>F</i> (2,72)
Sample size ( <i>N</i> )	30	30	30	0.0	25	25	25	0.0
Sex ( <i>N</i> female)	22	17	14	2.2	19	18	19	0.1
Age (years)	8.07 (0.69)	8.63 (0.67)	9.90 (0.31)	7.9***	19.60 (1.00)	23.92 (2.45)	24.44 (4.50)	19.4** *
Reading in- struction (years)	3.20 (0.50)	3.30 (0.0)	3.27 (0.37)	0.6	---	---	---	---
CFT ( <i>n</i> correct)	6.17 (2.45)	6.60 (2.01)	6.90 (2.38)	0.8	8.28 (2.13)	7.88 (1.83)	8.60 (2.47)	0.7
Word read- ing (norm val- ues)	48.23 (19.06)	54.40 (26.76)	48.00 (27.97)	0.6	41.80 (29.03)	51.00 (29.33)	---	1.2
Nonword reading (norm val- ues)	59.07 (21.33)	53.03 (29.56)	---	0.8	62.36 (22.99)	62.24 (25.17)	---	0.0
Comp. ac- curacy (% correct)	.89 (.31)	.91 (.29)	.92 (.27)	1.28	.98 (.13)	.98 (.13)	.96 (.21)	2.08

*Note.* *SDs* in parentheses. CFT = Culture Fair Intelligence Test (Catell, 1963).

### Sentence Reading Measures

We first ran a set of global analyses to examine overall indices of sentence interpretation and comprehension processes in both adults and children in the three languages. We analyzed mean sentence reading time, first-pass reading time per sentence, and rereading time per sentence, mean number of fixations per sentence, mean fixation duration and saccade length per sentence and mean skipping, refixation and regression probability aggregated over all words in the experimental sentences. These indices were computed for the experimental sentences separately. Prior to the analysis, sentences which were reread more than 10 times or received more than 50 fixations were removed (48 trials, corresponding to 0.5% of data). These values were determined post-hoc, based on the distributions of the two variables. These trials



were also not included in the following target word analysis. Overall, there were 2514 trials available for the English sample, 2512 trials for the German sample, and 2632 trials for the Finnish sample (corresponding to 95.2, 95.2, and 99.7% of all trials, respectively). Descriptive statistics for the experimental sentences and results from the corresponding (generalized) linear mixed-effects models are reported in Table 4.

Generally, the results replicated the typical developmental differences between children and adults in all languages, that is, adults read the sentences faster, made fewer and shorter fixations, had longer saccades, and showed higher word skipping but lower refixation rates.

In addition, the pattern in children's sentence reading time, first-pass reading time and rereading time, was quite clear. Finnish children generally read faster than either German or English children (who showed similar reading times). These overall differences might be related to the fact that Finnish children were older on average and generally read more fluently. We will return to this point in the Discussion section. In the remaining measures, English children showed a qualitatively different reading behavior than both German and Finnish children with longer fixation times, lower refixation probabilities, but higher skipping and regression rates. This finding replicates the pattern reported by Rau et al. (2016) that English children tend to show a more "explorer"-like reading style (characterized by higher word skipping rates, but also more regressions), while German children show a "plodder"-like pattern (characterized by lower word skipping rates and fewer regressions). The present study extends this finding by showing Finnish children behave more like German children in this respect.

Table 4

*Sentence Reading Measures for the Experimental Sentences for Children and Adults in the Three Languages (SDs are provided in parentheses).*

Measure	Children				Adults			
	English	German	Finnish	$\chi^2(2)$	English	German	Finnish	$\chi^2(2)$
Total sentence reading time (ms)	4879 <sup>a</sup> (2379)	4990 <sup>a</sup> (2291)	3764 <sup>b</sup> (1681)	15.25***	2086 <sup>a</sup> (765)	2442 <sup>b</sup> (969)	2059 <sup>a</sup> (823)	7.07*
First pass reading time (ms)	3590 <sup>a</sup> (1629)	3937 <sup>a</sup> (1724)	2862 <sup>b</sup> (1203)	16.78***	1852 <sup>a</sup> (578)	2142 <sup>b</sup> (724)	1704 <sup>a</sup> (584)	14.86**
Rereading time (ms)	1290 <sup>a</sup> (1395)	1053 <sup>a</sup> (1188)	901 <sup>b</sup> (1039)	7.33*	234 <sup>a</sup> (379)	300 <sup>a</sup> (464)	355 <sup>a</sup> (492)	2.75
Mean fixation duration per sentence (ms)	239 <sup>a</sup> (41)	221 <sup>b</sup> (33)	208 <sup>b</sup> (34)	18.60***	177 <sup>a</sup> (30)	182 <sup>a</sup> (23)	179 <sup>a</sup> (27)	1.60
Mean saccade length per sentence (characters)	6.02 <sup>a</sup> (2.02)	5.35 <sup>a</sup> (1.85)	6.05 <sup>a</sup> (1.94)	3.65	7.29 <sup>a</sup> (1.21)	7.31 <sup>a</sup> (1.99)	7.84 <sup>a</sup> (1.82)	3.24
Mean skipping probability (% words)	.27 <sup>a</sup> (.16)	.20 <sup>b</sup> (.19)	.16 <sup>b</sup> (.17)	25.44***	.34 <sup>a</sup> (.11)	.28 <sup>b</sup> (.15)	.24 <sup>b</sup> (.14)	21.84***
Mean refixation probability (% words)	.35 <sup>a</sup> (.19)	.41 <sup>b</sup> (.18)	.45 <sup>b</sup> (.20)	8.88*	.18 <sup>a</sup> (.14)	.21 <sup>a</sup> (.15)	.25 <sup>b</sup> (.19)	10.31**
Mean regression probability (% words)	.24 <sup>a</sup> (.15)	.19 <sup>b</sup> (.14)	.20 <sup>b</sup> (.16)	8.22*	.13 <sup>a</sup> (.12)	.13 <sup>a</sup> (.13)	.15 <sup>a</sup> (.14)	0.68

*Note.* Superscripts denote significant post-hoc differences between languages. Means with the same index do not significantly differ from each other (all  $|t| < 1.9$ ,  $p > .05$ ).

Table 5

*Sentence Reading Measures for the Filler Sentences for Children and Adults in the Three Languages (SDs are provided in parentheses).*

Measure	Children				Adults			
	English	German	Finnish	$\chi^2(2)$	English	German	Finnish	$\chi^2(2)$
Total sentence reading time (ms)	3233 <sup>a</sup> (1921)	3083 <sup>a</sup> (1181)	2489 <sup>b</sup> (1591)	8.61*	1409 <sup>a</sup> (555)	1522 <sup>a</sup> (563)	1501 <sup>a</sup> (587)	1.49
First pass reading time (ms)	2370 <sup>a</sup> (1224)	2376 <sup>a</sup> (1199)	1866 <sup>b</sup> (866)	10.25**	1253 <sup>a</sup> (446)	1346 <sup>a</sup> (452)	1213 <sup>a</sup> (451)	4.36
Rereading time (ms)	863 <sup>a</sup> (1291)	707 <sup>a</sup> (940)	623 <sup>a</sup> (814)	1.29	156 <sup>a</sup> (277)	176 <sup>a</sup> (304)	289 <sup>b</sup> (383)	7.57*
Mean fixation duration per sentence (ms)	240 <sup>a</sup> (48)	224 <sup>b</sup> (39)	215 <sup>b</sup> (39)	11.85***	180 <sup>a</sup> (32)	187 <sup>a</sup> (30)	189 <sup>a</sup> (38)	2.84
Mean saccade length per sentence (characters)	5.79 <sup>a</sup> (2.05)	5.14 <sup>a</sup> (1.86)	5.76 <sup>a</sup> (1.93)	3.30	6.97 <sup>a</sup> (1.45)	6.79 <sup>a</sup> (1.97)	7.19 <sup>a</sup> (1.89)	1.98
Mean skipping probability (% words)	.23 <sup>a</sup> (.16)	.16 <sup>b</sup> (.16)	.13 <sup>b</sup> (.16)	25.79***	.30 <sup>a</sup> (.12)	.24 <sup>b</sup> (.15)	.15 <sup>c</sup> (.14)	38.97***
Mean refixation probability (% words)	.35 <sup>a</sup> (.22)	.40 <sup>a</sup> (.22)	.50 <sup>b</sup> (.25)	13.53**	.18 <sup>a</sup> (.18)	.20 <sup>a</sup> (.18)	.30 <sup>b</sup> (.25)	16.57***
Mean regression probability (% words)	.24 <sup>a</sup> (.19)	.20 <sup>a</sup> (.18)	.23 <sup>a</sup> (.20)	3.13	.12 <sup>a</sup> (.15)	.13 <sup>a</sup> (.16)	.18 <sup>b</sup> (.18)	7.41*

*Note.* Superscripts denote significant post-hoc differences between languages. Means with the same index do not significantly differ from each other (all  $|t| < 1.9$ ,  $p > .05$ ).

Since the filler sentences consisted on average of shorter and fewer words than the target sentences (7.9, 7.5, and 5.6 words in English, German, and Finnish, respectively), we repeated the same set of analyses with the filler sentences to ensure that the pattern of effects generalizes to different sentence structures. The results of these analyses are reported in Table 5. As can be seen, the main pattern of effects was completely replicated. In particular, Finnish children had shorter total and first pass reading times than both English and German children. In addition, mean fixation duration and mean skipping probability decreased, and mean re-fixation probability increased with orthographic consistency. However, there were also two minor differences between the two analyses. Specifically, English children did not show higher regression rates for the filler sentences and, as a consequence, had similar rereading times than German and Finnish children. These discrepancies are presumably driven by the fact that the filler sentences were rather short and had very simple syntactic structures.

In contrast to children, cross-linguistic differences were generally less pronounced in adults' reading behavior in both the experimental and the filler sentences. Results showed that German adult readers generally had longer reading times than Finnish and English adult readers. These differences in overall reading speed might be related to the fact that 20% of the participants in the German sample were not university students, in contrast to the English and the Finnish sample. This difference might have been particularly important for the more challenging experimental sentences leading to the observed differences in first-pass reading time and total reading time.

There were also consistent cross-linguistic differences that occurred in both children and adults and these are likely to be related to overall linguistic differences between the three languages. First, Finnish readers had higher re-fixation rates than English readers with German readers falling in the middle the two extremes, although they were more similar to the English readers (see also Juhasz, 2008, who discusses systematic differences in re-fixation rates between English and Finnish adults' processing of compound words). Second, skipping rates

were highest for English readers and lowest for Finnish readers, again with German readers falling between the two other languages. Both differences can be attributed to differences in average word length between the three languages (i.e, average word length was shortest for English and longest for Finnish). The pattern also replicates the effects found by Liversedge et al. (2016) for Finnish and English adult readers and extends them to German adult readers.

Crucially, however, there was no indication of different reading styles in adult reading, suggesting that the explorer-like reading style relative to the plodder-like reading style seems to be a unique characteristic of developing readers. We will come back to this point in the General Discussion section.

### **Target Word Measures**

We ran a set of analyses on the target words to examine whether word length effects were moderated by frequency in a similar way in all three languages. We computed three standard eye movement measures for the target word in each sentence: First fixation duration (the duration of the first fixation on a word) was considered an early processing measure. By contrast, gaze duration (the sum of all first-pass fixations on a word) was considered a somewhat later measure, and beyond this, total reading time (the sum of all fixations on a word) was considered a late processing measure that reflects the total amount of time spent fixating a word to process it.

Prior to the analysis, trials with blinks before or after the target word (1109 trials, corresponding to 14.0% of the data), in which the target word was skipped during first-pass reading (604 trials, 7.8% of data) or fixated last in a trial (94 trials, 1.2% of the data) were removed. In addition, similar to the global analysis, target words that had very distant launch sites (> 16 letters; 333 trials, corresponding to 4.2% of the data), were reread more than 4 times (32 trials, 0.4% of the data) or received more than 10 fixations (54 trials, 0.8% of the data) were also removed. Overall, there were 1770 trials available for the English sample, 1634 trials for the

German sample, and 2030 trials for the Finnish sample (corresponding to 67.0, 61.9, and 76.9% of all trials, respectively).

Mean reading times for the remaining words are provided in Table 5 for children and adults. Inferential statistics from the corresponding linear mixed-effects models are reported in Table 6.

### *Children*

**First Fixation Duration.** The main effect of language was significant. First fixations in English,  $M=241$ ,  $SE=6$  ms, were substantially longer than in German,  $M=223$ ,  $SE=6$  ms, and Finnish,  $M=220$ ,  $SE=6$  ms, which did not differ from each other; English vs. German:  $t=2.17$ ,  $p=.030$ , English vs. Finnish:  $t=2.60$ ,  $p=.009$ , German vs. Finnish:  $t=0.41$ ,  $p=.683$ .

The length  $\times$  language interaction was significant. However, the length effect was not significant in any of the three languages; English:  $\Delta=-4$ ms,  $t=-1.46$ ,  $p=.144$ , German:  $\Delta=0$ ms,  $t=-0.11$ ,  $p=.911$ , Finnish:  $\Delta=6$  ms,  $t=1.58$ ,  $p=.114$ . The reason the interaction was significant was that the length effect in English and Finnish went in opposite directions.

The frequency  $\times$  language interaction was also significant: The frequency effect was not significant in English,  $\Delta=3$ ms,  $t=0.77$ ,  $p=.441$ , but it was significant in German,  $\Delta=11$ ms,  $t=2.69$ ,  $p=.007$ , and approached significance in Finnish,  $\Delta=7$ ms,  $t=1.82$ ,  $p=.069$ .

Finally, the length  $\times$  frequency  $\times$  language interaction was significant. Post-hoc-comparisons showed that the simple length  $\times$  frequency interaction approached significance in English,  $t=1.93$ ,  $p=.053$ , and was significant in German,  $t=-2.31$ ,  $p=.021$ , but non-significant in Finnish,  $t=0.50$ ,  $p=.615$ . In English, the length effect was only significant for low-frequency words,  $\Delta=-15$  ms,  $t=-2.35$ ,  $p=.019$ , but not for high-frequency words,  $\Delta=2$  ms,  $t=0.33$ ,  $p=.738$ . In German, by contrast, the length effect approached significance for high-frequency words,  $\Delta=-10$  ms,  $t=-1.71$ ,  $p=.087$ , but did not for low-frequency words,  $\Delta=9$  ms,  $t=1.54$ ,  $p=.123$ .

Table 6

*Mean Fixation Times (ms) for Target Words for Children and Adults*

	Children				Adults			
	High frequency		Low frequency		High frequency		Low frequency	
	Short	Long	Short	Long	Short	Long	Short	Long
<i>First Fixation Duration</i>								
English	242 (7)	244 (7)	247 (7)	232 (7)	188 (4)	185 (4)	193 (4)	196 (4)
German	222 (7)	213 (6)	224 (7)	233 (7)	174 (4)	176 (4)	172 (4)	184 (4)
Finnish	213 (6)	220 (6)	221 (6)	225 (7)	169 (4)	177 (4)	180 (4)	185 (4)
<i>Gaze Duration</i>								
English	320 (19)	412 (24)	347 (21)	401 (24)	201 (7)	228 (8)	213 (8)	231 (8)
German	281 (17)	350 (21)	303 (18)	511 (31)	184 (7)	211 (8)	196 (7)	229 (9)
Finnish	269 (17)	346 (21)	287 (18)	408 (31)	190 (7)	214 (8)	193 (7)	219 (8)
<i>Refixation Probability</i>								
English	.406 (.053)	.679 (.051)	.446 (.054)	.728 (.049)	.107 (.025)	.369 (.054)	.152 (.031)	.273 (.046)
German	.342 (.050)	.682 (.053)	.478 (.056)	.877 (.032)	.079 (.020)	.262 (.046)	.141 (.031)	.329 (.051)
Finnish	.301 (.054)	.610 (.053)	.354 (.048)	.728 (.046)	.128 (.020)	.280 (.046)	.117 (.031)	.315 (.051)
<i>Total Reading Time</i>								
English	423 (27)	523 (33)	541 (34)	555 (35)	226 (10)	244 (11)	240 (10)	249 (11)
German	360 (23)	406 (26)	473 (30)	639 (40)	194 (9)	225 (10)	218 (9)	266 (11)
Finnish	344 (21)	403 (25)	377 (23)	523 (32)	210 (9)	243 (10)	218 (10)	259 (11)

*Note.* Standard errors are provided in parentheses.

Table 7

*Results from Linear Mixed-Effects Models for the Three Dependent Measures for Children and Adults*

	Children				Adults			
	First	Gaze	Refixation Probability	Total	First	Gaze	Refixation Probability	Total
Language	7.95*	2.81	2.98	9.25**	8.93*	1.98	0.17	1.70
Frequency	4.35*	33.78***	25.67***	66.93***	21.23***	16.80***	1.18	20.42** *
Frequency × Language	6.58*	26.73***	9.02*	29.08***	2.66	2.21	5.30	5.70
Length	0.01	156.90***	248.85***	50.83***	9.24**	107.99***	107.00***	60.23** *
Length × Language	4.56	22.88***	5.55	9.38**	5.53	4.72	0.85	12.29**
Length × Frequency	0.02	8.55*	3.16	1.23	0.96	0.45	1.67	0.13
Language × Length × Frequency	9.51**	31.15***	1.58	25.04***	3.84	3.62	5.03	1.75

*Note.* \* $p < .05$ , \*\* $p < .01$ , \*\*\* $p < .001$ . Mixed-effects model with Language, Length, and Frequency as fixed effects and Participant, Sentence, and Word as random intercepts. Model evaluation using Type II sum of squares and Wald  $\chi^2$ -Tests.



Focusing on the simple frequency effects, post-hoc comparisons showed that there were no differences between languages for short words,  $\chi^2(2)=0.82$ ,  $p=.662$ : Frequency effects in all languages were not significant; English:  $\Delta=5$  ms,  $t=0.82$ ,  $p=.413$ , German:  $\Delta=2$  ms,  $t=0.27$ ,  $p=.784$ , Finnish:  $\Delta=8$  ms,  $t=1.64$ ,  $p=.102$ . By contrast, frequency effects varied between languages for long words,  $\chi^2(2)=15.28$ ,  $p<.001$ . The frequency effect did not achieve significance and was negative in English,  $\Delta= -12$  ms,  $t=1.92$ ,  $p=.055$ , significant and positive in German,  $\Delta= 21$  ms,  $t=3.51$ ,  $p<.001$ , and not significant in Finnish,  $\Delta= 5$  ms,  $t=0.93$ ,  $p=.351$ .

In sum, the length effects in first fixation duration, if present, were weak and inconsistent. In addition, the direction of the effects was reversed from the standard word length effect, as first fixations on long words were shorter than on short words. This does not indicate that long words are easier to process than short words. Instead, this inverse length effect is driven by a trade-off relationship between the number of fixations on a word and the duration of those fixations. If a word is processed using multiple fixations, each of them will be relatively short, but if the same word is processed in one fixation, it will be relatively long (see also Joseph et al., 2009).

**Gaze Duration.** Means and standard deviations for the four conditions are shown in Figure 1a for all three languages.

There were no overall differences in gaze duration between the three languages indicating that children in all languages spent a similar amount of time processing words during first-pass reading.

The main effect of length and the length  $\times$  language interaction were significant. Post-hoc comparisons showed that the length effect was significantly smaller in English,  $\Delta=74$ ms,  $t=6.65$ ,  $p<.001$ , than in both German,  $\Delta=132$  ms,  $t=12.08$ ,  $p<.001$ , and Finnish,  $\Delta=98$  ms,  $t=10.95$ ,  $p<.001$ .

The main effect of frequency and the frequency  $\times$  language interaction were also significant: The frequency effect was not significant in English,  $\Delta=10$  ms,  $t=0.89$ ,  $p=.372$ , but it was in German,  $\Delta=79$  ms,  $t=7.39$ ,  $p<.001$ , and in Finnish,  $\Delta=38$  ms,  $t=4.22$ ,  $p<.001$ .

Finally, the length  $\times$  frequency and length  $\times$  frequency  $\times$  language interactions were significant. Post-hoc comparisons showed that the length effect for high-frequency words was very similar for all three languages;  $\chi^2(2)=0.52$ ,  $p=.770$ ; English:  $\Delta=93$  ms,  $t=6.10$ ,  $p<.001$ , German:  $\Delta=69$  ms,  $t=5.06$ ,  $p<.001$ , Finnish:  $\Delta=77$  ms,  $t=6.47$ ,  $p<.001$ . By contrast, the length effect for low-frequency words differed substantially between languages,  $\chi^2(2)=47.11$ ,  $p<.001$ : It was significantly smaller in English,  $\Delta=54$  ms,  $t=3.37$ ,  $p=.001$ , than in Finnish,  $\Delta=121$  ms,  $t=9.02$ ,  $p<.001$ , and in German,  $\Delta=209$  ms,  $t=12.00$ ,  $p<.001$ .

Focusing on the simple frequency effects, post-hoc comparisons showed that there were no differences between languages for short words,  $\chi^2(2)=0.09$ ,  $p=.957$ : Frequency effects in short words approached significance in each language, English:  $\Delta=27$  ms,  $t=1.95$ ,  $p=.052$ , German:  $\Delta=21$  ms,  $t=1.70$ ,  $p=.090$ , Finnish:  $\Delta=19$  ms,  $t=1.73$ ,  $p=.084$ . By contrast, frequency effects varied between languages for long words,  $\chi^2(2)=53.60$ ,  $p<.001$ : The frequency effect was not significant and negative in English,  $\Delta= -12$  ms,  $t=0.68$ ,  $p=.494$ , but highly significant and positive in German,  $\Delta= 161$  ms,  $t=8.57$ ,  $p<.001$ , and in Finnish,  $\Delta= 62$  ms,  $t=4.23$ ,  $p<.001$ .

**Refixation Probability.** There were no overall differences in refixation probability between the three languages indicating that children refixated words similarly often.

The main effect of length was significant, indicating that long words,  $M=72.7$ ,  $SE=2.5$ , were refixated significantly more often than short words,  $M=38.6$ ,  $SE=2.7$ .

In addition, the main effect of frequency and the frequency  $\times$  language interaction were significant: The frequency effect was not significant in English,  $\Delta=4.8$  %,  $t=1.34$ ,  $p=.179$ , but it was in German,  $\Delta=20.6$  %,  $t=5.28$ ,  $p<.001$ , and in Finnish,  $\Delta=9.7$  %,  $t=2.86$ ,  $p=.004$ .

**Total Reading Time.** The main effect of language was significant: Total reading times were generally higher in English,  $M=508$  ms,  $SE=28$  ms, than in Finnish,  $M=406$  ms,  $SE=22$  ms, with German occupying an intermediate position,  $M=458$  ms,  $SE=25$  ms; English vs. German:  $t=1.36$ ,  $p=.174$ , English vs. Finnish:  $t=2.97$ ,  $p=.003$ , German vs. Finnish:  $t=1.60$ ,  $p=.109$ .

The main effect of length and the length  $\times$  language interaction were significant: The length effect was significant in all three languages, but it was substantially smaller in English,  $\Delta=60$  ms,  $t=3.18$ ,  $p=.001$ , than in German,  $\Delta=96$  ms,  $t=5.57$ ,  $p<.001$ , and in Finnish,  $\Delta=99$  ms,  $t=7.02$ ,  $p<.001$ .

The main effect of frequency and the frequency  $\times$  language interaction were also significant: The frequency effect was significant in all three languages, but it was substantially stronger in German,  $\Delta=168$  ms,  $t=9.69$ ,  $p<.001$ , than in either English,  $\Delta=77$  ms,  $t=4.12$ ,  $p<.001$ , or Finnish,  $\Delta=72$  ms,  $t=5.13$ ,  $p<.001$ .

Finally, the length  $\times$  frequency  $\times$  language interaction was significant. The length effect did not significantly differ for high-frequency words between languages,  $\chi^2(2)=2.29$ ,  $p=.319$ ; English:  $\Delta=100$  ms,  $t=4.08$ ,  $p<.001$ , German:  $\Delta=46$  ms,  $t=2.25$ ,  $p=.025$ , Finnish:  $\Delta=59$  ms,  $t=3.22$ ,  $p=.001$ . By contrast, there were substantial differences for low-frequency words,  $\chi^2(2)=32.74$ ,  $p<.001$ : The length effect was not significant in English,  $\Delta=14$  ms,  $t=0.47$ ,  $p=.639$ , but significant in German,  $\Delta=166$  ms,  $t=5.66$ ,  $p<.001$ , and in Finnish,  $\Delta=146$  ms,  $t=6.72$ ,  $p<.001$ .

Focusing on the simple frequency effects, post-hoc comparisons showed that there were differences between languages for short words,  $\chi^2(2)=12.15$ ,  $p=.002$ : Frequency effects were significant and strong in English,  $\Delta=118$  ms,  $t=4.70$ ,  $p<.001$ , and German,  $\Delta=114$  ms,  $t=5.22$ ,  $p<.001$ , but did not achieve significance in Finnish:  $\Delta=33$  ms,  $t=1.89$ ,  $p=.059$ . Similarly, frequency effects varied between languages for long words,  $\chi^2(2)=41.92$ ,  $p<.001$ . The frequency

effect was not significant in English,  $\Delta = 32$  ms,  $t = 1.13$ ,  $p = .260$ , but highly significant in German,  $\Delta = 233$  ms,  $t = 12.00$ ,  $p < .001$ , and in Finnish,  $\Delta = 121$  ms,  $t = 5.36$ ,  $p < .001$ .

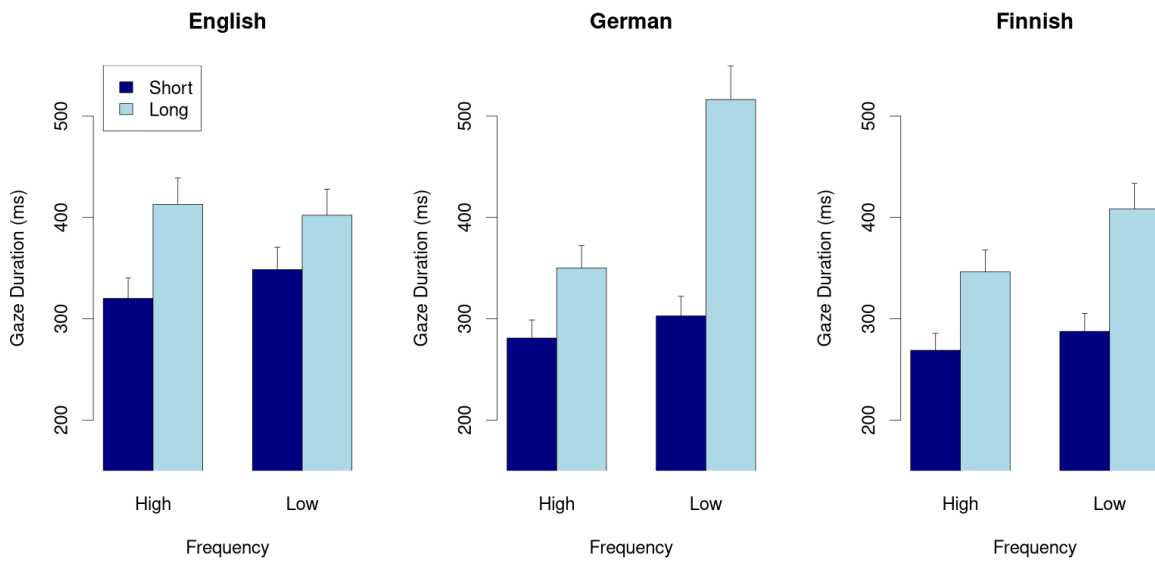
**Summary.** The target word analyses revealed that children in all languages spent a similar amount of time identifying words during first-pass reading. There were, however, some notable cross-linguistic differences in marker effects of word processing.

Table 8 summarizes our main findings. Here, children's simple length and frequency effects in each of the three languages and each outcome variable are shown next to each other. On the left side of the table, simple length effects are shown for high- and low-frequency words separately. In each row, the  $\chi^2$  indicates whether there were significant differences between languages. Results show that length effects for high-frequency words were very similar with no significant differences between languages for any of the outcome variables. Significant language differences only occurred for low-frequency words. Here, the size of the effect was modulated by orthographic consistency with larger length effects for German and Finnish than English.

The effects for refixation probability are particularly interesting. English children had similar length effects for high-frequency words as German and Finnish children. For low-frequency words, however, the length effect was smaller in English than in German or Finnish and of similar size as for the high-frequency words. Thus, English children did not appropriately adjust their refixation strategy for long, low-frequency words; that is, they refixated long words similarly, irrespective of their frequency. German and Finnish children, by contrast, showed particularly high refixation rates for long, low-frequency words.

Figure 1. Mean gaze duration for (a) children and (b) adults.

a) Children



b) Adults

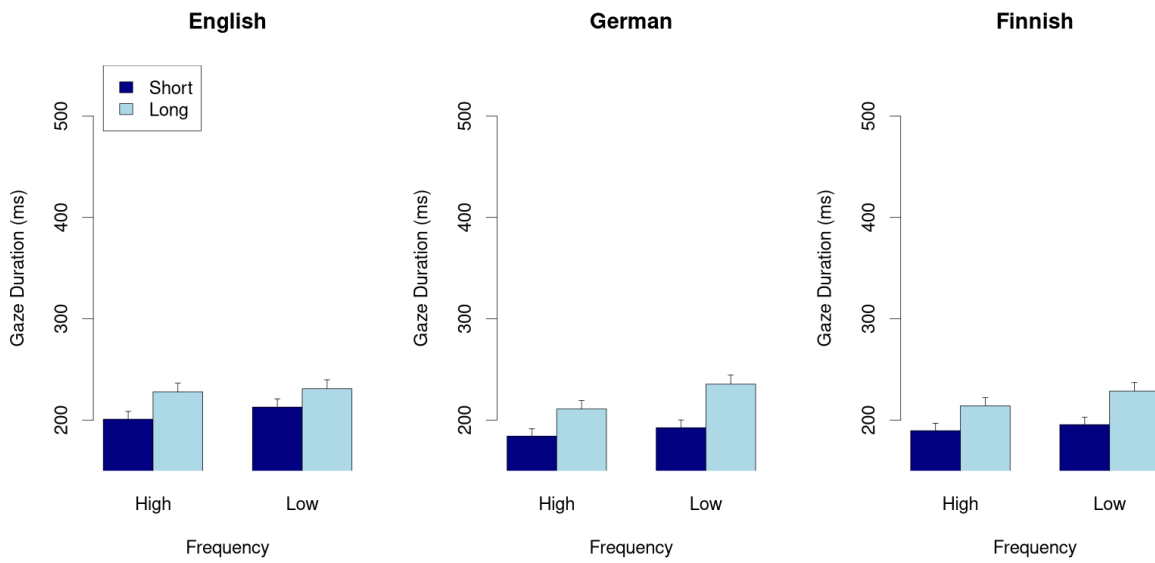


Table 8

*Children's Simple Length and Frequency Effects in all Languages (in ms for all Duration Measures and in % for Refixation Probability).*

Frequency	Length Effects				Frequency Effects				
	English	German	Finnish	$\chi^2(2)$	Length	English	German	Finnish	$\chi^2(2)$
					<i>First Fixation Duration</i>				
High	2 <sup>a</sup>	-10 <sup>a</sup>	7 <sup>a</sup>	5.32	Short	5 <sup>a</sup>	1 <sup>a</sup>	8 <sup>a</sup>	0.82
Low	<b>-15<sup>a</sup></b>	9 <sup>b</sup>	4 <sup>b</sup>	8.73*	Long	-12 <sup>a</sup>	<b>20<sup>b</sup></b>	5 <sup>c</sup>	15.28***
					<i>Refixation Probability</i>				
High	<b>27<sup>a</sup></b>	<b>34<sup>a</sup></b>	<b>31<sup>a</sup></b>	0.91	Short	4 <sup>a</sup>	<b>14<sup>a</sup></b>	5 <sup>a</sup>	2.24
Low	<b>28<sup>a</sup></b>	<b>40<sup>b</sup></b>	<b>37<sup>b</sup></b>	6.43*	Long	5 <sup>a</sup>	<b>20<sup>b</sup></b>	<b>12<sup>a</sup></b>	8.39*
					<i>Gaze Duration</i>				
High	<b>93<sup>a</sup></b>	<b>69<sup>a</sup></b>	<b>77<sup>a</sup></b>	0.52	Short	27 <sup>a</sup>	21 <sup>a</sup>	19 <sup>a</sup>	0.09
Low	<b>54<sup>a</sup></b>	<b>209<sup>b</sup></b>	<b>121<sup>c</sup></b>	47.11***	Long	-12 <sup>a</sup>	<b>161<sup>b</sup></b>	<b>62<sup>c</sup></b>	53.60***
					<i>Total Reading Time</i>				
High	<b>100<sup>a</sup></b>	<b>46<sup>a</sup></b>	<b>59<sup>a</sup></b>	2.34	Short	<b>118<sup>a</sup></b>	<b>114<sup>a</sup></b>	33 <sup>b</sup>	12.15**
Low	14 <sup>a</sup>	<b>166<sup>b</sup></b>	<b>146<sup>b</sup></b>	32.02***	Long	32 <sup>a</sup>	<b>233<sup>b</sup></b>	<b>121<sup>c</sup></b>	41.92***

*Note.* Significant effects ( $p < .05$ ) are printed in bold, effects approaching significance ( $p < .10$ ) in italics. Superscripts denote significant post-hoc differences between languages. Means with the same index do not significantly differ from each other (all  $t < 2.0$ ,  $p > .05$ ).

On the right side of the table, children's frequency effects are provided for short and long words separately. Again, a significant  $\chi^2$  value indicates that the frequency effects differed between languages. For short words, frequency effects were very similar with no significant differences between languages for most of the outcome measures. For long words, by contrast, frequency effects varied substantially between languages. In particular, German and Finnish children showed clear frequency effects in refixation probability and, as a consequence, in gaze duration and total reading time while English children did not. Thus, English children processed short words similarly as German and Finnish children, but they did not show consistent frequency effects for long words.

Taken together, these results indicate that the processing of short and high-frequency words is largely similar between the three languages. For long or low-frequency words, however, systematic cross-linguistic differences emerge with stronger length and frequency effects for both German and Finnish children who generally had similar effects. By contrast, English children showed a qualitatively different pattern and had particular difficulties in processing long, low-frequency words. This indicates that orthographic predictability affects eye movement development while the complexity of the writing system seems to be less important. We will come back to this point in the Discussion section.

### *Adults*

**First fixation duration.** The main effect of language was significant. First fixations in English were longer,  $M=191$ ,  $SE=4$ , than in German,  $M=177$ ,  $SE=34$  or Finnish,  $M=178$ ,  $SE=3$ , which did not differ from each other; English vs. German:  $t=2.72$ ,  $p=.007$ , English vs. Finnish:  $t=2.51$ ,  $p=.012$ , German vs. Finnish:  $t=0.22$ ,  $p=.823$ .

In addition, the main effects of length,  $\Delta=5$  ms, and frequency,  $\Delta=7$  ms, were both significant. The size of these effects did not differ between languages.

**Gaze duration.** Means and standard deviations for the four conditions are shown in Figure 1b for all three languages.

The main effects of word length,  $\Delta=29$  ms, and word frequency,  $\Delta=11$  ms, were both significant. Again, the size of these effects did not differ between languages.

### **Refixation Probability.**

There were no overall differences in refixation probability between the three languages indicating that adults refixated words similarly often.

Only the main effect of length was significant, indicating that adults in all three languages refixated long words,  $M=30.3$ ,  $SE=2.6$ , significantly more often than short words,  $M=11.9$ ,  $SE=1.4$ .

**Total reading time.** The main effect of length and the length  $\times$  language interaction were significant. The length effect in English,  $\Delta=14$ ms,  $t=2.12$ ,  $p=.034$ , was smaller than in both German,  $\Delta=36$ ms,  $t=5.59$ ,  $p<.001$ , or Finnish,  $\Delta=40$ ms,  $t=6.72$ ,  $p<.001$ . No other effects were significant.

**Summary.** In sum, adults in all languages showed stable word length and frequency effects for all reading time measures. Crucially, adults' reading behavior was remarkably similar across languages: There were few overall differences between languages and most interactions with the language factor were not significant. The contrast between this finding and the strong interactive effects observed for children is noteworthy.

## **Discussion**

In the present study, we investigated reading development in three alphabetic orthographies: English, German, and Finnish. In each language, children matched on years of reading instruction and reading ability as well as adults read identical stimulus materials in their native language. The length and frequency of target words embedded in closely matched sentences were orthogonally manipulated. The present study is the first eye-tracking study where developmental differences in children's silent reading in three different languages were comprehensively investigated.



With regard to overall developmental differences, we observed the typical pattern for all languages (Reichle et al., 2013; Schroeder et al., 2015, for an overview). Adults read the sentences faster than children, made fewer and shorter fixations, had longer saccades, and showed higher word skipping but lower refixation rates. Both children and adults showed reliable word frequency effects in all measures except refixation probability and strong word length effects in late reading measures. As expected, length and frequency effects were stronger for children than for adults, replicating earlier findings (Blythe et al., 2010; Joseph et al., 2009; Tiffin-Richards & Schroeder, 2015b).

The main aim of this study was to investigate cross-linguistic differences in children's and adults' eye movement behavior in well-matched participant groups and items. For children, we found clear and reliable differences between the three languages. First, as expected, the word length effect was stronger in German and Finnish than in English. This finding is in line with that of Rau et al. (2015), who also reported the length effect to be more pronounced in German than in English children. Our results extend this finding to Finnish and thus confirm that length effects are generally stronger in predictable than in unpredictable orthographies and not affected by orthographic complexity (Seymour et al., 2003).

Second, in contrast to our expectations, the word frequency effect was generally stronger in German and Finnish than in English. In particular, while frequency effects were similar in all languages for short words, English children did not show frequency effects for long words, in contrast to German and Finnish children. This effect was mainly driven by differences in refixation strategies by English children for long low-frequency words which were refixated similarly often as long high-frequency words. Thus, English children did not refixate long low-frequency words as would have been necessary in order to process them lexically. Thus, contrary to our prediction, English children did not rely more strongly on the lexical route. Instead, they seem to distribute their processing more widely across the sentence. In

line with this, total reading times were generally longer in English than in German and Finnish. Furthermore, English children were more likely to make regressions suggesting that they used the sentence context to further support their access of the meaning of low-frequency words (see also Rau et al., 2016). This fits with the view that reading development is delayed in unpredictable orthographies and visual word identification is more resource-demanding (Seymour et al., 2003). As a consequence, English children's processing is more likely to be influenced by the sentence context. This also fits with the overall idea that readers with less efficient word decoding skills need to rely more heavily on sentence context in order to identify words (see Stanovich, 1980, for a review).

Finally, we found reliable length by frequency interactions in German and Finnish with stronger length effects for low-frequency words than for high-frequency words. For German, this finding is very similar to the results of Rau et al. (2014) and Tiffin-Richards and Schroeder (2015a). In Finnish, length by frequency interactions have not previously been examined. The present study shows for the first time that German and Finnish children's eye behavior is very similar in this respect. In English, however, the interaction was less pronounced, a finding that contrasts with the results reported by Hyönä and Olson (1995) and Rau et al. (2015). It is important to note, however, that in these studies children read the texts aloud rather than silently. This is particularly relevant, because in oral reading words have to be pronounced overtly and sequentially, emphasizing a sublexical and linear reading strategy. Thus, it is not possible to use the upcoming sentence context to access the meaning of long, low-frequency words.

In general, the patterns of results found for German and Finnish children are consistent with the dual-route model of visual word recognition (Coltheart et al., 2001) and show that both the sublexical and the lexical route are well established in 4th grade readers. The two routes also interacted in the expected way with stronger word length effects for low- than for high-frequency words. However, our results are also consistent with the orthographic depth

hypothesis (Katz & Feldman, 1983), as we observed a different pattern for English children. As expected, they showed weaker length effects indicating that sublexical processing was less efficient (Ziegler et al., 2003). English children generally had problems identifying long words and did not show reliable frequency effects for long words. This indicates that word recognition is not yet sufficiently automatized to allow rapid lexical identification of long words, even if they are encountered relatively often. In line with this, English children seem to rely more strongly on the sentence context during reading, as evidenced by higher regression rates and longer rereading times in the global analyses. Although such context effects are generally outside the theoretical scope of models of single word recognition (but see Pritchard, Coltheart, & Castles, 2018), our results indicate that contextual information aids word identification in beginning readers which is particularly important in unpredictable orthographies.

Most importantly, as we have investigated children's eye-movements in three different orthographies, our results are informative with regard to the question of which aspects of a writing system are relevant to explain cross-linguistic differences. Our findings demonstrate that English children show a qualitatively different reading pattern than German or Finnish children when they process individual words. The observed differences between German and English children are consistent with the findings reported by Rau et al. (2015). One of our main findings is that the reading behavior of German children is very similar to that of Finnish children.

This allows us to form two important conclusions: First, as German and Finnish are rather similar in terms of their orthographic predictability but vary substantially in the complexity of the underlying GPC rules, our results thus suggest that orthographic complexity does not have a huge impact on reading development over and above orthographic predictability. Most GPC rules in Finnish involve only single letters, while German and English make abun-

dant use of position- and context-dependent multi-letter rules. Our results, thus, show that orthographic predictability, not complexity is the major factor driving children's eye movement development. Second, our findings also demonstrate that the syllabic complexity of a language does not have a strong impact on children's eye movements. Both English and German have rather complex syllable structures while Finnish syllables are usually open and do not have complex consonant clusters (e.g., Seymour et al., 2003). The finding that German and Finnish children showed very similar effects in the present study indicates that syllabic complexity does not strongly affect children's eye-movement development at the local word-level.

However, at the global level of sentence processing the picture is somewhat different: Finnish children were faster readers than German and English children, suggesting that they are more advanced in their reading development. German and English children, on the other hand, did not differ from each other in this respect. Together, this pattern of results indicates that there is an intricate interplay between the predictability of an orthography and its complexity when trying to explain differences in children's reading development. In our case, while the predictability of an orthography can explain the findings at the local level (German children patterning with Finnish children), the complexity of an orthography is a better predictor for the global findings (German children patterning with English children). Maybe, however, the difference in sentence reading may also be attributed to the fact that the children in the Finnish sample were slightly older than in the German and English samples.

This also points to a limitation of our study, which is that there were clear age differences between children in the three languages. As discussed in the Introduction, a particular methodological challenge in cross-linguistic studies is to ensure the comparability of the participants. This is particularly difficult for developmental studies as there are differences between the educational systems and the age when children enter school (5 years in England, 6 years in Germany, 7 years in Finland). Next to cultural and political factors, these differences are also related to the complexity of the orthographic systems; that is, English children start

school earlier because they need more time to crack the orthographic code, while children learning to read in Finnish can afford to enter school later as the orthographic system of the language is easier. This necessarily creates a confound between three important variables that can independently affect children's reading development: chronological age, the amount of reading instruction children have received (grade), and children's effective reading ability. Unfortunately, it is not possible to match children on all three variables at the same time. That is, when children are matched on chronological age (e.g., 8 years), they will necessarily attend different grades. Conversely, when children are matched on grade (e.g., grade 3), they will necessarily vary in age. Moreover, if children are matched on reading skill (e.g., on mean gaze duration as in Rau et al., 2016), they are likely to differ in both age and grade simultaneously.

We have opted to match children on the amount of reading instruction they have received which entails differences in chronological age across language groups. In particular, English children were approximately 7 months younger than the German children which were in turn 13 months younger than the Finnish sample. As children develop very rapidly at this age, it is likely that children also differed in other linguistic and non-linguistic variables such as print exposure, vocabulary size, or working memory. Unfortunately, it was not possible to include a measure of children's oral vocabulary in our study as there are no vocabulary tests available that allow cross-language comparisons. However, we did assess children's non-verbal intelligence as an overall indicator of their cognitive development. It is reassuring that the three samples did not differ significantly here, although non-verbal intelligence is admittedly a rather general construct and we cannot exclude the possibility that there were other relevant differences in cognitive skills that are related to the development of children's global reading skill.

At present, the relative importance of maturational vs. experience-specific variables on children's reading development is relatively unexplored. However, there is evidence from regression discontinuity designs (e.g., comparing children of different ages attending the same grade; Thistlethwaite & Campbell, 1960) that educational experience is more important than mere age-related factors (e.g., Ceci, 1991). The observation that Finnish and German children performed similarly in our study despite the fact that the Finnish children were over one year older is in line with this notion. In order to properly address this question, however, future studies should employ more advanced designs and sample children of different ages and in different grades and systematically compare their performance when matched on age, grade, or ability level.

In addition, future studies should investigate additional variables next to word length and word frequency that affect reading on the sentence level. One prime candidate would be the predictability of the upcoming word (Staub, 2015) as this variable is already explicitly included in existing models of eye movement control (such as E-Z Reader, Reichle, Warren, & McConnell, 2009; or SWIFT, Engbert, Nuthmann, Richter, & Kliegl, 2005). Moreover, it would be interesting to see whether English children make particular use of predictability information in order to process long, low-frequency words which would fit to their more global, explorer-like reading behavior (see Rayner et al., 2006).

In contrast to children, adults' reading behavior was remarkably consistent across languages. They showed stable main effects of word length and frequency in all eye-movement measures and a reliable length by frequency interaction in early processing. Crucially, none of the effects varied between orthographies. This indicates that adults' (in contrast to children's) eye movement behavior during reading is not strongly affected by the characteristics of their writing system, at least in alphabetic scripts. This similarity was most obvious in local processing of the target words. However, differences in global sentence reading times were also

rather minimal. Our data, thus, clearly favor a homogenization over a differentiation hypothesis in reading development. Children's reading behavior initially varies between languages but becomes continuously more automatized and more homogeneous during reading development. Thus, children's eye movement behavior is dominated by orthographic processing at first which, in turn, strongly depends on the predictability of a writing system. When children have learned to crack the orthographic code, higher-level and semantic factors become more important which are more similar between languages. This is consistent with previous cross-linguistic studies that found processing differences between children but not between adults (Feng et al., 2009; Rau et al., 2015).

This finding obviously stands in contrast to the results of isolated word recognition studies that have found cross-linguistic effects even in adults (e.g., Ziegler et al., 2001). There are two possible explanations for this discrepancy. First, as we aimed at investigating developmental effects, our stimulus materials were optimized for children. We therefore cannot exclude the possibility that our materials were too simple for the adults and not sensitive enough to detect orthographic effects in that population. However, we think this interpretation is unlikely, because we observed strong and reliable word length and frequency effects, similar to studies using materials written for adult readers (Liversedge et al., 2016). Second, and arguably more likely, there are obvious differences between the processing of single words versus sentences that might be responsible for the different pattern of findings. In particular, tasks such as word naming or lexical decision do not place much emphasis on semantic processing and do not involve linguistic processing beyond the level of the word (Kuperman et al., 2013; Seidenberg, 2011). It is therefore possible that orthographic differences at the word level are compensated by information from preceding and/or upcoming words in a sentence (e.g., via increased parafoveal preprocessing and/or stronger use of predictability information). Further research is needed to investigate this question.

In summary, this is the first eye-tracking study investigating children's reading behavior during normal reading in three different orthographies. Although the three languages differ substantially in their orthographic and linguistic properties, the experimental stimuli were directly comparable and allowed us to investigate differences in reading behavior not only at the sentence-level but also for well-controlled target words. In addition, participant groups were carefully matched for the amount of reading instruction they had received and showed typical reading performance for their age. We are thus able to draw firm conclusions about cross-linguistic differences in children's reading development. In particular, our results show that children's reading behavior is influenced by the nature of the orthography in which they are learning to read, while adults' reading behavior is rather homogenous across writing systems. In addition, our findings suggest that the predictability of the orthographic system is the main factor that affects children's reading development at the local level.



### References

- Baayen, R. H., Piepenbrock, R., & Gulikers, L. (1996). CELEX2 [CD-ROM]. Philadelphia, PA: Linguistic Data Consortium.
- Bates, D., Mächler, M., Bolker, B. M., & Walker, S. C. (2014). Fitting linear mixed-effects models using lme4. Advance online publication: <http://arxiv.org/abs/1406.5823>
- Blythe, H. I. (2014). Developmental changes in eye movements and visual information encoding associated with learning to read. *Current Directions in Psychological Science*, 23, 201-207. <http://dx.doi.org/10.1177/0963721414530145>
- Blythe, H. I., & Joseph, H. S. S. I. (2011). Children's eye movements during reading. In S. P. Liversedge, I. D. Gilchrist & S. Everling (Eds.), *The Oxford handbook of eye movements* (pp. 643-662). Oxford, UK: Oxford University Press. <http://dx.doi.org/10.1093/oxfordhb/9780199539789.013.0036>
- Blythe, H. I., Liang, F., Zang, C., Wang, J., Yan, G., Bai, X., & Liversedge, S. P. (2012). Inserting spaces into Chinese text helps readers to learn new words: An eye-movement study. *Journal of Memory and Language*, 67, 241-254. <http://dx.doi.org/10.1016/j.jml.2012.05.004>
- Blythe, H. I., Liversedge, S. P., Joseph, H. S. S. L., White, S. J., Findlay, J. M., & Rayner, K. (2006). The binocular coordination of eye movements during reading in children and adults. *Vision Research*, 46, 3898-3908. <http://dx.doi.org/10.1016/j.visres.2006.06.006>
- Blythe, H. I., Pagan, A., & Dodd, M. (2015). Beyond decoding: Phonological processing during silent reading in beginning readers. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. Advance online publication. <http://dx.doi.org/10.1037/xlm0000080>
- Blythe, H. I., Häikiö, T., Bertram, R., Liversedge, S. P., & Hyönä, J. (2011). Reading disappearing text: Why do children refixate words? *Vision Research*, 51, 84-92. <http://dx.doi.org/10.1016/j.visres.2010.10.003>

- Blythe, H. I., Liversedge, S. P., Joseph, H. S. S. L., White, S. J., & Rayner, K. (2009). Visual information capture during fixations in reading for children and adults. *Vision Research*, *49*, 1583-1591. <http://dx.doi.org/10.1016/j.visres.2009.03.015>
- Borgwaldt, S. R., Hellwig, F. M., & de Groot, A. M. B. (2005). Onset entropy matters – Letter-to-phoneme mappings in seven languages. *Reading & Writing*, *18*, 211-229. <http://dx.doi.org/10.1007/s11145-005-3001-9>
- Bretz, F., Hothorn, F., & Westfall, P. (2011). *Multiple comparisons using R*. Boca Raton, FL: Chapman & Hall.
- Caravolas, M., Lervag, A., Defior, S., Malkova, G. S., & Hulme, C. (2012). Different patterns, but equivalent predictors, of growth in reading in consistent and inconsistent orthographies. *Psychological Science*, *24*, 1398-1407. <http://dx.doi.org/10.1177/0956797612473122>
- Cattell, R. B. (1963). Theory of fluid and crystallized intelligence: A critical experiment. *Journal of Educational Psychology*, *54*, 1-22. <http://dx.doi.org/10.1037/h0046743>
- Ceci, S. J. (1991). How much does schooling influence general intelligence and its cognitive components? A reassessment of the evidence. *Developmental Psychology*, *27*, 703-722. <https://doi.org/10.1037/0012-1649.27.5.703>
- Chen, M., & Ko, H. (2011). Exploring the eye-movement patterns as Chinese children read texts: A developmental perspective. *Journal of Research in Reading*, *34*, 232-246. <http://dx.doi.org/10.1111/j.1467-9817.2010.01441.x>
- Coltheart, M., & Rastle, K. (1994). Serial processing in reading aloud: Evidence from dual-route models of reading. *Journal of Experimental Psychology: Human Perception and Performance*, *20*, 1197-1211. <http://dx.doi.org/10.1037/0096-1523.20.6.1197>
- Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. (2001). DRC: A Dual Route Cascaded Model of visual word recognition and reading aloud. *Psychological Review*, *108*, 204-256. <http://dx.doi.org/10.1037/0033-295X.108.1.204>

- Engbert, R., Nuthmann, A., Richter, E. M., & Kliegl, R. (2005). SWIFT: a dynamical model of saccade generation during reading. *Psychological review*, *112*, 777-813.  
<https://doi.org/10.1037/0033-295X.112.4.777>
- Feng, Miller, Shu, & Zhang, (2009). Orthography and the development of reading processes: An eye-movement study of Chinese and English. *Child Development*, *80*, 720-735.  
<http://dx.doi.org/10.1111/j.1467-8624.2009.01293.x>
- Fox, J., & Weisberg, S. (2011). *An R companion to applied regression* (2<sup>nd</sup> ed.). Thousand Oaks, CA: Sage.
- Grainger, J., & Ziegler, J. C. (2011). A dual-route approach to orthographic processing. *Frontiers in Psychology*, *2*, 1-13. <http://dx.doi.org/10.3389/fpsyg.2011.00054>
- Häikiö, T., Bertram, R., Hyönä, J., & Niemi, P. (2009). Development of the letter identity span in reading: Evidence from the eye movement moving window paradigm. *Journal of Experimental Child Psychology*, *102*, 167-181.  
<http://dx.doi.org/10.1016/j.jecp.2008.04.002>
- Häyrynen, T., Serenius-Sirve, S. & Korkman, M. 1999. *Lukilasse. Lukemisen, kirjoittamisen ja laskemisen seulontatestistö peruskoulun ala-asteen luokille 1–6*. [Lukilasse. Screening test for reading, writing and counting for the primary school grades 1-6.] Helsinki, Finland: Psykologien Kustannus Oy.
- Huestegge, L., Radach, R., Corbic, D., & Huestegge, S. M. (2009). Oculomotor and linguistic determinants of reading development: A longitudinal study. *Vision Research*, *49*, 2948-2959. <http://dx.doi.org/10.1016/j.visres.2009.09.012>
- 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*, 1430-1440. <http://dx.doi.org/10.1037/0278-7393.21.6.1430>
- Jincho, N., Feng, G., & Mazuka, R. (2014). Development of text reading in Japanese: An eye

movement study. *Reading & Writing*, 27, 1437-1465. <http://dx.doi.org/10.1007/s11145-014-9500-9>

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. <http://dx.doi.org/10.1016/j.visres.2009.05.015>

Joseph, H. S. S. L., Nation, K., & Liversedge, S. P. (2013). Using eye movements to investigate word frequency effects in children's sentence reading. *School Psychology Review*, 42, 207-222.

Juhasz, B. (2008). The processing of compound words in English: Effects of word length on eye movements during reading. *Language and Cognitive Processes*, 23, 1057-1088. <https://doi.org/10.1080/01690960802144434>

Karlsson, F. (1999). *Finnish: An essential grammar*. Routledge.

Katz, L., & Feldman, L. B. (1983). Relation between pronunciation and recognition of printed words in deep and shallow orthographies. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 9, 157-166. <http://dx.doi.org/10.1037/0278-7393.9.1.157>

Kuperman, V. & Bertram, R. (2013). Moving spaces: Spelling alternations in English noun-noun compounds. *Language and Cognitive Processes*, 28, 939-966. <http://dx.doi.org/10.1080/01690965.2012.701757>

Kuperman, V., Drieghe, D., Keuleers, E., & Brysbaert, M. (2013). How strongly do word reading times and lexical decision times correlate? Combining data from eye movement corpora and megastudies. *Quarterly Journal of Experimental Psychology*, 66, 563-580. <http://dx.doi.org/10.1080/17470218.2012.658820>

Kuperman, V., Stadthagen-Gonzalez, H., & Brysbaert, M. (2012). Age-of-acquisition ratings for 30,000 English words. *Behavior Research Methods*, 44, 978-990. <http://dx.doi.org/10.3758/s13428-012-0210-4>

Laine, M., & Virtanen, P. (1999). *WordMill lexical search program*. Center for Cognitive

Neuroscience, University of Turku.

Li, X., Bicknell, K., Liu, P., Wei, W., & Rayner, K. (2014). Reading is fundamentally similar across disparate writing systems: A systematic characterization of how words and characters influence eye movements in Chinese. *Journal of Experimental Psychology: General*, *143*, 895-913. <http://dx.doi.org/10.1037/a0033580>

Liang, F. F., Blythe, H. I., Bai, X. J., Yan, G. L., Li, X., Zang, C. L., & Liversedge, S. P. (2017). The role of character positional frequency on Chinese word learning during natural reading. *PloS One*, *12*, e0187656.

<https://doi.org/10.1080/20445911.2014.1000918>

Liang, F. F., Blythe, H. I., Zang, C. L., Bai, X. J., Yan, G. L., Li, X., & Liversedge, S. P. (2015). Positional character frequency and word spacing facilitate the acquisition of novel words during Chinese children's reading. *Journal of Cognitive Psychology*, *27*, 594–608. <http://dx.doi.org/10.1080/20445911.2014.1000918>

Liang, F., Ma, J., Bai, X., & Liversedge, S.P. (2021) Initial landing position effects on Chinese word learning in children and adults. *Journal of Memory and Language*. Advance online publication. <https://doi.org/10.1016/j.jml.2020.104183>

Liversedge, S. P., Drieghe, D., Li, X., Yan, G., Bai, X., & Hyönä, J. (2016). Universality in eye movements and reading: A trilingual investigation. *Cognition*, *147*, 1-20. <http://dx.doi.org/10.1016/j.cognition.2015.10.013>

Liversedge, S. P., Gilchrist, I. D., & Everling, S. (Eds.). (2011). *The Oxford handbook of eye movements*. Oxford, UK: Oxford University Press. <http://dx.doi.org/10.1093/oxfordhb/9780199539789.001.0001>

Mancheva, L., Reichle, E. D., Lemaire, B., Valdois, S., Ecalle, J., & Guérin-Dugué, A. (2015). An analysis of reading skill development using E-Z Reader. *Journal of Cognitive Psychology*, *27*, 657-676. <http://dx.doi.org/10.1080/20445911.2015.1024255>

- Milledge, S.V., Blythe, H.I. & Liversedge, S.P. (2021) Parafoveal pre-processing in children reading English: The importance of external letters. *Psychonomic Bulletin & Review*, 28, 197-208. <https://doi.org/10.3758/s13423-020-01806-8>
- Milledge, S. V., & Blythe, H. I. (2019). The changing role of phonology in reading development. *Vision*, 3(2), 23. <https://doi.org/10.3390/vision3020023>
- Moll, K. & Landerl, K. (2009). SLRT-II. Lese- und Rechtschreibtest. [SLRT-II. Reading and Writing test.] Göttingen, Germany: Hogrefe.
- Perfetti, C. A., & Harris, L. N. (2013). Universal reading processes are modulated by language and writing system. *Language Learning and Development*, 9, 296-316. <http://dx.doi.org/10.1080/15475441.2013.813828>
- Perry, C., & Ziegler, J. C. (2002). Cross-language computational investigation of the length effect in reading aloud. *Journal of Experimental Psychology: Human Perception and Performance*, 28, 990-1001. <http://dx.doi.org/10.1037/0096-1523.28.4.990>
- Perry, C., Ziegler, J. C., & Zorzi, M. (2007). Nested incremental modeling in the development of computational theories: The CDP+ model of reading aloud. *Psychological Review*, 114, 273-315. <http://dx.doi.org/10.1037/0033-295X.114.2.273>
- Pritchard, S.C., Coltheart, M., & Castles, A. (2018). A computational model of the self-teaching hypothesis based on the dual-route cascaded model of reading. *Cognitive Science*, 42, 722-770. <http://dx.doi.org/10.1111/cogs.12571>
- Rau, A. K., Moeller, K., & Landerl, K. (2014). The transition from sublexical to lexical processing in a consistent orthography: An eye-tracking study. *Scientific Studies of Reading*, 18, 224-233. <http://dx.doi.org/10.1080/10888438.2013.857673>
- Rau, A. K., Moll, K., Snowling, M. J., & Landerl, K. (2015). Effects of orthographic consistency on eye movement behavior: German and English children and adults process the same words differently. *Journal of Experimental Child Psychology*, 130, 92-105. <http://dx.doi.org/10.1016/j.jecp.2014.09.012>

- Rau, A. K., Moll, K., Moeller, K., Huber, S., Snowling, M. J., & Landerl, K. (2016). Same same, but different: Word and sentence reading in German and English. *Scientific Studies of Reading, 20*, 203-219. <http://dx.doi.org/10.1080/10888438.2015.1136913>
- Rayner, K., Reichle, E. D., Stroud, M.J., Williams, C.C., & Pollatsek, A. (2006). The effect of word frequency, word predictability, and font difficulty on the eye movements of young and older readers. *Psychology & Aging, 21*, 448-465. <http://dx.doi.org/10.1037/0882-7974.21.3.448>
- Reichle, E. D., Livversedge, S. P., Drieghe, D., Blythe, H. I., Joseph, H. S. S. L, White, S. J., & Rayner, K. (2013). Using E-Z Reader to examine the concurrent development of eye-movement control and reading skill. *Developmental Review, 33*, 110-149. <http://dx.doi.org/10.1016/j.dr.2013.03.001>
- Reichle, E., Warren, T., & McConnell, K. (2009). Using E-Z reader to model the effects of higher-level language processing on eye movements during reading. *Psychonomic Bulletin Review, 16*, 1-21. <http://doi:10.3758/PBR.16.1.1>
- Schmalz, X., Marinus, E., Coltheart, M., & Castles, A. (2015). Getting to the bottom of orthographic depth. *Psychonomic Bulletin Review, 22*, 1614-1629. <http://dx.doi.org/10.3758/s13423-015-0835-2>
- Schroeder, S. Hyönä, J., & Livversedge, S. P. (2015). Developmental eye-tracking research in reading: Introduction to the special issue. *Journal of Cognitive Psychology, 27*, 500-510.
- Schroeder, S., Würzner, K.-M., Heister, J., Geyken, A., & Kliegl, R. (2015a). childLex: A lexical database for German read by children. *Behavior Research Methods*. Advance online publication. <http://dx.doi.org/10.3758/s13428-014-0528-1>
- Schroeder, S., Würzner, K.-M., Heister, J., Geyken, A., & Kliegl, R. (2015b). childLex – Eine lexikalische Datenbank zur Schriftsprache für Kinder im Deutschen [childLex – a lexical database for German read by children.] *Psychologische Rundschau, 66*, 155-165.

<https://doi.org/10.1026/0033-3042/a000275>

Seidenberg, M. S. (2011). Reading in different writing systems: One architecture, multiple solutions. In P. McCardle, J. Ren, & O. Tzeng (Eds.), *Dyslexia across languages: Orthography and the gene-brain-behavior link* (pp. 151-174). Baltimore, MD: Brookes.

Seymour, P. H. K., Aro, M., & Erskine, J. M. (2003). Foundation literacy acquisition in European orthographies. *British Journal of Psychology*, *94*, 143-174.

<http://dx.doi.org/10.1348/000712603321661859>

Sperlich, A., Schad, D. J., & Laubrock, J. (2015). When preview information starts to matter: Development of the perceptual span in German beginning readers. *Journal of Cognitive Psychology*, *27*, 511-530. <http://dx.doi.org/10.1080/20445911.2014.993990>

Staub, A. (2015). The effect of lexical predictability on eye movements in reading: Critical review and theoretical implications. *Language and Linguistics Compass*, *9*, 311-327.

<https://doi.org/10.1111/lnc3.12151>

Sun, F., Morita, M. & Stark, L. W. (1985). Comparative patterns of reading eye movement in Chinese and English. *Perception & Psychophysics*, *37*, 502-506.

<http://dx.doi.org/10.3758/BF03204913>

Stanovich, K. E. (1980). Toward an interactive-compensatory model of individual differences in the development of reading fluency. *Reading Research Quarterly*, *16*, 32-71.

<http://dx.doi.org/10.2307/747348>

Tiffin-Richards, S. P. & Schroeder, S. (2015a). Children's and adults' parafoveal processing in German: Phonological and orthographic effects. *Journal of Cognitive Psychology*, *27*, 531-548. <http://dx.doi.org/10.1080/20445911.2014.999076>

Tiffin-Richards, S. P. & Schroeder, S. (2015b). Word length and frequency effects on children's eye movements during silent reading. *Vision Research*, *113*, 33-43.

<http://dx.doi.org/10.1016/j.visres.2015.05.008>



- Thistlethwaite, D. L., & Campbell, D. T. (1960). Regression-discontinuity analysis: An alternative to the ex post facto experiment. *Journal of Educational Psychology, 51*, 309-317.  
<https://doi.org/10.1037/h0044319>
- Torgesen, J. K., Wagner, R., & Rashotte, C. (1999). *TOWRE 2. Test of word reading efficiency* (2<sup>nd</sup> ed.). Kitchener, ON: Mind Resources.
- Verhoeven, L. & Perfetti, C. (Eds.). (2017). *Learning to read across languages and writing systems*. Cambridge, UK: Cambridge University Press.
- Verhoeven, L. & Perfetti, C. (2021). Universals in learning to read across languages and writing systems. *Scientific Studies of Reading*. Advance online publication.  
<https://doi.org/10.1080/10888438.2021>
- Vorstius, C., Radach, R., & Lonigan, C. J. (2014). Eye movements in developing readers: A comparison of silent and oral sentence reading. *Visual Cognition, 22*, 458-485.  
<http://dx.doi.org/10.1080/13506285.2014.881445>
- Weekes, B. S. (1997). Differential length effects in number of letters on word and nonword naming latency. *Quarterly Journal of Experimental Psychology A, 50*, 439-456.  
<http://dx.doi.org/10.1080/713755710>
- Zang, C., Liang, F., Bai, X., Yan, G., & Liversedge, S. P. (2013). Interword spacing and landing position effects during Chinese reading in children and adults. *Journal of Experimental Psychology: Human Perception and Performance, 39*, 720-739.  
<http://dx.doi.org/10.1037/a0030097>
- Zeno, S. M., Ivens, S. H., Millard, R. T., & Duvvuri, R. (1995). *The educator's word frequency guide*. Brewster, NJ: Touchstone Applied Science.
- Ziegler, J. C., & Goswami, U. (2005). Reading acquisition, developmental dyslexia, and skilled reading across languages: A psycholinguistic grain size theory. *Psychological Bulletin, 131*, 3-29. <http://dx.doi.org/10.1037/0033-2909.131.1.3>

Ziegler, J. C., Bertrand, D., Tóth, D., Csépe, V., Reis, A., Fásca, L. ... Blomert, L. (2010). Orthographic depth and its impact on universal predictors of reading: A cross language investigation. *Psychological Science, 21*, 551-559.

<http://dx.doi.org/10.1177/0956797610363406>

Ziegler, J. C., Perry, C., Jacobs, A. M., & Braun, M. (2001). Identical words are read differently in different languages. *Psychological Science, 12*, 379-384.

<http://dx.doi.org/10.1111/1467-9280.00370>

Ziegler, J. C., Perry, C., & Coltheart, M. (2000). The DRC model of visual word recognition and reading aloud: A extension to German. *European Journal of Cognitive Psychology, 12*, 413-430. <http://dx.doi.org/10.1080/09541440050114570>

Ziegler, J. C., Perry, C., Ma-Wyatt, A., Ladner, D., & Schulte-Körne, G. (2003). Developmental dyslexia in different languages: Language specific or universal? *Journal of Experimental Child Psychology, 86*, 169-193. [http://dx.doi.org/10.1016/S0022-0965\(03\)00139-5](http://dx.doi.org/10.1016/S0022-0965(03)00139-5)

## Appendix A

Table A1

*Target Words in the Three Languages*

Length	Frequency	Language				
		English	German	Finnish		
Short	High	aunt	Tante	täti		
		cabin	Hütte	mökki		
		goat	Ziege	vuohi		
		goose	Gans	hanhi		
		doll	Puppe	nukke		
		wine	Wein	viini		
		bay	Bucht	lahti		
		oven	Ofen	uuni		
		puppy	Welp	pentu		
		bike	Rad	pyörä		
		skirt	Rock	hame		
		rat	Ratte	rotta		
		Short	Low	widow	Witwe	leski
				ferry	Fähre	lossi
panda	Panda			panda		
ape	Affe			apina		
hint	Tipp			vihje		
rum	Rum			rommi		
haze	Dunst			usva		
bass	Bass			basso		
hare	Hase			jänis		
lever	Hebel			vipu		
jewel	Juwel			koru		
virus	Virus			virus		
Long	High			captain	Kapitän	kapteeni
				airplane	Flugzeug	lentokone
		elephant	Elefant	elefantti		
		emperor	Kaiser	keisari		
		policeman	Polizist	poliisi		
		fisherman	Fischer	kalastaja		
		drawing	Zeichnung	piirustus		
		compass	Kompass	kompassi		
		neighbour	Nachbar	naapuri		
		telephone	Telefon	puhelin		
		disease	Krankheit	sairaus		
		principal	Direktor	rehtori		
		Long	Low	preacher	Prediger	saarnaaja
				canteen	Kantine	ruokala
giraffe	Giraffe			kirahvi		
ballerina	Ballerina			ballerina		
allergy	Allergie			allergia		
drummer	Trommler			rumpali		
glacier	Gletscher			jäätikkö		
battery	Batterie			paristo		
hamster	Hamster			hamsteri		
monitor	Monitor			monitori		
diamond	Diamant			timantti		
scorpion	Skorpion			skorpioni		

## Appendix B

### Examples of Carrier Sentences and Comprehension Questions in the Three Languages

Four words from each set of target words were grouped together (here: *goat/panda/elephant/giraffe*). Each target word was embedded in a carrier sentence that was identical up to word  $n + 1$  for all four target words. Four different sentence frames were created for each set of target words. For each sentence, a Yes/No comprehension question (Q) was created (see text for further explanations).

#### Sentence Frame A

##### *English:*

1. The children were happy because the **goat** had given birth to three small kids.  
Q1. Did the goat give birth to six kids? (No)
2. The children were happy because the **panda** had carried its cubs out of its cave.  
Q2. Were the children sad? (No)
3. The children were happy because the **elephant** had sprayed water on the teacher.  
Q3. Did the elephant spray water on the teacher? (Yes)
4. The children were happy because the **giraffe** had eaten leaves out of their hands.  
Q4. Did the giraffe eat the leaves? (Yes)

##### *Finnish (literal translation provided in brackets):*

1. Lapset olivat iloisia, koska **vuohi** oli saanut kolme pientä poikasta.  
[Children were joyful because the goat was gotten three small kids.]  
Q1. Saiko vuohi kuusi poikasta?
2. Lapset olivat iloisia, koska **panda** oli kantanut pentunsa esille luolastaan.  
[Children were joyful because the panda was carried its puppies from its cave.]  
Q2. Olivatko lapset surullisia?
3. Lapset olivat iloisia, koska **elefantti** oli suihkuttanut vettä opettajan päälle.  
[Children were joyful because the elephant was sprayed water teacher's over.]  
Q3. Suihkuttiko elefantti vettä opettajan päälle?
4. Lapset olivat iloisia, koska **kirahvi** oli syönyt heidän ojentamiaan lehtiä.  
[Children were joyful because the giraffe was eaten their handed leaves.]  
Q4. Söikö kirahvi lehdet?

##### *German (literal translation provided in brackets):*

1. Die Kinder waren glücklich, denn die **Ziege** hatte drei kleine Junge bekommen.  
[The children were happy because the goat had three small kids gotten.]  
Q1. Hat die Ziege sechs Junge bekommen?
2. Die Kinder waren glücklich, denn der **Panda** hatte seine Jungen aus der Höhle hervorgeholt.  
[The children were happy because the panda had its cubs out of the cave carried out.]  
Q2. Waren die Kinder traurig?
3. Die Kinder waren glücklich, denn der **Elefant** hatte Wasser auf den Lehrer gespritzt.  
[The children were happy because the elephant had water onto the teacher sprayed.]  
Q3. Hat der Elefant den Lehrer mit Wasser bespritzt?

4. Die Kinder waren glücklich, denn die **Giraffe** hatte Blätter aus ihren Händen gefressen.  
[The children were happy because the giraffe had leaves from their hands eaten.]  
Q4. Hat die Giraffe die Blätter gegessen?

Sentence Frame B

*English:*

1. The trip to the zoo was funny because the **goat** had run into the teacher.  
Q1. Was the trip to the zoo sad? (No)
2. The trip to the zoo was funny because the **panda** had spit at the teacher.  
Q2. Did the panda spit at the teacher? (Yes)
3. The trip to the zoo was funny because the **elephant** had sprayed water on dad.  
Q3. Was the trip to the zoo funny? (Yes)
4. The trip to the zoo was funny because the **giraffe** had spit on the teacher's head.  
Q4. Was the trip to the zoo sad? (No)

*Finnish (literal translation provided in brackets):*

1. Retki eläintarhaan oli hauska, koska **vuohi** oli juossut opettajaa päin.  
[The trip to the zoo was funny because the goat was run of teacher against.]  
Q1. Oliko retki eläintarhaan ikävä?
2. Retki eläintarhaan oli hauska, koska **panda** oli sylkäissyt opettajaa päin.  
[The trip to the zoo was funny because the panda was spat of teacher against.]  
Q2. Sylkäisikö panda opettajaa päin?
3. Retki eläintarhaan oli hauska, koska **elefantti** oli suihkuttanut vettä isän päälle.  
[The trip to the zoo was funny because the elephant was sprayed water dad's over.]  
Q3. Oliko retki eläintarhaan hauska?
4. Retki eläintarhaan oli hauska, koska **kirahvi** oli sylkäissyt opettajaa päähän.  
[The trip to the zoo was funny because the giraffe was spat of teacher in head.]  
Q4. Oliko retki eläintarhaan ikävä?

*German (literal translation provided in brackets):*

1. Der Ausflug in den Zoo war lustig, denn die **Ziege** hatte den Lehrer umgerannt.  
[The trip to the zoo was funny because the goat had the teacher run over.]  
Q1. War der Ausflug in den Zoo blöd?
2. Der Ausflug in den Zoo war lustig, denn der **Panda** hatte mit dem Esel getanzt.  
[The trip to the zoo was funny because the panda had with the donkey danced.]  
Q2. Hat der Panda den Lehrer angespuckt?
3. Der Ausflug in den Zoo war lustig, denn der **Elefant** hatte Papa mit Wasser bespritzt.  
[The trip to the zoo was funny because the elephant had dad with water sprayed.]  
Q3. War der Ausflug in den Zoo lustig?
4. Der Ausflug in den Zoo war lustig, denn die **Giraffe** hatte dem Lehrer auf den Kopf gespuckt.  
[The trip to the zoo was funny because the giraffe had the teacher on the head spit.]  
Q4. War der Ausflug in den Zoo blöd?

Sentence Frame C

*English:*

1. Anna saw the animals in the zoo and the **goat** was really cute.  
Q1. Was the goat cute? (Yes)
2. Anna saw the animals in the zoo and the **panda** was the most interesting.  
Q2. Was the panda boring? (No)
3. Anna saw the animals in the zoo and the **elephant** was the most interesting.  
Q3. Did Anna see the animals in the zoo? (Yes)
4. Anna saw the animals in the zoo and the **giraffe** was the most interesting.  
Q4. Was the giraffe boring? (No)

*Finnish (literal translation provided in brackets):*

1. Anna näki kaikki eläintarhan eläimet, ja **vuohi** oli todella suloinen.  
[Anna saw all the zoo's animals and the goat was really cute.]  
Q1. Oliko vuohi suloinen?
2. Anna näki kaikki eläintarhan eläimet, ja **panda** oli kaikista kiinnostavin.  
[Anna saw all the zoo's animals and the panda was of all the most interesting.]  
Q2. Oliko panda tylsä?
3. Anna näki kaikki eläintarhan eläimet, ja **elefantti** oli kaikista kiinnostavin.  
[Anna saw all the zoo's animals and the elephant was of all the most interesting.]  
Q3. Näkikö Anna kaikki eläintarhan eläimet?
4. Anna näki kaikki eläintarhan eläimet, ja **kirahvi** oli kaikista kiinnostavin.  
[Anna saw all the zoo's animals and the giraffe was of all the most interesting.]  
Q4. Oliko kirahvi tylsä?

*German (literal translation provided in brackets):*

1. Anna sah die Tiere im Zoo und die **Ziege** war wirklich süß.  
[Anna saw the animals in the zoo and the goat was really cute.]  
Q1. War die Ziege süß?
2. Anna sah die Tiere im Zoo und der **Panda** war am interessantesten.  
[Anna saw the animals in the zoo and the panda was most interesting.]  
Q2. War der Panda langweilig?
3. Anna sah die Tiere im Zoo, und der **Elefant** war am interessantesten.  
[Anna saw the animals in the zoo and the elephant was most interesting.]  
Q3. Hat Anna die Tiere im Zoo gesehen?
4. Anna sah die Tiere im Zoo und die **Giraffe** war am interessantesten.  
[Anna saw the animals in the zoo and the giraffe was most interesting.]  
Q4. War die Giraffe langweilig?

Sentence Frame D

*English*

1. The bear was angry because the **goat** had bit him.  
Q1. Was the bear happy? (No)
2. The bear was angry because the **panda** had eaten his honey.  
Q2. Did the panda eat the bear's honey? (Yes)
3. The bear was angry because the **elephant** had kicked him.  
Q3. Was the bear happy? (No)
4. The bear was angry because the **giraffe** had spit on his head.  
Q4. Did the giraffe spit on the bear's head? (Yes)

*Finnish (literal translation provided in brackets):*

1. Karhu oli vihainen, koska **vuohi** oli purrut sitä.  
[The bear was angry because the goat was bitten of it.]  
Q1. Oliko karhu tyytyväinen?
2. Karhu oli vihainen, koska **panda** oli syönyt tämän hunajan.  
[The bear was angry because the panda was eaten its honey.]  
Q2. Söikö panda karhun hunajan?
3. Karhu oli vihainen, koska **elefantti** oli potkaissut sitä.  
[The bear was angry because the elephant was kicked of it.]  
Q3. Oliko karhu tyytyväinen?
4. Karhu oli vihainen, koska **kirahvi** oli sylkäissyt sitä päähän.  
[The bear was angry because the giraffe was spat of it in head.]  
Q4. Sylkäisikö kirahvi karhua päähän?

*German (literal translation provided in brackets):*

1. Der Bär war wütend, denn die **Ziege** hatte ihn gebissen.  
[The bear was angry because the goat had it bit.]  
Q1. War der Bär fröhlich?
2. Der Bär war wütend, denn der **Panda** hatte seinen Honig gegessen.  
[The bear was angry because the panda had its honey eaten.]  
Q2. Hat der Panda den Honig des Bären gegessen?
3. Der Bär war wütend, denn der **Elefant** hatte ihn getreten.  
[The bear was angry because the elephant had it kicked.]  
Q3. War der Bär fröhlich?
4. Der Bär war wütend, denn die **Giraffe** hatte ihm auf den Kopf gespuckt.  
[The bear was angry because the giraffe had it on the head spit.]  
Q4. Hat die Giraffe dem Bären auf den Kopf gespuckt?