UNIVERSITY OF AMSTERDAM

UvA-DARE (Digital Academic Repository)

Density and length in the neighborhood: Explaining cross-linguistic differences in learning to read in English and Dutch

Marinus, E.; Nation, K.; de Jong, P.F.

DOI
10.1016/j.jecp.2015.05.006

Publication date
2015

Document Version
Final published version

Published in
Journal of Experimental Child Psychology

Link to publication

Citation for published version (APA):
Marinus, E., Nation, K., & de Jong, P. F. (2015). Density and length in the neighborhood: Explaining cross-linguistic differences in learning to read in English and Dutch. *Journal of Experimental Child Psychology*, *139*, 127-147. https://doi.org/10.1016/j.jecp.2015.05.006

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: https://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

UvA-DARE is a service provided by the library of the University of Amsterdam (https://dare.uva.nl)

Download date:28 Nov 2021

Journal of Experimental Child Psychology 139 (2015) 127–147

Contents lists available at ScienceDirect

Journal of Experimental Child Psychology

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

Density and length in the neighborhood: Explaining cross-linguistic differences in learning to read in English and Dutch

Eva Marinus [a,*], Kate Nation [b], Peter F. de Jong [c]

[a] Department of Cognitive Science, ARC Centre of Excellence in Cognition and Its Disorders (CCD), Macquarie University, Sydney, Australia
[b] Department of Experimental Psychology, University of Oxford, Oxford OX1 3UD, UK
[c] Research Institute of Child Development and Education, University of Amsterdam, 1018 VZ Amsterdam, The Netherlands

ARTICLE INFO

Article history:
Received 7 August 2014
Revised 28 April 2015
Available online 20 June 2015

Keywords:
Cross-linguistic
Reading development
Length effects
Orthographic neighborhood size effects
Orthographic depth
Orthographic density

ABSTRACT

Two experiments examined underlying cognitive processes that may explain why it is harder to learn to read in English than in more transparent orthographies such as German and Dutch. Participants were English and Dutch readers from Grades 3 and 4. Experiment 1 probed the transition from serial to more parallel processing, as measured by the word length effect for words and pseudowords. English children took longer to make the transition to more parallel reading strategies for words than Dutch children. In contrast, Dutch children continued to use more serial reading strategies for pseudowords. Experiment 2 investigated children's sensitivity to the orthographic overlap between words, as measured by the size of orthographic neighborhood effects for words and pseudowords. Children reading Dutch showed greater sensitivity to the overlap between both words and pseudowords than English children. Cross-linguistic differences in the transition from serial to parallel reading strategies are discussed within the framework offered by the self-teaching hypothesis and the orthographic depth hypothesis. Finally, it is argued that differences between the two languages in the effect of orthographic neighborhood size are a result of cross-linguistic differences in orthographic density and not cross-linguistic differences in orthographic transparency.
© 2015 Elsevier Inc. All rights reserved.

* Corresponding author.
E-mail address: eva.marinus@mq.edu.au (E. Marinus).

http://dx.doi.org/10.1016/j.jecp.2015.05.006
0022-0965/© 2015 Elsevier Inc. All rights reserved.

Introduction

Cross-linguistic studies have consistently demonstrated that it is harder for children to learn to read in English than in other European languages (Landerl, Wimmer, & Frith, 1997; Patel, Snowling, & de Jong, 2004; Seymour, Aro, & Erskine, 2003). This finding has mainly been attributed to the fact that letter–sound correspondences in English are much less transparent than in other alphabetic languages, making it harder to decode and recognize words (e.g., Ziegler & Goswami, 2005). The aim of the current study was to go beyond the well-established finding that reading acquisition in English takes more time. Instead, we focused on how cross-linguistic differences in the reliability of letter–sound mappings affect the development of two processes thought to underlie the development of skilled word recognition, namely the transition from serial to parallel processing and sensitivity to overlap between words. To do this, we compared the naming speed of English and Dutch children as they read words and pseudowords that varied in length and neighborhood size.

The majority of relevant cross-linguistic experiments in the literature have compared children learning to read in English and German (Goswami, Ziegler, Dalton, & Schneider, 2003; Landerl et al., 1997). This is an interesting comparison; although the two languages differ in orthographic depth, both are Germanic languages and highly comparable in terms of syllable complexity (Ziegler, Perry, Ma-Wyatt, Ladner, & Schulte-Körne, 2003). For instance, they boast numerous different complex consonant clusters in both the onset and coda positions of words. The same holds for English versus Dutch. In addition, Dutch is even more similar to English with regard to the complexity of letter–phoneme alignment (van den Bosch, Content, Daelemans, & de Gelder, 1994) and the frequency of vowel digraphs. However, the languages differ markedly in the transparency of their letter-to-sound mappings (van den Bosch et al., 1994). In sum, although the English and Dutch orthographies are very similar, especially in comparison with very shallow orthographies with a simple syllable structure such as pointed Hebrew, Italian, and Finnish (Share, 2004), they differ on an important aspect of orthographic depth, namely the transparency of letter-to-sound mappings.

Given that English and Dutch vary in spelling–sound transparency, how might this influence the developmental time course of visual word recognition? Two contrasting perspectives can be gleaned from the literature. According to the self-teaching hypothesis (Share, 1995), successful phonological decoding facilitates the development of orthographic representations. Due to the greater transparency of Dutch, children learning to read Dutch are able to decode sooner and with greater ease than children learning to read English (Patel et al., 2004; Seymour et al., 2003). Thus, we would expect Dutch readers to make a more rapid transition to a more skillful mode of word recognition than those learning to read in English. An alternative view stems from the orthographic depth hypothesis (Frost, Katz, & Bentin, 1987; Share, 2004). This proposes that whereas readers of transparent languages continue to use nonlexical reading strategies (i.e., serial decoding), those reading in deep scripts are pressured to recruit lexical strategies given the less transparent spelling–sound mappings (Ziegler, Perry, Jacobs, & Braun, 2001). In line with this, Share (2004) found that children learning to read in pointed Hebrew—a very transparent script—remained insensitive to word-specific details for a relatively long time. On this view, we would expect English children to make a faster transition to direct word recognition than those learning to read Dutch.

To investigate and compare the development of visual word recognition in English and Dutch, we examined cross-linguistic differences in two experiments. In Experiment 1, we focused on the transition from serial to more parallel reading strategies, as indexed via the length effect, based on the assumption that the difference in naming times between longer and shorter words is a marker of serial processing (Marinus & de Jong, 2010b; Spinelli et al., 2005; Zoccolotti et al., 2005). In Experiment 2, we investigated children's sensitivity to the overlap between words, as indexed by the effect of neighborhood size on reading speed.

Length effects

To consider length effects first, if Dutch children make a faster transition from serial to parallel processing, as predicted by the self-teaching hypothesis, word length effects should be smaller in children learning to read Dutch compared with those learning to read English. In addition, the difference in the size of word length effect between relatively advanced and less advanced readers should be smaller in Dutch than in English. To test this prediction, we split each language sample into a group of relatively fast readers (indicating the use of more parallel reading strategies; e.g., Yap & van der Leij, 1993) and a group of relatively slow readers.

Only one cross-linguistic study has explored length effects in single-word reading in developing readers to date. Ziegler and colleagues (2003) compared dyslexic children who were learning to read in English or German with typical readers of the same age (chronological age controls) and younger readers at the same reading level (reading age controls). Both the English and German dyslexic children showed larger length effects (across words and pseudowords) in terms of naming times than the control groups and the magnitude of the difference between the dyslexic and control groups did not depend on language. Unfortunately, no comparisons were made of length effects across the two orthographies for the typically developing children. This was investigated by Rau, Moll, Snowling, and Landerl (2015), who monitored eye movements as children read aloud sentences containing words and pseudowords of different lengths. The difference in gaze duration between shorter and longer words was larger for the German children than for the English children, compatible with the conclusion that German children show a stronger length effect than children learning to read English. However, just like Ziegler and colleagues (2003), Rau and colleagues (2015) did not compare the performance of children of different age or reading ability, neither did they conduct separate analyses for words and pseudowords.

According to some theories of word reading and its development (e.g., dual route model: Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001; self-teaching hypothesis: Share, 1995), different cognitive mechanisms contribute to the reading words and pseudowords. Because these might have different time courses in development, it is appropriate to examine words and pseudowords separately when investigating the influence of reading proficiency. Based on the orthographic depth hypothesis and the finding that Dutch children do not use larger units in pseudoword reading (Marinus & de Jong, 2008), we predicted that length effects for the Dutch children for pseudowords would be stronger than those for the English children. The psycholinguistic grain size theory (Ziegler & Goswami, 2005) makes a similar prediction. According to the psycholinguistic grain size theory, the lack of transparency of the English letter-to-sound mapping system will shape the reading system in English such that young readers rely on larger units when reading, particularly body–rime units. In contrast, children reading more transparent languages, such as Dutch, rely on smaller units, particularly grapheme–phoneme units (Ziegler et al., 2001, 2003). Thus, although it differs from the orthographic depth hypothesis with regard to underlying cognitive processes, the psycholinguistic grain size theory would also predict larger length effects for pseudowords in Dutch than for pseudowords in English.

Neighborhood size

The orthographic neighborhood size of a given word (its N) is defined as all the existing words that can be created by replacing one of its letters (Coltheart, Davelaar, Jonasson, & Besner, 1977). Although children learning to read in English (Laxon, Coltheart, & Keating, 1988) and in Dutch (Marinus & de Jong, 2010a) are faster and more accurate at naming words that have more neighbors, there has been no direct cross-linguistic comparison of neighborhood effects in developing readers. Based on the orthographic depth hypothesis, which postulates that children learning to read in a deep orthography such as English rely more on lexical reading strategies (e.g., analogy with other words) than children learning to read a more transparent orthography such as Dutch or German, we predicted that sensitivity to the overlap between words would be stronger for English children than for Dutch children.

It is important to stress that the orthographic N-size metric, used for this study, is different from the body N-size metric used in an earlier study by Ziegler and colleagues (2003). Body N counts as neighbor words that share an orthographic rime (e.g., *bat* and *cat*) and, in contrast to orthographic N, words that differ in length are neighbors so long as the rime unit is shared (e.g., *beam* and *stream*). Ziegler and colleagues found that English children showed larger body N-size effects than German children. This is consistent with the idea that children learning to read in English are more sensitive to larger units than children learning to read in German. Note, however, that this does not necessarily mean that English children are more sensitive to orthographic overlap between words than their German peers. Ziegler and colleagues interpreted their results as evidence for the psycholinguistic grain size theory, namely that English children use larger (body) units in reading, whereas German children stick to smaller units (see also Ziegler & Goswami, 2005). However, there is another factor, independent of orthographic depth and grain size, that may be underlying cross-linguistic differences in orthographic neighborhood size effects. According to the CELEX database (Baayen, Piepenbrock, & van Rijn, 1993), German three-letter and four-letter words have on average fewer neighbors and body neighbors than English words.[1] This could explain why German children showed less facilitation by body neighbors. Dutch words have on average more neighbors than English words (see Norris & Kinoshita, 2012, p. 533, Fig. 10, data drawn from the British Language Project and Dutch Language Project databases and also statistics from the CELEX database). In other words, at the level of monosyllables, Dutch orthography is denser than English. If it is not the use of larger units that is driving the differences between the two languages but rather differences in orthographic density, Dutch children should show more facilitation by orthographic neighbors than English children. We tested this prediction in Experiment 2.

Experiment 1

This experiment examined the effect of length on word and pseudoword reading in children learning to read in Dutch or English. Cross-linguistic experiments bring with them the challenge of how to match participants and items across languages. Ziegler and colleagues (2003) matched English and German children for chronological age. Inevitably, however, this matching strategy leads to big differences across the two groups in amount of formal reading instruction. In the Netherlands, as in Germany, formal instruction starts during the year that children turn 7 years old; by that time, English children have already had more than a full year of reading instruction. Studies on the effects of schooling suggest that the number of years of instruction is a more reliable predictor of reading skill than age (e.g., Christian, Morrison, Frazier, & Massetti, 2000; Cunningham & Carroll, 2011). Therefore, we chose to match the two language groups on number of years of reading instruction rather than age, meaning that the Dutch children were older than the English children. We also matched the English and Dutch children on relative reading proficiency; each English child was matched to a Dutch child with the same relative proficiency (see Method for details). Another methodological difference between our experiment and the experiments reported by Ziegler and colleagues (2003) concerns the items. Ziegler and colleagues used cognates (e.g., *tea* and *Tee*) so that most words were matched across languages on first letter, syllable complexity, and meaning. However, cognates are not matched for frequency and neighborhood size, both of which are variables that are potential confounds for length effects. Therefore, we matched words on frequency, taking overall frequency differences in Dutch and English into account (Ellis & Hooper, 2001; Ellis et al., 2004).

Method

Participants

The Dutch and English samples each consisted of 44 children: 24 from Grade 3 (12 boys and 12 girls) and 20 from Grade 4 (7 boys and 13 girls). The English children were recruited from six primary schools in Oxford, United Kingdom. The Dutch children came from two large primary schools in Oostzaan, The Netherlands. The children were tested during the first half of the school year (October–December). The Dutch and English children had received the same amount of reading instruction (2.3 years for Grade 3 and 3.3 years for Grade 4), but the English children were on average 1 year younger (M = 8.24 years, SD = 0.50) than the Dutch children (M = 9.10 years, SD = 0.70) because primary education in the United Kingdom starts 1 year earlier when children enter the reception year.

[1] Average orthographic neighborhood size for monosyllabic words for word lengths of three, four, and five letters in English: 12.3, 8.8, 4.4; in German: 5.0, 5.5, 5.02; in Dutch: 16.1, 9.8, 5.2.

We could not perfectly control for reading instruction method and focus; however, both English and Dutch schools used a phonics instruction approach to teach reading in Grade 1.

The English children were recruited and tested first, and so the Dutch children were matched to them. To this end, 154 Grade 3 and 4 Dutch children were screened on tests of single-word reading, receptive vocabulary, and nonverbal reasoning. From this group, 44 children were selected to match the English children. Matching was based on similar position for each language group in grade-norm scores on single-word reading performance, age-norm scores in receptive vocabulary and nonverbal reasoning, grade (3 or 4), and gender. Descriptive data are shown in Table 1.

Single-word reading performance was measured with the Sight Word subtest (Version A) of the Test of Word Reading Efficiency (TOWRE; Torgesen, Wagner, & Rashotte, 1999) and Version A of the Dutch One-Minute Test (Brus & Voeten, 1995). In both tests, words are presented in lists and the participants are asked to read aloud as many words as quickly as they can within a set time (45 s for the English test and 60 s for the Dutch test). Receptive vocabulary was measured with the British Picture Vocabulary Scale (Dunn, Dunn, Whetton, & Pintilie, 1982) and the Vocabulary subtest of the RAKIT, a Dutch intelligence test battery for children (Bleichrodt, Drenth, Zaal, & Resing, 1987). Both tests contain words of increasing difficulty, and the children are asked to choose the corresponding picture out of four alternatives. To compare performance across the two groups, raw scores were converted to standard scores ($M = 100$, $SD = 15$) using the standardization data provided in the test manuals.

Finally, nonverbal reasoning was assessed in the English children with the Matrix Reasoning subtest from the Wechsler Intelligence Scale for Children (Wechsler, 1992) and in the Dutch sample with Raven's Standard Progressive Matrices (Raven, Court, & Raven, 1986). In both tests children choose a pattern from a set of answer options to complete a series of patterns. There are no norms for the Dutch

Table 1
Means (and standard deviations) of the reading-level variable and matching variables of the relatively slow and relatively fast English and Dutch readers.

Language	Reading group, (*n*), and number of children per grade	Reading level and matching variables				
		Word reading speed[a]	Word reading standard score	Receptive vocabulary standard score	Nonverbal reasoning[b]	Age (years)
English	Overall ($n = 44$) Grade 3 $n = 24$, Grade 4 $n = 20$	60.8 (8.6)	103.6 (9.2)	108.4 (11.4)	12.1 (2.8)	8.2 (0.5)
	Relatively slow ($n = 19$) Grade 3 $n = 13$, Grade 4 $n = 6$	52.7 (5.6)	97.0 (5.0)	107.0 (9.7)	12.6 (2.7)	8.1 (0.5)
	Relatively fast ($n = 25$) Grade 3 $n = 11$, Grade 4 $n = 14$	67.0 (4.1)	108.6 (8.5)	110.6 (12.7)	11.8 (3.0)	8.2 (0.5)
Dutch	Overall ($n = 44$) Grade 3 $n = 24$, Grade 4 $n = 20$	63.8 (9.2)	104.9 (9.7)	109.2 (15.1)	38.8 (5.5)	9.1 (0.6)
	Relatively slow ($n = 19$) Grade 3 $n = 13$, Grade 4 $n = 6$	55.4 (5.2)	98.2 (6.7)	110.4 (14.4)	38.7 (5.1)	9.0 (0.7)
	Relatively fast ($n = 25$) Grade 3 $n = 11$, Grade 4 $n = 14$	70.2 (5.8)	110.0 (8.5)	108.3 (15.8)	39.4 (5.8)	9.1 (0.5)

[a] For the English children, this is the number of TOWRE sight words read within 45 s. For the Dutch children, this is the average number of One-Minute Test words read in 60 s.
[b] Scaled scores ($M = 10$, $SD = 3$) for the English sample; raw scores for the Dutch sample.

test available; however, the mean raw scores of the Grade 3 and 4 children selected were slightly above the mean raw score of approximately 1000 Dutch Grade 4 children used in reading research at the University of Amsterdam ($M = 37.92$, $SD = = 7.47$; see also Marinus, 2010). This complements the performance of the English children according to population norms.

Children from both language groups were split into 19 relatively slow readers and 25 relatively fast readers based on a mean split on the *raw score* on the single-word reading task. Assignment to reading group was based on raw scores and not on year group or norm scores because we were interested in examining cross-linguistic differences in reading strategies at different developmental levels of reading. The mean raw reading scores of the relatively slow readers (±53 words per 45 s in English; ±55 words per 60 s in Dutch) map onto the average reading speed of children during the first half of Grade 3. The mean raw reading of the relatively fast readers (±67 words per 45 s in English; ±70 words per 60 s in Dutch) map onto the average reading speed of children during the first half of Grade 4. We refer to these groups as the relatively slow and relatively fast readers.

Materials

Matching word sets across language

One of the challenges in conducting cross-linguistic research is to match item sets across languages on factors such as length, frequency, and syllable structure (Share, 2008). Previous research (mostly comparing English and German) has tackled this in part by using cognates or even identical words (Landerl et al., 1997; Ziegler et al., 2003). Although this approach does ensure that the item sets are equivalent in terms of first letter, length, word structure, and meaning, it does not automatically control for other important factors that are known to influence word reading speed, particularly frequency and neighborhood size.

Matching for frequency across languages is not straightforward because word corpora (especially in different languages) vary in size. Ellis and Hooper (2001) dealt with this by creating frequency-matched lists compiled by sampling words from 100 successive strata of decreasing written word frequency for the languages they were comparing in their study (English and Welsh). In the current study, we used a slightly different but comparable approach.

To create Dutch item sets that were matched for frequency to the English items, we first extracted all one- to five-letter monosyllabic words from the Children's Printed Word Database (Masterson, Stuart, Dixon, & Lovejoy, 2003), the word corpus that was used to select the items for the English experiment. Next, we calculated the natural log of the frequency of each monosyllabic English word and then converted these natural log frequency scores into *Z*-scores. After this, we looked up the *Z*-scores for every item in the English experiment. Next, we followed the same procedure for Dutch monosyllabic words by extracting all one- to five-letter monosyllabic words from a Dutch word corpus of child literature (Schrooten & Vermeer, 1994) and converting the frequencies of all monosyllabic words into *Z*-scores of the natural logarithm of the frequency. Finally, we selected Dutch words matching the standardized frequency scores of the English words.

The length manipulation

The length task consisted of a word block and a pseudoword block, each with three-, four-, and five-letter items. Each length condition comprised 12 items. Within each block, the words and pseudowords were presented in random order. Half of the children started with the word block, whereas the other half first read the pseudoword block.

Words were matched across length conditions and language on first letter (except for a few items in the three-letter condition), word structure (CVC, CCVC, and CCVCC, where C is consonant and V is vowel), and frequency. It was impossible to match for orthographic neighborhood size because in both languages three-letter words have many more neighbors than four- and five-letter words. Pseudowords were matched across length conditions on first letter, word structure, and orthographic neighborhood size. Across languages, it was not possible to match pseudowords for neighborhood size because Dutch monosyllabic pseudowords also have more neighbors than their English counterparts. The reported N sizes (see Table 6 in Discussion) were taken from the Dutch and English CELEX database. See Appendix A for the word and pseudoword sets.

Procedure

The experiment was programmed and run in E-Prime (Version 1.0) (Schneider, Eschman, & Zuccolotto, 2002). Items were presented one at a time in the middle of a 14.1-inch XGA LCD screen of a D600 Pentium-M 1.3-GHz white background. A fixation point (+) was projected in the middle of the screen, and 750 ms later an item appeared printed in Arial font 46. The stimulus disappeared from the screen as soon as the voice key was triggered. Naming latencies (time between appearance of the stimulus and the onset of the voice key) were registered, and the tester labeled the responses (correct, incorrect, or invalid) via a serial response box. Children were instructed to read aloud the words and pseudowords as quickly as they could without making mistakes. The task started with 6 to 10 practice trials.

The English children were tested individually in two sessions of 30 min each in a quiet room at their school. In the first session, the children completed the nonverbal reasoning task, the single-word reading task, and Experiment 1 (words and pseudowords). Experiment 2 and receptive vocabulary were completed in the second session.

Following screening, the Dutch children were tested individually in a quiet room in their school in a session of 30 min in which they completed Experiment 1 and Experiment 2, respectively.

Results

Method of data analysis

The data were analyzed with multilevel models with MLwiN 2.12 (Rasbash, Steele, Browne, & Prosser, 2004). In a word reading experiment like this, dependent variables are embedded in a two-level hierarchical structure with responses to items (Level 1) nested under participants (Level 2). Within a multilevel model, random factors from items and participants can be tested within one model, meaning that separate analyses across both participants and items are not necessary. In addition, the model has more statistical power than a standard analysis of variance (ANOVA) because response times (RTs) to all items are used. Moreover, it is robust against missing data because parameters are estimated based on all available responses to items instead of means per condition. Finally, it allows for the effect of unmatched stimulus variables to be controlled (Quené & van den Berg, 2004).

RTs were estimated via dummy variables that were specified for the separate within-participant conditions of *length* (three, four, or five letters) and *lexicality* (words or pseudowords) and for the between-participants variables of *reading level* (relatively slow or relatively fast) and *language* (English or Dutch). Because the condition variables within each reading level by language group were highly correlated, we also estimated the covariances to get a more reliable estimate of mean RT. We also estimated between-participants and within-participant variability. The estimated variances and covariances can be obtained from the first author.

Several contrasts were specified to test mean differences in latency. The differences were subsequently tested with the chi-square test statistic. This approach is comparable to the M-MATRIX and L-MATRIX options that can be used in the General Linear Model function in SPSS. The contrasts are described in more detail below.

Data cleaning

Invalid responses were excluded. Latencies were considered invalid when the response was too quick (<325 ms) or too slow (>6000 ms), when the response was caused by a voice key error (when a sound other than a naming response triggered the microphone), when self-corrections were made, and when the time to respond was more than 2 standard deviations from a child's individual mean score. The deviation scores were calculated separately for words and pseudowords and per length condition. The percentage of invalid latencies for the English children was 9.9% for words and 9.3% for pseudowords. For the Dutch children, it was 10.2% for words and 12.9% for pseudowords.

Mean latency scores and accuracy rates for the valid trials per length and lexicality condition for the faster and slower English and Dutch readers are presented in Table 2. Because the latency data were strongly skewed to the right (which is typically the case for reaction time data; see Ratcliff, 1993), RTs were converted into number of items (words/pseudowords) read per second (invRT transformation; Baayen & Milin, 2010) before the data were subjected to the multilevel analyses. An

Table 2
Mean latency scores (in ms) and accuracy rates of the relatively slow and relatively fast English and Dutch readers for words and pseudowords in every length (three, four, or five letters, L3–L5) condition.

	Latencies (ms)							
	Relatively slow readers				Relatively fast readers			
	Words		Pseudowords		Words		Pseudowords	
	English	Dutch	English	Dutch	English	Dutch	English	Dutch
L3	678 (196)	568 (87)	828 (313)	676 (174)	587 (139)	547 (80)	695 (438)	607 (113)
L4	734 (324)	578 (133)	930 (416)	689 (193)	566 (151)	551 (96)	778 (554)	631 (141)
L5	790 (420)	597 (142)	975 (483)	737 (190)	585 (146)	565 (109)	705 (341)	656 (146)
	Accuracy rates (%)							
	Relatively slow readers				Relatively fast readers			
	Words		Pseudowords		Words		Pseudowords	
	English	Dutch	English	Dutch	English	Dutch	English	Dutch
L3	88 (33)	99 (10)	85 (36)	89 (32)	97 (16)	100 (0)	90 (30)	97 (18)
L4	91 (29)	100 (7)	81 (40)	92 (28)	96 (20)	100 (6)	91 (28)	95 (21)
L5	96 (19)	99 (10)	91 (29)	94 (24)	95 (21)	99 (9)	95 (21)	95 (22)

Note. Standard deviations are in parentheses.

additional benefit of this transformation is that it allowed us to test absolute (as opposed to proportional) differences between the reading and language groups. For instance, as can be seen in Table 2, the English children were slower than the Dutch children. As a result, potential significant interactions between the between-participants variables (reading group and language) and the within-participant variables (length and lexicality) could be an effect of differences in overall speed. By analyzing reading speed (items per second) instead of how long a certain item takes to read, we can be reasonably sure that any interaction is an absolute effect and not a proportional effect.

We did not analyze the error data because our research questions were concerned with the transition from serial to parallel reading, which shows in latency outcomes and not in accuracy outcomes. In addition, the Dutch children were at ceiling, as were the relatively fast English children.

Specification and justification of the models

The design comprised two within-participant variables (length and lexicality) and two between-participants variables (reading level and language). A full multilevel model with all conditions, therefore, would consist of 3 (Length) ×2 (Lexicality) ×2 (Reading Level) ×2 (Language) = 24 dummy variables. Because it is hard to interpret four-way interactions, we constructed smaller models, each with three variables (12 dummy variables). Each contained the central two variables for this experiment, namely length and language. In the first model, we examined the interaction among length, language, and lexicality across all reading levels. In the second model, we focused on interactions among length, language, and reading level. To examine the length effect, we specified two contrasts concerning the effect of length three- and four-letter words and length four- and five-letter words. These contrasts were tested simultaneously with a chi-square test with 2 degrees of freedom.

Length × Lexicality × Language model

The estimated mean reading speed and the estimated standard errors of Model 1.1 are presented in Table 3. The three-way interaction among length, lexicality, and language was significant, $\chi^2(2) = 8.29$, $p < .05$. To interpret this, we first examined the Length × Language interaction separately for words and pseudowords and specified follow-up contrasts, and we conducted simple analyses to further pinpoint the meaning of these effects. Next, we analyzed the Length × Lexicality two-way interaction separately for the English and Dutch children and specified subsequent follow-up contrasts and simple analyses.

Table 3
Model 1.1: Estimated mean (and standard error) reading speed in items per second for each of the 12 Length × Lexicality × Language conditions.

	Words		Pseudowords	
	English	Dutch	English	Dutch
L3	1.70 (0.044)	1.84 (0.029)	1.52 (0.054)	1.64 (0.033)
L4	1.72 (0.057)	1.84 (0.030)	1.40 (0.058)	1.61 (0.032)
L5	1.67 (0.054)	1.80 (0.032)	1.42 (0.059)	1.52 (0.033)

The Length × Language interaction effect was not significant for words, $\chi^2(2) < 1$, *ns*, demonstrating that the length effect for words did not differ between the English and Dutch children. For words, simple analysis showed that the main effect of length was significant, $\chi^2(2) = 8.49$, $p < .05$, showing that children were slower to read longer words. The Length × Language interaction was significant for pseudowords, $\chi^2(2) = 17.60$, $p < .001$. Follow-up contrasts indicated that both the English and Dutch children showed significant length effects for pseudowords (English children: $\chi^2(2) = 23.10$, $p < .001$; Dutch children: $\chi^2(2) = 33.00$, $p < .001$), but the interaction indicated that the length effect was smaller for English children than for Dutch children.

The Length × Lexicality interaction effect was significant for both the English children, $\chi^2(2) = 13.80$, $p < .01$, and the Dutch children, $\chi^2(2) = 6.53$, $p < .05$. This indicates that for both groups, the length effect was different for words versus pseudowords. Simple analyses showed that the main effect of length was smaller for words, $\chi^2(2) = 8.49$, $p < .05$ (see above), than for pseudowords, $\chi^2(2) = 37.80$, $p < .001$.

As mentioned in the Method section, we were unable to match the word items across languages for orthographic neighborhood size. Therefore, we calculated two additional one-level coefficients controlling for orthographic neighborhood size for the English and Dutch words. Before the coefficients were inserted into the model, we converted the raw orthographic neighborhood variables into their natural logarithmic values. Because it is impossible to perform logarithmic transformations on zero values, we first added +1 to each neighborhood size variable (see also Marinus & de Jong, 2010b). Inspection of the two neighborhood size coefficients showed that the word reading of the Dutch children was significantly influenced by neighborhood size, $z = -4.9$, $p < .001$; they were faster at reading words from a large neighborhood than from a sparse one. This was not the case for the English children, $z = -0.58$, $p > .10$. After controlling for orthographic neighborhood size, the length effect for words for Dutch readers was significant, $\chi^2(2) = 15.00$, $p < .001$. However, the direction of the length effect was reversed; when orthographic neighborhood size was controlled, the Dutch children were faster to name the longer words than the shorter words.

In sum, all children showed clear length effects when reading pseudowords, and this effect was stronger for the Dutch children. For both language groups, length effects for words were smaller than those for pseudowords. The Dutch children were faster to read words with a large neighborhood size than to read words with a small neighborhood size, and when orthographic N was controlled the children showed reversed length effects. Cross-linguistic differences in sensitivity to orthographic neighborhood size are examined in more detail in Experiment 2.

The results for Model 1.1 show that the cross-linguistic differences in the time course of the expression of length effects differ for words and pseudowords. This empirical finding is in line with our theoretical rationale to specify separate models for words and pseudowords. In the next section, we consider the length and language effects for relatively slow and fast readers.

Length × Language × Reading Level models

We consider first the results for words (Model 1.2a) and then the results for pseudowords (Model 1.2b). The estimated reading speed and standard errors of the models are summarized in Tables 4 and 5.

Table 4
Model 1.2a: Estimated mean (and standard error) reading speed in items per second for each of the 12 Length × Language × Reading Level conditions for words.

	Relatively slow readers		Relatively fast readers	
	English	Dutch	English	Dutch
L3	1.58 (0.071)	1.80 (0.037)	1.79 (0.050)	1.87 (0.030)
L4	1.51 (0.080)	1.81 (0.052)	1.88 (0.063)	1.87 (0.033)
L5	1.49 (0.084)	1.75 (0.055)	1.82 (0.056)	1.84 (0.036)

Table 5
Model 1.2b: Estimated mean (and standard error) reading speed in items per second for each of the 12 Length × Language × Reading Level conditions for pseudowords.

	Relatively slow readers		Relatively fast readers	
	English	Dutch	English	Dutch
L3	1.34 (0.058)	1.56 (0.056)	1.65 (0.073)	1.71 (0.035)
L4	1.22 (0.069)	1.54 (0.048)	1.54 (0.076)	1.66 (0.040)
L5	1.18 (0.067)	1.43 (0.046)	1.59 (0.073)	1.60 (0.041)

Words. The Length × Language × Reading Level three-way interaction was significant, $\chi^2(2) = 7.60$, $p < .05$. To interpret this effect, we analyzed the Length × Reading Group interaction separately for the English and Dutch readers. This analysis directly addresses the following question: In what way is the development of the length effect for words different for English and Dutch readers? For the English children, we found a significant Length × Reading Level interaction, $\chi^2(2) = 10.30$, $p < .01$, caused by a larger effect of length for the slower readers. In contrast, in Dutch there was no difference in length effect between the relatively slow and relatively fast Dutch readers, $\chi^2(2) < 1$, *ns*.

We also found a significant two-way interaction between language and reading group. The relatively slow English readers were indeed slower than the relatively fast English readers, $\chi^2(1) = 10.60$, $p < .01$. However, the two groups of Dutch children did not differ in word reading speed, $\chi^2(2) = 1.86$, $p > .10$.

Pseudowords. The Length × Language × Reading Level three-way interaction was not significant, $\chi^2(2) < 1$, *ns*. Both the two-way interaction of Length × Language, $\chi^2(2) = 16.42$, $p < .001$, and Length × Reading Level, $\chi^2(2) = 7.02$, $p < .05$, were significant. In line with the results for the first model (Length × Lexicality × Language; see Model 1.1 in Table 3), follow-up analyses showed that the length effect for pseudowords was significant for both English and Dutch readers but was larger for the Dutch readers. Follow-up analyses on the second two-way interaction effect also confirmed findings from previous studies (Marinus & de Jong, 2010b; Spinelli et al., 2005; Zoccolotti et al., 2005) that the length effect for pseudowords is present for all children but larger for relatively slow readers, $\chi^2(2) = 29.71$, $p < .001$, than for relatively fast readers, $\chi^2(2) = 24.46$, $p < .001$.

We also found a significant two-way interaction between language and reading group. For readers of both orthographies, there was a difference in overall reading speed between lower and faster readers (English: $\chi^2(1) = 14.01$, $p < .001$; Dutch: $\chi^2(1) = 5.75$, $p < .05$), but this difference was larger for English readers than for Dutch readers.

Discussion

It is important to compare the different outcomes of the two models in order to interpret the length effects for words. In an initial analysis that did not include a comparison between fast and slow readers, length effects were similar across languages (Model 1.1). This makes sense, especially because the relatively slow readers in both languages were still at the relatively early stages of reading development and arguably still building orthographic knowledge for some of our stimuli. Hence, it is not surprising that as a group both the English and Dutch children were still showing length effects, reflecting serial processing. However, when reading level was taken into account (Model 1.2a), a different pattern emerged, with the length effect being larger for the slower reading group relative to the faster group. Still, this was the case only for the English children; there was no difference in length effect for the fast versus slow readers of Dutch.

Together, these results indicate that English children follow the expected pattern of gradually using more parallel strategies as reading level increases. In contrast, the Dutch children, including the subgroup with relatively slow reading, showed only a small length effect for words. This supports the idea that Dutch children become more proficient in word reading earlier than English children, thereby diminishing the difference between good and poor readers in word length effects. This explanation is also supported by our finding that the differences in overall reading speed for the relatively slow versus relatively fast readers is significantly larger in English than in Dutch.

The outcomes for pseudowords were straightforward; both English and Dutch readers showed length effects, with these being stronger for the Dutch children (Models 1.1 and 1.2b). This concurs with the prediction from the orthographic depth hypothesis (Frost et al., 1987), namely that a transparent language encourages more use of serial sublexical processing strategies than a more opaque language. In line with previous studies (Marinus & de Jong, 2010b; Spinelli et al., 2005; Zoccolotti et al., 2005), relatively slow readers showed a larger length effect for pseudowords than relatively fast readers (Model 1.2b), a finding that held across both languages.

An additional finding was that Dutch children were affected by neighborhood size, whereas English children were not. Consequently, for the English children, controlling for N-size did not change the length effect. For the Dutch children, however, adding neighborhood size as a control variable served to reverse the length effect for words, with longer words being named more quickly than shorter words. This reversed length effects for the Dutch children makes sense both theoretically and statistically.

Theoretically, it is known that length and N-size independently affect word reading speed (Balota, Cortese, Sergent-Marshall, Spieler, & Yap, 2004). However, their impact steers the reaction times in the same direction; the shorter the word, the faster it will be read, and the higher the N-size, the faster it will be read. Because N-size is much higher for the three-letter words than for the five-letter words, RTs for three-letter words will be "compensated" more strongly than those for five-letter words once N-size is controlled. Thus, the reversal of length effects in Dutch, especially in smaller (three- to five-letter) words, may be a reflection of length effects in Dutch being determined by orthographic N-size rather than word length.

In addition to this theoretical explanation, the item characteristics in Experiment 1 might also have played a role in finding a reversed length effect. In other words, the reversed length effects could be partly statistically driven. In the current study, the differences in orthographic neighborhood size were more extreme across length conditions for the Dutch stimuli than for the English stimuli (see Table 6).

Table 6
Neighborhood size statistics per language, lexicality, and length condition in Experiment 1.

	Dutch					
	Words			Pseudowords		
	3L	4L	5L	3L	4L	5L
Mean N-size	17.58	5.33	3.75	9	8.5	8.25
SD	3.34	2.81	1.82	2.04	2.39	1.6
Range	14–23	2–12	0–7	6–12	6–13	6–11
	English					
	Words			Pseudowords		
	3L	4L	5L	3L	4L	5L
Mean N-size	8.83	2.83	2.83	6	5.83	5.58
SD	3.61	1.7	1.19	2.49	1.27	0.67
Range	5–14	1–7	1–3	1–9	3–7	5–7

It is possible that this asymmetrical distribution affected the estimations of the statistical models. Of interest here are the outcomes of an earlier study focusing on length effects in Dutch children in which the N-size differences between the length conditions were less extreme (Marinus & de Jong, 2010b). This study did not report reversal of length effects after controlling for N-size. However, the Dutch words used in the current study were also less than half the frequency as the stimuli of Marinus and de Jong (2010b). Previous studies showed that the effect of neighborhood size is stronger when words are less frequent or familiar (Marinus & de Jong, 2010a; Weekes, 1997). As per our theoretical explanation, this could have given the neighborhood effect the needed additional strength to override the effects of length, resulting in reversed length effects for the Dutch children (see also van den Boer, de Jong, & Haentjens-van Meeteren, 2012).

Experiment 2

The aim of Experiment 2 was to examine cross-linguistic differences in sensitivity to orthographic overlap between words. Although this has not been explored before, it is possible to make two contrasting predictions based on the orthographic depth theory and differences in density between the two languages. If it is true that English readers use more lexical processing (including reading by analogy) than Dutch readers, as postulated by the orthographic depth theory, they should benefit more from words having more neighbors than the Dutch children. In contrast, however, if the underlying factor is orthographic density rather than lexical reading strategies steered by orthographic depth, Dutch children should benefit more than English children from words and pseudowords having more neighbors. Stronger facilitation by overlap for the Dutch children can be explained by assuming that their reading system is more strongly tuned toward processing words that are highly similar as a result of their reading experience.

Based on the results of within-language studies in both English (Andrews, 1997) and Dutch (Marinus & de Jong, 2010a), we expected the facilitating effect of orthographic overlap to be stronger for pseudowords than for words for both groups of children. Until now, studies investigating general N-size effects in children learning to read in English have focused on accuracy and not on reading speed. The only study we know of that compared the reading speed of beginning (Grade 2) and more advanced (Grade 4) readers in Dutch found no differences in orthographic N-size effects (Marinus & de Jong, 2010a). We expected to replicate this effect in the current experiment for the Dutch Grade 3 and 4 readers.

Method

Participants

The same children who participated in Experiment 1 also completed Experiment 2.

Materials and procedure

For both English and Dutch, we selected a set of words and pseudowords, with 30 four-letter items and 30 five-letter items in each set. Half of the items at each length had many orthographic neighbors (6–20); the other half had few orthographic neighbors (0–3). To calculate number of orthographic neighbors, we used the English and Dutch corpus of the CELEX database. The same procedures were used as in Experiment 1 to match and control stimulus properties across the two languages. Stimuli were matched across languages for frequency (words only) and mostly for first letter/phoneme across language (Dutch and English). In addition, the words and pseudowords were matched on first letter and word structure (e.g., CCVC). Pseudowords were created by changing the final letter of the words and occasionally by changing the vowel or second consonant of the consonant onset cluster. As can be seen in Table 7, the manipulation of neighborhood size was equally strong for words and pseudowords in both languages. See Appendix B for the stimuli sets.

The general procedure was the same as in Experiment 1. Words and pseudowords were presented as two separate blocks, and within each block items were presented in a random order. Within each language group, half of the children read the words first and half read the pseudowords first.

Table 7
Neighborhood size statistics for Experiment 2.

	Dutch				English			
	Words		Pseudowords		Words		Pseudowords	
	High N-size	Low N-size	High N-size	Low N-size	High N-size	Low N-size	High N-size	Low N-size
Mean N-size	9.63	1.80	9.53	2.00	9.20	2.09	8.93	1.83
SD	3.09	1.00	2.36	0.91	2.61	0.78	2.26	0.83
Range	6–20	0–3	6–15	0–3	6–17	1–3	6–13	1–3

Results

Data cleaning and model structure

Our analytic approach and data cleaning procedures were the same as for Experiment 1. The percentage of invalid latencies for the English children was 11.6% for words and 11.7% for pseudowords. For the Dutch children it was 10.7% for words and 14.2% for pseudowords. Error rates are shown in Table 8 along with latency data. The Dutch children were at ceiling for all conditions (accuracy > 92%). The English children were less accurate when reading pseudowords than when reading words, and they made fewer errors when items came from a dense neighborhood (high N-size) compared with a sparse neighborhood (low N-size).

The design comprised two within-participant variables, N-*size* (high or low) and *lexicality* (words or pseudowords) and two between-participants variables, *reading level* (relatively slow or relatively fast) and *language* (English or Dutch). As with Experiment 1, we simplified the analyses by estimating two separate models containing the two primary variables, N-size and language. In Model 2.1, we examined the interaction among N-size, language and lexicality. Again we ran separate analyses for the words (Model 2.2a) and pseudowords (Model 2.2b) when examining interactions with reading proficiency and language.

N-Size × Lexicality × Language model

The estimated mean reading speed and standard errors of Model 2.1 are presented in Table 9. The three-way interaction among N-size, lexicality, and language was not significant, $\chi^2(1) = 1.97$, $p > .10$. However, all two-way interactions were statistically significant, as detailed below.

Table 8
Mean latency scores and accuracy rates of the relatively slow and relatively fast English and Dutch readers for words and pseudowords in both (low and high) orthographic neighborhood size conditions.

	Latencies (ms)							
	Relatively slow readers				Relatively fast readers			
	Words		Pseudowords		Words		Pseudowords	
	English	Dutch	English	Dutch	English	Dutch	English	Dutch
High N	843 (413)	607 (121)	1067 (505)	673 (184)	662 (230)	589 (133)	814 (490)	638 (175)
Low N	792 (275)	634 (143)	1177 (573)	721 (203)	656 (231)	606 (148)	868 (570)	669 (206)
	Accuracy rates (%)							
	Relatively slow readers				Relatively fast readers			
	Words		Pseudowords		Words		Pseudowords	
	English	Dutch	English	Dutch	English	Dutch	English	Dutch
High N	88 (33)	98 (13)	79 (41)	97 (18)	96 (19)	99 (9)	90 (30)	97 (17)
Low N	83 (38)	94 (24)	72 (45)	94 (23)	95 (22)	94 (23)	83 (38)	93 (25)

Note. Standard deviations are in parentheses.

Table 9
Model 2.1: Estimated mean (and standard error) reading speed in items per second for each of the eight N-Size × Lexicality × Language conditions.

	Words		Pseudowords	
	English	Dutch	English	Dutch
High N	1.52 (0.051)	1.74 (0.031)	1.33 (0.060)	1.61 (0.035)
Low N	1.54 (0.049)	1.69 (0.032)	1.25 (0.062)	1.53 (0.033)

The strength of the N-size effect differed for words and pseudowords, $\chi^2(1) = 13.00$, $p < .001$, being stronger for pseudowords, $\chi^2(1) = 40.70$, $p < .001$, than for words, $\chi^2(1) = 3.86$, $p < .05$. The N-size effect also differed for the two language groups, $\chi^2(1) = 13.37$, $p < .001$, being significant for the Dutch children, $\chi^2(1) = 61.60$, $p < .001$, but absent for the English children, $\chi^2(1) = 2.53$, $p > .10$, replicating Experiment 1. Finally, we also found a significant Lexicality × Language effect, $\chi^2(1) = 12.42$, $p < .001$. The difference in word versus pseudoword reading speed was larger for the English children than for the Dutch children, but it was significant for both language groups, $\chi^2(1) = 80.50$, $p < .001$, and $\chi^2(1) = 80.09$, $p < .001$, respectively.

N-Size × Language × Reading Level model

The estimated mean reading speed and standard errors of the models can be found in Table 10 (words, Model 2.2a) and Table 11 (pseudowords, Model 2.2b).

Words. The three-way interaction effect among N-size, language, and reading level was not significant, $\chi^2(1) < 1$, *ns*. The two-way interaction between N-size and language was significant, $\chi^2(1) = 13.44$, $p < .001$. Simple follow-up analyses showed that there was no effect of N-size for the English children, $\chi^2(1) = 1.25$, $p > .01$, but a significant effect for the Dutch children, $\chi^2(1) = 24.72$, $p < .001$. The interaction between language and reading level was marginal, $\chi^2(1) = 3.72$, $p = .0539$. Simple follow-up analyses showed that for the English readers there was a significant difference between the relatively slow and relatively fast readers, $\chi^2(1) = 15.09$, $p < .001$, whereas the two subgroups of Dutch readers did not differ in their word reading speed, $\chi^2(1) = 1.91$, $p > .10$, mirroring findings from Experiment 1. The N-Size × Reading Level interaction was not significant.

Pseudowords. The three-way interaction effect among N-size, language, and reading level was not significant, $\chi^2(1) < 1$, *ns*. In contrast to the results for word stimuli, there was no interaction between N-size and language; for pseudowords, the N-size effects were the same for both English and Dutch readers. The main effects for N-size and language were significant; children were faster to read pseudowords with a large N-size than to read pseudowords with a small N-size, $\chi^2(1) = 35.69$, $p < .001$, and Dutch children were faster than English readers, $\chi^2(1) = 29.96$, $p < .001$. We also found a significant two-way interaction between language and reading level, $\chi^2(1) = 5.21$, $p < .05$. The difference between the relatively fast and relatively slow readers was larger for the English children, $\chi^2(1) = 36.15$, $p < .001$, than for the Dutch children. However, in contrast to the findings for words, but replicating our findings from Experiment 1 for pseudowords, this effect was also apparent in Dutch, $\chi^2(1) = 4.36$, $p < .05$.

Table 10
Model 2.2a: Estimated mean (and standard error) reading speed in items per second for each of the eight N-Size × Language × Reading Level conditions for words.

	Relatively slow readers		Relatively fast readers	
	English	Dutch	English	Dutch
High N	1.35 (0.079)	1.70 (0.038)	1.64 (0.058)	1.78 (0.050)
Low N	1.37 (0.067)	1.65 (0.041)	1.66 (0.058)	1.73 (0.048)

Table 11
Model 2.2b: Estimated mean (and standard error) reading speed in items per second for each of the eight N-Size × Language × Reading Level conditions for pseudowords.

	Relatively slow readers		Relatively fast readers	
	English	Dutch	English	Dutch
High N	1.08 (0.064)	1.57 (0.042)	1.46 (0.077)	1.66 (0.055)
Low N	1.02 (0.074)	1.47 (0.039)	1.41 (0.078)	1.60 (0.051)

Discussion

In Experiment 2, we aimed to directly investigate differences in sensitivity to overlap between words in children learning to read either English or Dutch. Replicating the observation from Experiment 1, the Dutch children were sensitive to overlap between words, whereas the English children were not. Both groups of children were sensitive to orthographic neighborhood size when reading pseudowords, but the effect was stronger for the Dutch children.

Once again, we found that the slow and fast Dutch children performed more similarly to each other than the two English subgroups. The English relatively slow children were slower than their relatively fast peers for both words and pseudowords. In the Dutch children this effect was much smaller, and for word reading speed the performances of the reading level groups did not differ. To make sure that the English and Dutch samples were equally proficient in reading, we individually matched all participating children on single-word reading speed. So why do these cross-linguistic differences in general reading speed between the relatively slow and relatively fast reading children arise? We think that the explanation lies in differences in the complexity of items (i.e., in terms of morphology and number of syllables) used for the matching procedure and the one-syllable items used in the experiments. From this perspective, the cross-linguistic differences in general reading speed tie in nicely with the view that reading development occurs faster in children learning to read in more transparent languages. The Dutch children in the relatively slow reading group may still perform at the same level as the relatively slow English children on a standardized reading test that contains more *complex words.* At the same time, however, their word recognition speed for *simpler words* (i.e., monosyllabic words, mostly nouns that were used for the experiments) has already reached the level of relatively fast Dutch readers. Similarly, the relatively slow Dutch readers have also started to close the gap with their faster reading language peers with regard to reading speed for simple pseudowords. The relatively slow English readers, on the other hand, need more time and reading experience to progress and to close these gaps. Using a cross-sectional study, Frith, Wimmer, and Landerl (1998) already demonstrated that differences in pseudoword reading speed between German and English children no longer exist by 12 years of age. Ideally these findings should be confirmed using a longitudinal design.

Let us return to our main question: How could the cross-linguistic difference in sensitivity to overlap between words and pseudowords be explained? As raised in the introduction to Experiment 2, a possibility is that the reading system of a child learning to read in Dutch is better trained to benefit from overlap than that of a child learning to read in English because Dutch has a denser orthography than English. We further elaborate on the mechanisms in the General Discussion.

In addition, neighbors of Dutch words are consistent in that they facilitate recognition of not only the target word's orthographic form but also its pronunciation; that is, neighbors are almost always friends. This is not the case in English. In line with this idea, Frauenfelder, Baayen, and Hellwig (1993) demonstrated that there is more overlap in Dutch, as compared with English, in the regression curves for the neighborhood density of orthographic and phonemic representations. When selecting the English items for Experiment 2, we did not control for the presence of enemies and friends. It might be that neighborhood effects were not so strong in English due to interference from enemies (e.g., *pint, mint*) conspiring against facilitation from friends. We return to this possibility in more detail in the General Discussion.

No other cross-linguistic study has compared N-size effects. However, relevant data were collected by Ziegler and colleagues (2003), who compared English and German children's reading using a different metric, namely body neighborhood size. In contrast to our findings, Ziegler and colleagues found that English children showed larger neighborhood size effects than German children. It should be noted, however, that their body neighborhood size manipulation was stronger for the English items (low body N: on average 2.9; high body N: on average 10.9; difference of 8.0) than for the German items (low body N: on average 2.5; high body N: on average 8.5; difference of 6.0). This is not surprising because German monosyllabic words have on average fewer neighbors and body neighbors than English monosyllabic words (CELEX database; Baayen et al., 1993; see also Footnote 1). Nevertheless, it could be that this difference in manipulation strength provides an alternative explanation for the cross-linguistic differences observed by Ziegler and colleagues. We return to the theoretical implications of these alternative explanations later.

General discussion

An important question is why reading development takes longer for children learning to read in English than for children learning to read in more transparent languages such as Dutch. To address this question, we compared the influence of orthographic transparency on cross-linguistic differences in the development of (a) the transition from serial decoding to more parallel reading strategies and (b) sensitivity to overlap between words in beginning readers.

The results of Experiment 1 showed that the transition from serial processing to more parallel word reading occurs more slowly in English children than in Dutch children. This was shown by the fact that the relatively slow English readers displayed larger length effects than the relatively fast English readers, whereas the two Dutch subgroups did not differ in the strength of their length effects. These findings are consistent with the predictions from the self-teaching hypothesis, namely that the transparent and easier to decode Dutch orthography promotes opportunities for orthographic learning and, hence, faster transition to more parallel reading strategies. Therefore, our first main conclusion is that the transparency of letter-to-sound mappings directly affects orthographic learning, forming one explanation for why it takes longer to learn to read in English than in a transparent language such as Dutch.

For pseudowords, we found stronger length effects in Dutch than in English. This concurs with the prediction that readers of transparent languages will be more proficient at using serial or sublexical processes when reading, stemming from the orthographic depth hypothesis (Frost et al., 1987). These outcomes are also in line with the rationale of the psycholinguistic grain size theory (Ziegler & Goswami, 2005) that Dutch readers will continue to rely more strongly on smaller sublexical units (i.e., grapheme–phoneme units), whereas English readers are postulated to rely more strongly on body–rime units. Finally, we also replicated the finding of earlier studies that slower readers show larger effects of pseudoword length than faster readers (Marinus & de Jong, 2010b; Zoccolotti et al., 2005).

In addition to cross-linguistic differences in serial processing revealed by differential length effects, Experiment 1 also provided initial evidence of language differences associated with the effect of orthographic neighborhood size. Dutch children were sensitive to orthographic neighborhood size in this length experiment; they were faster to read words from a larger neighborhood than those from a sparser neighborhood. English children did not show such an effect. A likely explanation for this cross-linguistic difference is that the English words in Experiment 1 were generally lower in orthographic neighborhood size than the Dutch words, in line with the fact that, overall, Dutch monosyllabic words tend to have higher neighborhood sizes than English words (Baayen et al., 1993; Norris & Kinoshita, 2012).

In Experiment 2, we matched English and Dutch words and pseudowords on absolute N-size so that we could directly address whether Dutch and English readers differ in the degree to which they benefit from orthographic overlap between words. Because we found that the reading speed of Dutch children benefited more strongly from higher N-sizes than that of the English children, our conclusion is that cross-linguistic differences in orthographic density influence word and pseudoword recognition speed. But how might differences in density affect reading aloud?

One developmental theory that explains how the reading system adapts in reaction to exposure to highly similar words (i.e., neighbor words) is the lexical tuning hypothesis (Castles, Davis, Cavalot, & Forster, 2007). According to this theory, if a child learns a new orthographic representation (e.g., *cot*) that is highly similar to an already acquired representation (e.g., *cat*), the existing orthographic representation will become more "finely tuned." This is because if two or more words look alike, their representations must be specific in order to be mapped onto the correct pronunciation and meaning. In a dense orthography such as Dutch, such tuning is more critical than in a less dense orthography such as English. Thus, experience with reading Dutch might induce lexical tuning more strongly and from an earlier point in reading development (see also Marinus & de Jong, 2010a) than experience with reading English, leading to the cross-linguistic differences in word and pseudoword naming speed seen in our experiments.

The lexical tuning hypothesis discusses lexical tuning from an item-based point of view; a specific word becomes more strongly tuned in response to the creation of a highly similar representation. Item-based tuning, however, is not enough to explain our cross-linguistic findings because we matched across language on absolute N-size. To explain our findings, we need to go a step further and suggest that when an orthography is denser, the reading system will respond with even more tuning. In other words, the system might be "better trained to tune."

But then, how exactly would more strongly tuned representations lead to faster naming speed of words and pseudowords? One speculation might be that due to the "hyper" tuning of high N-size words in Dutch, the difference in resting levels between the more strongly tuned Dutch representations (i.e., the high N-size words) and the less strongly tuned ones (i.e., low N-size words) is larger than that for the English representations. As a result, the differences in naming speed for high and low N-size words are smaller for English children than for Dutch children.

Another important question is whether the sensitivity to the similarity between words is driven by phonological overlap or by orthographic overlap. These factors are difficult to pull apart, but there is reason to argue that orthographic overlap might be driving the effect. First, consider previous work comparing children learning English versus German (Ziegler et al., 2003). English is orthographically denser than German, and English children showed greater orthographic neighborhood effects as estimated with the body N metric. Had this been due to sensitivity to phonological overlap, the relatively transparent German language, with its larger proportion of phonologically overlapping neighbors, should have steered the effect into the other direction—resulting in stronger sensitivity to overlap for the German children. However, this interpretation must be treated with caution. As discussed earlier, the manipulation strength of the body N metric was not matched across the two languages in Ziegler and colleagues' (2003) experiment; it was stronger for English than for German.

Additional evidence that orthographic overlap drives the emergence of neighborhood effects comes from Experiment 2. Here, the majority of the English items were consistent; that is, in most cases, all the body neighbors of the English words and pseudowords had the same pronunciation as the target word (words: 43/60; pseudowords: 40/60), just like in Dutch. Moreover, as demonstrated by Jared, McRae, and Seidenberg (1990), it is not the case that the presence of enemies itself slows down word recognition; this happens only when the enemies are highly frequent. Taking a look at our word sets in this more restricted way, only three words and seven pseudowords had enemies of a higher frequency than their friends. This strongly suggests that differences in phonological overlap did not contribute to our results. Instead, we suggest that cross-linguistic differences in orthographic density, and not phonological density, caused the different orthographic neighborhood size effects that we saw between the English and Dutch children. This leads to our third conclusion, namely that it is easier to learn to read in Dutch, not only because its orthography is more transparent than that of English but also because of its higher orthographic density.

How do our cross-linguistic differences in sensitivity to orthographic neighborhood size fit within the orthographic depth hypothesis? Following Frost and colleagues (1987), this hypothesis states that children learning to read English will use more lexical strategies, such as reading words by analogy with other words, than children learning a more transparent language such as German or Dutch. This is not what we found. The Dutch children clearly benefited more from orthographic overlap than the English children, which indicates stronger lexical processing in the Dutch children than in the English children.

As explained in the Introduction, neighborhood size effects cannot be discussed directly within the psycholinguistic grain size theory (Ziegler and Goswami, 2005) because this theory focuses on cross-linguistic differences in the use of larger units (e.g., bodies) versus smaller units (i.e., grapheme–phoneme units). However, there is considerable overlap between the body N and general N metrics, leading to the question of whether there might be an alternative account of Ziegler and colleagues' (2003) body N-size data. Because German is less orthographically dense (and therefore will also have fewer body neighbors) than English at the monosyllabic level, this might cause readers of English to show larger body-N effects than readers of German. Plausibly, this effect was magnified in Ziegler and colleagues' experiment because the body neighborhood size manipulation was stronger for the English items than for the German items. To reliably address this issue, further research is needed. A key experiment would be to compare general N-size effects among English, German, and Dutch readers while controlling for body N.

In conclusion, conducting cross-linguistic research is a challenging undertaking that needs to take many potential within-language and across-language factors into account. In this research, we showed that alongside the well-known dimension of orthographic depth, another factor—namely orthographic density—contributes to cross-linguistic differences in learning to read.

Acknowledgments

We thank Steven Saunders for retrieving the density statistics from the CELEX database. The English data were collected when the first author visited Oxford University in 2007, partly funded by the Experimental Psychology Society. The Dutch data were collected while the first author was appointed as a post-doctoral researcher in the Department of Educational Sciences at the University of Amsterdam in 2009. The data were analyzed, and the article was written, while she was funded by an NWO Rubicon grant (Project No. 446-09-014) at the Macquarie Centre for Cognitive Science and during her appointment as a post-doctoral researcher within the ARC Research Centre of Excellence for Cognition and Its Disorders at Macquarie University. Kate Nation was supported by the Economic and Social Research Council.

Appendix A.

A.2. English and Dutch word and pseudoword sets of the length task.

Words						Pseudowords					
Three letters		Four letters		Five letters		Three letters		Four letters		Five letters	
English	Dutch	English	Dutch	English	Dutch	English	Dutch	English	Dutch	English	Dutch
bus	bus	blot	brug	blink	blank	bef	bif	blam	blak	blant	blets
cab	kam	drag	druk	drink	drank	bym	bun	brin	brof	brond	brank
cup	kip	drum	drum	drops	dwerg	kel	cag	clum	kles	climp	klans
dam	dal	frog	fles	frown	flits	kas	kur	crof	krum	crant	krang
dug	dek	scar	snor	scarf	slurf	gof	gig	gral	gral	grast	grans
fog	fel	scat	smak	skirt	sterk	gek	gur	gris	gres	grint	grent
pup	pup	spit	step	spark	speld	pof	pug	prip	prip	prist	prank
sag	set	spun	slag	spent	spits	pif	def	prak	pras	pronk	prant
sir	sip	stem	spul	sport	sport	sif	sug	slig	slom	sland	storp
sob	sok	step	spek	start	start	sef	sif	swog	sleg	swink	slomp
tag	tal	stew	sluw	storm	slang	tes	teb	trin	trin	traft	trons
ton	ton	swim	smal	swamp	stomp	tuv	tur	treg	treg	trank	trank

Appendix B.

B.1. English and Dutch word and pseudoword sets of the neighborhood task.

High N-size four letters				Low N-size four letters			
English		Dutch		English		Dutch	
Words	Pseudowords	Words	Pseudowords	Words	Pseudowords	Words	Pseudowords
bent	belk	berg	belk	bulb	bubs	club	bulm
bunk	bunt	bast	burk	club	clus	darm	klun
clap	clag	krak	klag	desk	dems	dorp	derp
dust	dunt	dolk	dust	drip	dreg	durf	delm
flat	flam	flop	flas	drop	dron	film	dolg
lamp	lans	last	lank	film	fisc	fles	fist
prop	prot	plak	prok	glen	glak	geld	glun
pump	pusk	park	pust	gulp	guft	gids	gurp
rink	rilt	rest	rint	plug	plis	golf	pleb
scat	scag	stem	stas	plus	plit	pech	plin
slip	slan	slap	slit	smog	smup	smid	smof
slot	slom	slok	slom	twig	twol	terp	twos
spat	sput	stek	sput	twin	twap	turf	twep
swag	swit	stal	stin	wolf	wons	wilg	wops
west	welf	west	werm	yelp	yemp	wurm	jept
High N-size five letters				Low N-size five letters			
beast	baunt	beest	beels	brief	braim	breuk	breig
bound	binch	boord	boorn	broad	broaf	beurs	broof
brain	brate	bries	braan	clear	clerp	kleed	kleup
creek	creat	kliek	kraat	cloud	cloaf	kling	kleip
grate	grash	graat	groes	clown	clows	clown	kluug
grave	grake	griep	griel	drool	drote	dreun	drien
paste	palse	paars	paats	frame	frage	frame	fraas
prime	prine	praal	pries	sleek	slobe	stuur	sloen
rains	raint	kluis	reest	snail	snase	sjaal	snuuf
scout	scark	speen	slook	sneer	sneem	snauw	sniep
snout	snool	snoet	snoel	speed	speef	spoed	spief
spine	spale	stier	spoer	spoil	sparb	spuug	spaam
steep	steen	steek	steer	stoat	sterg	steil	steum
stoop	stoot	sloot	stoek	troop	troin	toost	trool
train	treak	troef	truis	waist	waims	weids	woers

References

Andrews, S. (1997). The effect of orthographic similarity on lexical retrieval: Resolving neighborhood conflicts. *Psychonomic Bulletin & Review, 4*, 439–461.

Baayen, R. H., & Milin, P. (2010). Analyzing reaction times. *International Journal of Psychological Research, 3*, 12–28.

Baayen, R. H., Piepenbrock, R., & van Rijn, H. (1993). *CELEX lexical database [computer software]*. Philadelphia: Linguistic Data Consortium, University of Pennsylvania.

Balota, D. A., Cortese, M. J., Sergent-Marshall, S. D., Spieler, D., & Yap, M. J. (2004). Visual word recognition of single-syllable words. *Journal of Experimental Psychology: General, 133*, 283–316.

Bleichrodt, N. D., Drenth, P. J. D., Zaal, J. N., & Resing, W. C. M. (1987). *Revisie Amsterdamse Kinder Intelligentie Test [Revised Amsterdam Child Intelligence Test]*. Lisse, The Netherlands: Swets & Zeitlinger.

Brus, B. T., & Voeten, M. J. M. (1995). *Eén minuut test [One-Minute Test]*. Lisse, The Netherlands: Swets & Zeitlinger.

Castles, A., Davis, C., Cavalot, P., & Forster, K. (2007). Tracking the acquisition of orthographic skills in developing readers: Masked priming effects. *Journal of Experimental Child Psychology, 97*, 165–182.

Christian, K., Morrison, F. J., Frazier, J. A., & Massetti, G. (2000). Specificity in the nature and timing of cognitive growth in kindergarten and first grade. *Journal of Cognition and Development, 1*, 429–448.
Coltheart, M., Davelaar, E., Jonasson, J. T., & Besner, D. (1977). Access to the internal lexicon. In S. Dornic (Ed.), *Attention and performance VI* (pp. 535–555). Hillsdale, NJ: Lawrence Erlbaum.
Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. C. (2001). DRC: A dual route cascaded model of visual word recognition and reading aloud. *Psychological Review, 108*, 204–256.
Cunningham, A. J., & Carroll, J. M. (2011). The development of early literacy in Steiner- and standard-educated children. *British Journal of Educational Psychology, 81*, 475–490.
Dunn, L. M., Dunn, L. M., Whetton, C., & Pintilie, D. (1982). *The British Picture Vocabulary Scale (BPVS)*. Windsor, UK: NFER–Nelson.
Ellis, N. C., & Hooper, M. (2001). Why learning to read is easier in Welsh than in English: Orthographic transparency effects evinced with frequency-matched tests. *Applied Psycholinguistics, 22*, 571–599.
Ellis, N. C., Natsume, M., Stavropoulou, K., Hoxhallari, L., van Daal, V. H. P., Polyzoe, N., et al (2004). The effects of orthographic depth on learning to read alphabetic, syllabic, and logographic scripts. *Reading Research Quarterly, 39*, 438–468.
Frauenfelder, U. H., Baayen, R. H., & Hellwig, F. M. (1993). Neighborhood density and frequency across languages and modalities. *Journal of Memory and Language, 32*, 781–804.
Frith, U., Wimmer, H., & Landerl, K. (1998). Differences in phonological recoding in German- and English-speaking children. *Scientific Studies of Reading, 2*, 31–54.
Frost, R., Katz, L., & Bentin, S. (1987). Strategies for visual word recognition and orthographic depth: A multilingual comparison. *Journal of Experimental Psychology: Human Perception and Performance, 13*, 104–115.
Goswami, U., Ziegler, J. C., Dalton, L., & Schneider, W. (2003). Nonword reading across orthographies: How flexible is the choice of reading units? *Applied Psycholinguistics, 24*, 235–247.
Jared, D., McRae, K., & Seidenberg, M. S. (1990). The basis of consistency effects in word naming. *Journal of Memory and Language, 29*, 687–714.
Landerl, K., Wimmer, H., & Frith, U. (1997). The impact of orthographic consistency on dyslexia: A German-English comparison. *Cognition, 63*, 315–334.
Laxon, V., Coltheart, V., & Keating, C. (1988). Children find friendly words friendly too: Words with many orthographic neighbors are easier to read and spell. *British Journal of Educational Psychology, 58*, 103–119.
Marinus, E. (2010). *Word-recognition processes in normal and dyslexic readers* (Doctoral thesis). University of Amsterdam.
Marinus, E., & de Jong, P. F. (2008). The use of sublexical clusters in normal and dyslexic readers. *Scientific Studies of Reading, 12*, 253–280.
Marinus, E., & de Jong, P. F. (2010a). Orthographic neighborhood effects in normal and dyslexic readers: Size doesn't matter, frequency does. *Journal of Experimental Child Psychology, 106*, 129–144.
Marinus, E., & de Jong, P. F. (2010b). Variability in the word-reading performance of dyslexic readers: Effects of letter length, phoneme length, and digraph presence. *Cortex, 46*, 1259–1271.
Masterson, J., Stuart, M., Dixon, M., & Lovejoy, S. (2003). *Children's printed word database*. Retrieved from: <http://www.essex.ac.uk/psychology/cpwd>.
Norris, D., & Kinoshita, S. (2012). Reading through a noisy channel: Why there's nothing special about the perception of orthography. *Psychological Review, 119*, 517–545.
Patel, T. K., Snowling, M. J., & de Jong, P. F. (2004). A cross-linguistic comparison of children learning to read in English and Dutch. *Journal of Educational Psychology, 96*, 785–797.
Quené, H., & van den Berg, H. (2004). On multi-level modeling of data from repeated measures designs: A tutorial. *Speech Communication, 43*, 103–121.
Rasbash, J., Steele, F., Browne, W., & Prosser, B. (2004). *A user's guide to MLwiN Version 2.0*. London: Institute of Education.
Ratcliff, R. (1993). Methods for dealing with reaction time outliers. *Psychological Bulletin, 114*, 510–532.
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.
Raven, J. C., Court, J. H., & Raven, J. (1986). *Standard progressive matrices*. London: H. K. Lewis.
Schneider, W., Eschman, A., & Zuccolotto, A. (2002). *E-Prime reference guide*. Pittsburgh, PA: Psychology Software Tools.
Schrooten, W., & Vermeer, A. (1994). *Woorden in het basisonderwijs: 15,000 woorden aangeboden aan leerlingen [Words in primary education: 15,000 words presented to students]*. Tilburg, The Netherlands: Tilburg University Press.
Seymour, P. H. K., Aro, M., & Erskine, J. M. (2003). Foundation literacy acquisition in European languages. *British Journal of Psychology, 94*, 143–174.
Share, D. L. (1995). Phonological recoding and self-teaching: Sine qua non of reading acquisition. *Cognition, 55*, 151–218.
Share, D. L. (2004). Orthographic learning at a glance: On the time course and developmental onset of self-teaching. *Journal of Experimental Child Psychology, 87*, 267–298.
Share, D. L. (2008). On the anglocentricities of current reading research and practice: The perils of overreliance on an "outlier" orthography. *Psychological Bulletin, 134*, 584–615.
Spinelli, D., de Luca, M., di Filippo, G., Mancini, M., Martelli, M., & Zoccolotti, P. (2005). Length effect in word naming in reading: Role of reading experience and reading deficit in Italian readers. *Developmental Neuropsychology, 27*, 217–235.
Torgesen, J. K., Wagner, R. K., & Rashotte, C. A. (1999). *Test of word reading efficiency*. Austin, TX: Pro-Ed.
van den Boer, M., de Jong, P. F., & Haentjens-van Meeteren, M. M. (2012). Lexical decision in children: Sublexical processing or lexical search? *Quarterly Journal of Experimental Psychology, 65*, 1214–1228.
van den Bosch, A., Content, A., Daelemans, W., & de Gelder, B. (1994). Measuring the complexity of writing systems. *Journal of Quantitative Linguistics, 1*, 178–188.
Wechsler, D. (1992). *Manual for the wechsler intelligence scale for children–third edition U.K. edition*. Sidcup, UK: Psychological Corporation.
Weekes, B. S. (1997). Differential effects of number of letters on word and nonword naming latency. *Quarterly Journal of Experimental Psychology A, 50*, 439–456.
Yap, R., & van der Leij, A. (1993). Word processing in dyslexics: An automatic decoding deficit? *Reading and Writing, 5*, 261–279.

Ziegler, J., & Goswami, U. (2005). Reading acquisition, developmental dyslexia, and skilled reading across languages: A psycholinguistic grain size theory. *Psychological Bulletin, 131*, 3–29.
Ziegler, J. C., Perry, C., Jacobs, A. M., & Braun, M. (2001). Identical words are read differently in different languages. *Psychological Science, 12*, 379–384.
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.
Zoccolotti, P., de Luca, M., di Pace, E., Gasperini, F., Judica, A., & Spinelli, D. (2005). Word length effect in early reading and in developmental dyslexia. *Brain and Language, 93*, 369–373.