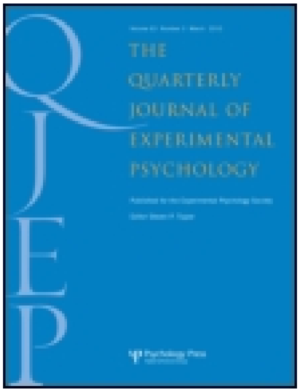
Lexical decision in children: Sublexical processing or lexical search?

van den Boer, M.; de Jong, P.F.; Haentjens-van Meeteren, M.M.

DOI
10.1080/17470218.2011.652136
Publication date
2012
Document Version
Final published version
Published in
The Quarterly Journal of Experimental Psychology



The Quarterly Journal of Experimental Psychology
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/pqje20

Lexical decision in children: Sublexical processing or lexical search?

Madelon van den Boer [a], Peter F. de Jong [a] & Marleen M. Haentjens-van Meeteren [a]

[a] Research Institute of Child Development and Education, University of Amsterdam, Amsterdam, The Netherlands

Accepted author version posted online: 21 Dec 2011.Version of record first published: 10 Apr 2012.

To cite this article: Madelon van den Boer, Peter F. de Jong & Marleen M. Haentjens-van Meeteren (2012): Lexical decision in children: Sublexical processing or lexical search?, The Quarterly Journal of Experimental Psychology, 65:6, 1214-1228

To link to this article: http://dx.doi.org/10.1080/17470218.2011.652136

THE QUARTERLY JOURNAL OF EXPERIMENTAL PSYCHOLOGY
2012, 65 (6), 1214–1228

Psychology Press
Taylor & Francis Group

Lexical decision in children: Sublexical processing or lexical search?

Madelon van den Boer, Peter F. de Jong, and Marleen M. Haentjens-van Meeteren

Research Institute of Child Development and Education, University of Amsterdam, Amsterdam, The Netherlands

Length effects in the lexical decision latencies of children might indicate that children rely on sublexical processing and essentially approach the task as a naming task. We examined this possibility by means of the effects of neighbourhood size and articulatory suppression on lexical decision performance. Sixty-six beginning and 62 advanced readers performed a lexical decision task in a standard, articulatory suppression, or tapping condition. We found length effects on words and nonwords in the children's lexical decisions. However, the effects of neighbourhood size were similar to those reported for adult lexical decisions, rather than the effects previously found in children's naming. In addition, no effect was found of articulatory suppression. Both findings suggest that, despite clear length effects, children do not adopt a naming task approach but, like adults, base lexical decisions mainly on a lexical search. These results pose a challenge for several computational models of reading.

Keywords: Reading development; Lexical decision; Sublexical processing; Word length effect; Neighbourhood size.

Over the course of reading development, sensitivity to the number of letters in a word tends to decrease. In beginning readers, word-naming latencies increase when words become longer (e.g., Spinelli et al., 2005; Zoccolotti et al., 2005). In advanced readers, however, the effect of word length is much smaller, although a length effect can still be observed, especially in nonwords (e.g., Bates, Burani, d'Amico, & Barca, 2001; Weekes, 1997; Ziegler, Perry, Jacobs, & Braun, 2001). The decrease of the length effect has often been taken to reflect a gradual shift from a sublexical letter-by-letter reading strategy towards a lexical strategy in which all letters in a word are processed in parallel.

Length effects have been a key issue in the cognitive modelling of reading. The interpretation of length effects as a marker of sublexical processing fits well within the dual-route cascaded (DRC) model, as an example of a larger class of dual-route models (Coltheart, Rastle, Perry, Langdon, & Ziegler, 2000; Jackson & Coltheart, 2001). The DRC model distinguishes between two routes of orthography to phonology conversion. The nonlexical route serially decodes the letters in a string into phonemes via the application of knowledge of grapheme–phoneme relationships. In the lexical route, all the letters in a string are processed in parallel and activate successively a word's

Correspondence should be addressed to Madelon van den Boer, University of Amsterdam, Research Institute of Child Development and Education, P.O. Box 94208, 1090 GE Amsterdam, The Netherlands. E-mail: m.vandenboer@uva.nl

http://www.psypress.com/qjep
http://dx.doi.org/10.1080/17470218.2011.652136

representation in the orthographic and phonological lexicon. Both routes work in parallel, but familiar words are read mainly through the lexical route, whereas the nonlexical route dominates in processing unfamiliar words and nonwords. Thus, length effects in reading aloud are explicitly modelled within the framework of the DRC model as a consequence of serial processing in the nonlexical route.

In parallel distributed processing (PDP) models, in contrast, the effect of length has generally not been modelled explicitly (Plaut, McClelland, Seidenberg, & Patterson, 1996; Seidenberg & McClelland, 1989; but for a recent exception, see Monaghan & Ellis, 2010). In PDP models, lexical knowledge is represented as patterns of activation over orthographic, phonological, and semantic layers of units. Lexical processing occurs via the connections among these units. The weights on these connections are formed and adjusted during the learning process and come to reflect the degree of consistency in the input. Within this associative network, there is no lexical memory for individual words. Instead, the system's knowledge of spelling–sound correspondences, anchored in the connection weights, is used to process all types of letter strings, words and nonwords alike.

In PDP models, length effects in reading aloud are mainly ascribed to orthographic neighbourhood size (Seidenberg & Plaut, 1998). The orthographic neighbourhood of a given word consists of all existing words that can be created by changing one of its letters (Coltheart, Davelaar, Jonasson, & Besner, 1977; but see for an alternative, Yarkoni, Balota, & Yap, 2008). Generally, a word is less well represented in the connection weights of the model if it has a smaller number of neighbours. As longer words tend to have fewer neighbours than shorter words (e.g., Balota, Cortese, Sergent-Marshall, Spieler, & Yap, 2004; Frauenfelder, Baayen, & Hellwig, 1993; Monaghan & Ellis, 2010), it follows that naming longer words will take more time.

There is extensive evidence for the effect of length on children's naming latencies. For example, for first and second graders, as well as children with dyslexia, length effects have been observed for words of two to five letters (e.g., Marinus & de Jong, 2010b; Ziegler, Perry, Ma-Wyatt, Ladner, & Schulte-Körne, 2003; Zoccolotti et al., 2005). For sixth through eighth graders, length affected naming latencies for words ranging from three to eight letters (Spinelli et al., 2005). In the current study, however, we considered length effects in lexical decisions.

The most common and well-documented finding for lexical decisions in adults is that words and nonwords of varying length are responded to equally fast (e.g., Coltheart et al., 2000; Juphard, Carbonnel, & Valdois, 2004). Although Balota et al. (2004) showed that length did predict lexical decision performance for nonwords of two to eight letters, the effects of length in lexical decisions were smaller than in naming latencies. A number of studies focused on the effects of (non)word length on lexical decisions in children. Martens and de Jong (2006) reported a length effect for words and nonwords of three to six letters in Dutch children. These length effects were larger for nonwords than for words, and larger for second-grade beginning readers than for fourth-grade advanced readers. Similar effects have been reported in children learning to read Italian (Filippo, de Luca, Judica, Spinelli, & Zoccolotti, 2006) or Spanish (Acha & Perea, 2008).

If length effects in lexical decisions should be interpreted in the same way as length effects in naming, within the DRC framework, an effect of word length would imply that children use the nonlexical route. However, in the DRC model, (adult) lexical decisions are modelled solely within the lexical route. As the result of a search of the orthographic lexicon, either a word's representation is activated, and a "yes" response is made, or no match is found between an entry in the lexicon and the stimulus, in which case "no" will be the answer. Thus, different from reading aloud, for lexical decision no specific role is proposed for the nonlexical route, in line with the absence of length effects in adult lexical decisions. For children, however, length effects are found in lexical decisions. To accommodate their findings within the DRC model, Martens and de Jong (2006) suggested that a lexical decision task might be approached by children as a naming task. Children arguably have a smaller orthographic lexicon. By identifying the (non)word through the nonlexical route, children access their (larger) phonological lexicon before reaching a decision.

The aim of the current study was to replicate and extend previous findings on length effects in lexical decisions in children. More specifically, we focused on how these length effects should be interpreted by examining whether lexical decision in children is similar to naming. Our focus was twofold; we investigated the effects of neighbourhood size and articulatory suppression. The effects of neighbourhood size are of interest for two reasons. First, according to PDP models, neighbourhood size is a possible alternative explanation for length effects, yet previous studies on length effects in children have not controlled for neighbourhood size (e.g., Martens & de Jong, 2006; Spinelli et al., 2005; Zoccolotti et al., 2005). In addition, in a large data set of adult lexical decisions, it has been shown that controlling for neighbourhood size is important when studying length effects. When neighbourhood size was controlled for, contrary to what has previously been reported, New, Ferrand, Pallier, and Brysbaert (2006) found that length effects are present in lexical decision latencies of adults, but follow a U-shaped curve, with a decrease in reaction times for words up to 5 letters, no length effect for 6- to 8-letter words, and the traditional increasing length effect for words of 9 to 13 letters.

Second, the type of neighbourhood size effect can shed light on how children approach the task. To our knowledge, effects of neighbourhood size on children's lexical decisions have not been studied before. In adults, the general finding is that lexical decisions for words are faster as the neighbourhood size increases. These effects tend to be most pronounced for words of low frequency. In contrast, for nonwords, larger neighbourhoods yield longer latencies (Balota et al., 2004; Carreiras, Perea, & Grainger, 1997; Fiebach, Ricker, Friederici, & Jacobs, 2007). In naming, the effects of neighbourhood size have been shown to be similar in children and adults. Naming latencies of high frequent words were not affected by neighbourhood size, whereas nonwords with a larger neighbourhood were named faster than nonwords with a smaller neighbourhood (Andrews, 1997; Balota et al., 2004; Marinus & de Jong, 2010a, 2010b). Accordingly, if children indeed approach the lexical decision task as a naming task, the effects of neighbourhood size are expected to mirror these effects in naming. However, if the effects of neighbourhood size resemble the effects found in adult lexical decision, it seems more likely that children, like adults, base lexical decisions mainly on lexical search.

The second focus of this study concerned the effects of articulatory suppression on lexical decision performance. As said, within the DRC model, length effects are mainly interpreted in terms of the use of the nonlexical route. If the effects of neighbourhood size indicate a naming task approach, length effects could be ascribed to sublexical processing. However, to further strengthen this interpretation, it seems important to find independent evidence for the involvement of the nonlexical route. Therefore, we focus on the key aspect of the nonlexical route, namely, the serial recoding of letters into sounds, for which we use the term *phonological recoding*. One way to investigate whether phonological recoding occurs during lexical decision has been articulatory suppression (e.g., Arthur, Hitch, & Halliday, 1994; Tenjović & Lalović, 2005). Articulatory suppression, or the continuous repetition of a nonsense word, is expected to interfere with phonological recoding.

In Serbian, an orthographically shallow language, it was found that articulatory suppression led to an increase in both reaction times and error rates in a lexical decision task (Tenjović & Lalović, 2005). In English, however, it has been shown that phonology can be derived from print and used for lexical access without interference from articulatory suppression (Besner, 1987). But, if the suppression rate was fast enough, accuracy did decrease in a variety of tasks that required the manipulation of the phonology derived from print (Besner, Davies, & Daniels, 1981). However, these studies all included adult participants, who according to the DRC model, rely on the lexical route only to make a lexical decision.

One of the few studies focusing on children is the study by Arthur et al. (1994), who tested 7- to 9-year-olds learning to read English on a lexical decision task. The authors assumed that the children would use a phonological recoding strategy to identify words and nonwords and therefore expected that articulatory suppression would hamper lexical decision performance. Surprisingly, articulatory suppression appeared to have a positive effect on lexical decisions. Children in the articulation condition showed shorter response latencies than children who were subjected to a standard lexical decision task. Arthur et al. suggested that children, like adults, do not need to phonologically recode stimuli to reach a lexical decision. However, in the standard lexical decision condition, the phonology of a stimulus was used as a check on its lexical status. According to Arthur et al., this check explained the difference in processing time between the standard and articulation conditions. Unfortunately, Arthur et al. did not provide independent evidence for a phonological check, because they did not determine whether a length effect was present in the standard lexical decision condition. In addition, the authors note that a phonological recoding strategy is most likely when instruction of this type is common in education, which was not the case for the children in their study.

In the current study, we investigated the effect of articulatory suppression during lexical decision in Dutch children in Grade 2 and Grade 5, representing beginning and advanced readers, respectively. Dutch beginning readers are explicitly instructed to use phonological recoding during word reading, because in orthographically shallow languages, like Dutch, grapheme–phoneme conversion leads to reliable word identification. In a standard lexical decision control condition, we established whether length effects were present. We also included a control condition in which children performed a nonphonological tapping task. Thereby we were able to rule out the possibility that it is performing a secondary task in general rather than articulatory suppression per se that disrupts phonological recoding. If beginning readers phonologically recode stimuli, and essentially approach the lexical decision task as a naming task, we expected length effects in the standard condition, whereas we expected that articulatory suppression would disrupt phonological recoding and force children to adopt a lexical search strategy, causing the length effect to disappear. Advanced readers, in contrast, who have developed a more extensive orthographic lexicon, were expected, like adults, to primarily rely on lexical processing when performing a lexical decision task. Therefore, we expected that advanced readers would not show a length effect, nor be disrupted by articulatory suppression.

Method

Participants

Sixty-seven second-grade children and 62 fifth-grade children participated in this study. These children attended regular elementary education in one of three participating schools in the western part of the Netherlands. All children had normal or corrected-to-normal vision.

The children were selected from a larger group of 88 second-grade and 71 fifth-grade children, based on word reading ability, assessed with the One Minute Reading Test (*Eén Minuut Test*; Brus & Voeten, 1995). This test consists of two lists of 116 words of increasing difficulty, which children were asked to read aloud as quickly and accurately as possible, within the given timeframe of one minute. These reading scores were used to select normally reading children who were either beginning or advanced readers. A total of 3 second-grade children and 9 fifth-grade children were excluded from participation due to a reading score below the 10th percentile, indicating a reading lag of at least 1.5 years. In addition, to avoid overlap between the groups of readers, 18 second-grade children were excluded because of a reading score above the 90th percentile, indicating that the reading levels of these children corresponded to Grade 5 or above.

The remaining 67 second-grade children with a mean age of 7 years 11 months (40 boys, 27 girls) represent normal but beginning readers, with a reading age of 1 year 6 months at the time of testing. The remaining 62 fifth-grade children with a mean age of 11 years 0 months (36 boys, 26 girls) represent normal but advanced readers, with a reading age of 4 years 6 months.

Materials

The stimulus set consisted of 45 words and 45 nonwords. The items varied in length from three to five letters. Fifteen monosyllabic words and nonwords were selected for every word length. The words were all high-frequency words, selected from a corpus of child literature (Schrooten & Vermeer, 1994). Across lengths, words were matched on onset (i.e., the first phoneme; always a single-letter grapheme) and frequency.

The nonwords were created by interchanging onsets and rhymes of the selected words. For example the words “grap”, “kwal”, and “druk” (meaning *joke, jellyfish*, and *busy*, respectively) were used to create the nonwords “gruk”, “kwap”, and “dral”. When the resulting nonword was unpronounceable or also a Dutch word, a rhyme was chosen that was as close as possible to the original. Due to this procedure, words and nonwords within a length condition were matched on onset and consonant–vowel (CV) structure. Also, within each length condition, words and nonwords were matched on number of orthographic and phonological neighbours (Baayen, Piepenbrock, & van Rijn, 1993). Table 1 presents the characteristics of the words and nonwords.

In addition to the stimulus set, 15 filler words and 15 filler nonwords were created that were similar to the stimulus words and nonwords. These items were used as a break in the experiment (as explained below) and are not included in the analyses.

Design and procedure

The words and nonwords were presented in a lexical decision task. Within both grades, children were randomly assigned to three conditions. Triplets of children of the same gender and with comparable age and reading scores were formed, and children within a triplet were then randomly assigned to the conditions. In the first condition, children performed a standard lexical decision task. In the second condition, the articulatory suppression condition, children performed the lexical decision task, while continuously repeating the nonsense word “DUBBA” at a rate of approximately one time per second. In the third condition, the tapping condition, children performed the lexical decision task, while continuously tapping their feet on the floor, alternating between the left and right foot, at a rate of approximately one tap per second.

The task was programmed in E-prime Version 1.0 (Schneider, Eschman, & Zuccolotto, 2002). Words and nonwords were presented one by one in the middle of a laptop screen of 14.1 inches (35.8 cm). The stimuli were printed in 46-point Arial font with black, lower-case letters, on a white background. The computer registered latencies and accuracy. Response latencies were recorded from the onset of the presentation of the stimulus until the child responded by pressing a button.

Every trial started with a fixation sign (“ + ”), presented in the middle of the screen. After 2000 ms, a stimulus appeared, which remained on the screen until the child made a response. Immediately after the child responded, a mask (“XXXXX”) appeared for 1500 ms, followed by a minus sign (“–”), which remained on the screen until the experimenter initiated the following trial by clicking the mouse key. Children in the articulatory suppression and tapping conditions were instructed to perform their secondary task from the onset of the fixation sign until the appearance of the minus sign.

Every child was tested individually on a total of four experimental blocks. The blocks consisted of, respectively, 25, 20, 25, and 20 lexical decision trials. Children were instructed to read each letter string carefully and indicate whether the letter string on the screen was a word (green button; located over the “c” on the keyboard) or a nonword (red button; located over the “m” on the keyboard). They were instructed to guess for stimuli they could not read. The blocks of trials were separated by a block of filler items and a

Table 1. *Frequency, orthographic neighbourhood, and phonological neighbourhood of the words and nonwords*

	Length		
Frequency and neighbourhood	*3*	*4*	*5*
Words			
Word frequency[a]	200 (186)	199 (181)	199 (183)
Orthographic *N*[b]	19.5 (5.0)	6.9 (2.7)	3.9 (2.2)
Phonological *N*[b]	26.3 (8.9)	10.5 (5.0)	3.3 (2.1)
Nonwords			
Orthographic *N*[b]	18.9 (5.1)	8.5 (3.4)	4.1 (2.9)
Phonological *N*[b]	22.3 (7.3)	10.4 (4.7)	3.1 (2.1)

Note: N = neighbourhood. Means (standard deviations in parentheses).
[a]Schrooten & Vermeer, 1994. [b]Baayen, Piepenbrock, & van Rijn, 1993.

break. During the 10 filler trials, children in the articulatory suppression and tapping conditions were allowed to temporarily stop their secondary task, to keep them motivated, even if they were struggling with the task. The filler trials were followed by a fixed break of 1.5 minutes. After the break, the next experimental block would start, until the child had been presented with each of the 90 stimuli and 30 filler items once.

Preceding the experimental trials, four blocks of practice were used to gradually familiarize the child with the procedure. During the first four trials, the children in the standard condition would simply look at the screen, while the children in the articulatory suppression and tapping conditions practised articulation and tapping, respectively. During the next four trials, children were presented with an arrow and indicated whether the arrow pointed to the left or to the right by pressing either the red (left) or the green (right) button. Children in the articulatory suppression and tapping conditions kept performing their secondary task, throughout all practice blocks. In the final two blocks of practice, children were introduced to the lexical decision task and responded to a total of 6 words and 6 nonwords.

Data analysis

The data were analysed using multilevel modelling. The reaction times and error rates collected in the current study are embedded in a hierarchical structure with two levels. The various responses to items (Level 1) are nested under individuals (Level 2). The tutorial by Quené and van den Bergh (2004) explains that within a multilevel model, or hierarchical regression model, random factors from participants and items can be captured within one model, rendering separate analyses by participants and items redundant. The model has more statistical power than alternatives such as the repeated measures analysis of variance (RM-ANOVA), because analyses are based on the reaction times to all separate items, instead of a mean latency score per person per condition. Other advantages of a multilevel approach are the opportunity to model variances and covariances explicitly at each level of the hierarchy, as well as the robustness against missing data, since parameters are estimated based on all available responses to items, instead of means per condition.

Apart from these less stringent assumptions, however, multilevel models remain a form of general linear modelling. Various variables can be included, and all effects within a model are estimated simultaneously. We first specified a model for the length effect. Next, we added neighbourhood size at Level 1 and estimated the effect of (non)word length corrected for neighbourhood size. This model is similar to an analysis of covariance with neighbourhood size as a covariate.

The analyses were conducted with MLwiN 2.12 (Rasbash, Steele, Browne, & Goldstein, 2008). Differences among parameter estimates were tested with a chi-square test statistic. Different models were specified to test the various hypotheses. In each model, between-item differences (or within-subject differences) were modelled on the first level and between-subject differences on the second level. Separate models were specified for response latency and error data, assuming a normal and binomial distribution, respectively. Reaction times were modelled via dummy variables that were specified for every condition. Parameters for these dummy variables represent the condition means. The error data were dummy coded (0 is incorrect, 1 is correct). Therefore, to analyse these data, a logistic regression procedure was used, but the models were the same as those for the reaction data.

Initially, the response latency models were specified to the raw data. However, because we aim to directly compare beginning readers in Grade 2 to advanced readers in Grade 5, especially for the effects of neighbourhood size, we controlled for the longer overall reaction times that are expected for beginning versus advanced readers. This difference might affect the interpretation of possible interactions between the groups and experimental manipulations, because a significant effect can reflect an absolute difference or just a proportional difference in reaction to the experimental manipulation. Because responses to each item are included in the analysis, we calculated within-subjects *z* scores, by subtracting the subject's overall mean latency score from every item and dividing by the standard deviation of the subject's latency score distribution, based on the 90 word and nonword items.

Results

The results are presented in four sections. First, we describe the process of data cleaning. Second, we focus on the effects of experimental condition (standard, articulatory suppression, and tapping) on error rates and latencies. Words and nonwords are analysed separately. In the third section, data on words and nonwords are combined, and we focus on interactions between length, grade, and lexicality. With these analyses, we aim to replicate the length effects found in previous studies. In the fourth and final section, however, we include neighbourhood size in the model and discuss the main effects of neighbourhood size on word and nonword error rates and latencies, as well as the specific effect of length after neighbourhood size is added as a covariate.

Data cleaning

Data from one child in Grade 2 were excluded from the analyses, because both the reaction times and the accuracy scores of this particular child were identified as an outlier at the group level. The Grade 2 analyses were based on data from the remaining 66 children. Furthermore, the analyses were conducted on valid and correct trials only. Trials were excluded from further analysis if the response was incorrect (12.7% in Grade 2 and 6.5% in Grade 5). In addition, reaction times lower than 350 ms or higher than 6000 ms (0.8% in Grade 2 and 0.0% in Grade 5) as well as reaction times 3 standard deviations above the participants' mean (1.3% in Grade 2 and 1.4% in Grade 5) were removed. The reported analyses were based on 85.2% of the trials for Grade 2 and on 92.1% of the trials for Grade 5. The mean error rates per grade, condition, and stimulus type can be found in Table 2. Table 3 contains the mean reaction times.

Condition effects

Two models were specified: one for words and another for nonwords. For each model, separate dummy variables were created for each condition by length by grade combination, amounting to a total of 18 variables per model. To test for the main effects of length and condition, as well as their interaction, each effect was split in two contrasts, which were tested simultaneously in a multivariate test, using a chi-square statistic (χ^2) with two degrees of freedom (Tabachnick & Fidell, 2001). For length, the contrasts specified the differences between three- and four- and between four- and five-letter words. Condition effects were tested with contrasts denoting the differences between the standard and both of the "extra task" conditions and between the articulatory suppression and tapping conditions. The reported analyses were

Table 2. *Mean error rates per grade, condition, and stimulus type*

	Words			*Nonwords*		
Grade and condition	*3*	*4*	*5*	*3*	*4*	*5*
Grade 2						
Standard	12.4 (8.8)	11.5 (11.9)	14.2 (10.3)	18.8 (12.8)	13.9 (6.9)	12.4 (10.1)
Articulation	9.4 (7.4)	13.9 (11.5)	10.3 (9.4)	21.2 (17.5)	17.3 (10.2)	17.0 (11.2)
Tapping	12.7 (7.5)	10.0 (9.2)	7.9 (11.1)	23.6 (16.6)	19.1 (13.1)	20.0 (12.7)
Total	11.5 (8.1)	11.8 (11.0)	10.8 (10.6)	21.2 (15.9)	16.8 (10.6)	16.5 (11.8)
Grade 5						
Standard	9.0 (7.7)	5.0 (7.3)	6.0 (6.6)	6.0 (7.0)	9.3 (11.6)	9.3 (10.4)
Articulation	7.3 (6.1)	7.0 (6.7)	7.0 (6.7)	5.7 (5.2)	11.4 (9.5)	9.2 (6.3)
Tapping	8.3 (6.8)	7.6 (9.0)	7.9 (7.9)	8.9 (6.9)	7.9 (7.3)	8.6 (7.7)
Total	8.2 (6.9)	6.6 (7.8)	7.0 (7.1)	6.9 (6.6)	9.6 (9.7)	9.0 (8.3)

Note: Error rates in percentages. Standard deviations in parentheses.

Table 3. *Mean reaction times per grade, condition, and stimulus type*

	Words			*Nonwords*		
Grade and condition	*3*	*4*	*5*	*3*	*4*	*5*
Grade 2						
Standard	1363 (610)	1464 (646)	1555 (736)	1739 (699)	1834 (719)	1968 (720)
Articulation	1483 (592)	1485 (585)	1526 (628)	1741 (647)	1938 (804)	1953 (846)
Tapping	1390 (592)	1501 (771)	1556 (851)	1761 (834)	1992 (1010)	2015 (1010)
Total	1413 (600)	1484 (673)	1546 (744)	1747 (729)	1920 (852)	1978 (862)
Grade 5						
Standard	850 (261)	822 (247)	823 (293)	943 (276)	962 (297)	978 (280)
Articulation	903 (314)	896 (260)	872 (282)	1027 (305)	1036 (349)	1043 (361)
Tapping	888 (227)	846 (202)	856 (229)	1007 (276)	1025 (246)	1008 (241)
Total	881 (271)	855 (240)	851 (269)	993 (288)	1008 (301)	1010 (299)

Note: Reaction times in milliseconds. Standard deviations in parentheses.

based on the unstandardized data, presented in Table 2 (error rates) and Table 3 (reaction times).

Words and nonwords; errors and response latencies. The results for words and nonwords were identical for both error rates and reaction times. The main effects of condition were not significant, nor were there significant interactions between condition and grade, or condition and length. Taken together, these results indicate that the error rates and mean reaction times of second and fifth graders were not affected by task condition, nor did the conditions of the task affect the length effects for words or nonwords.

Lexicality and length effects

Because the length effects for words and nonwords in Grade 2 and Grade 5 did not differ across task conditions, we specified a new model without the factor task condition. Both words and nonwords were included in this model to examine the effects of lexicality. Dummy variables were created for each length by lexicality by grade combination, amounting to a total of 12 variables.

Errors. The significant three-way interaction between length, lexicality, and grade, $\chi^2(2) = 8.14$, $p < .05$, indicated that length effects in the error rates were not the same for words and nonwords, nor for Grade 2 and Grade 5. Second graders made more errors in responding to nonwords than to words, $\chi^2(1) = 15.65$, $p < .001$. For words, error rates were similar across different lengths. For nonwords, however, the length effect was significant, $\chi^2(2) = 15.59$, $p < .001$. More specifically, second graders made more errors in responding to three-letter nonwords than to four- or five-letter nonwords, $\chi^2(1) = 9.46$, $p < .01$. For fifth graders, error rates did not differ between words and nonwords. For words, error rates were also similar across different lengths. For nonwords, however, the length effect approached significance, $\chi^2(2) = 5.39$, $p = .068$. More specifically, different from second graders, fifth graders made more errors in responding to four- or five-letter nonwords than to three-letter nonwords, $\chi^2(1) = 4.51$, $p < .05$.

Response latencies. The main effects of length and lexicality were both significant, $\chi^2(2) = 32.09$, $p < .001$, and $\chi^2(1) = 213.77$, $p < .001$, respectively, indicating that in general longer items yielded longer reaction times than shorter items, and nonwords yielded longer reaction times than words. The three-way interaction between length, lexicality, and grade was not significant. But, the two-way interactions between lexicality and length, $\chi^2(2) = 13.97$, $p < .001$, length and grade, $\chi^2(2) = 35.34$, $p < .001$, and lexicality and grade, $\chi^2(1) = 46.84$, $p < .001$, were significant.

In Grade 2, the overall length effect was significant for both words, $\chi^2(2) = 18.26$, $p < .001$, and nonwords, $\chi^2(2) = 38.38$, $p < .001$. Words of three letters yielded shorter reaction times than words of four letters, $\chi^2(1) = 6.00$, $p < .05$, which in turn yielded shorter reaction times than words of five letters, $\chi^2(1) = 7.68$, $p < .01$. The same pattern was found for nonwords of three versus four letters, $\chi^2(1) = 24.41$, $p < .001$, and of four versus five letters, $\chi^2(1) = 10.21$, $p < .01$. The length effect for nonwords was larger than the length effect for words, $\chi^2(2) = 7.15$, $p < .05$, but this difference was specific to the length effect from three to four letters, $\chi^2(1) = 4.97$, $p < .05$. The length effect from four to five letters did not differ significantly between words and nonwords.

In Grade 5, the overall length effect was significant for words, $\chi^2(2) = 10.76$, $p < .01$, but not for nonwords. Surprisingly, words of three letters yielded longer reaction times than words of four letters, $\chi^2(1) = 6.17$, $p < .05$, which were responded to equally as fast as five-letter words. Reaction times to nonwords with three, four, and five letters were similar. The length effect for words, characterized by shorter reaction times to longer words, was significantly different from the length effect for nonwords, characterized by similar reaction times to nonwords of varying length, $\chi^2(2) = 9.20$, $p < .05$, but this difference was specific to the length effect from three to four letters, $\chi^2(1) = 5.64$, $p < .05$. The length effect from four to five letters did not differ significantly between words and nonwords.

The overall reaction times were longer for children in Grade 2 than for children in Grade 5. As explained before, the interaction effects between lexicality and grade and between length and grade might be a consequence of this difference in overall reaction time. In other words, the interaction effects might reflect proportional differences. Therefore, the analyses were repeated using within-subject standardized reaction times. Due to this standardization, parameter estimates were also standardized and can be interpreted as effect sizes. The first and second columns of values of Table 4 contain the parameter estimates for the effects in this model. The interactions between lexicality and length, $\chi^2(2) = 17.93$, $p < .001$, and length and grade, $\chi^2(2) = 35.88$, $p < .001$, remained significant, indicating that these are true interactions, not proportional differences. However, the interaction between lexicality and grade was no longer significant. Thus, differences in the effects of lexicality are proportional. When overall reaction times were controlled for, children in Grade 2 and Grade 5 were similarly affected by lexicality.

Neighbourhood size

Next, we added neighbourhood size as a covariate. Because the distribution of the number of neighbours was skewed, orthographic neighbourhood size was log-transformed and standardized for words and nonwords separately. Four dummy variables were

Table 4. *Effect sizes and beta-coefficients of the standardized models*

	N size excluded		*N size included*	
Parameters	*Grade 2*	*Grade 5*	*Grade 2*	*Grade 5*
Words				
3 letters	−.432 (.033)	−.224 (.033)	−.369 (.050)	−.045 (.052)
4 letters	−.335 (.033)	−.334 (.034)	−.346 (.033)	−.365 (.035)
5 letters	−.229 (.037)	−.384 (.032)	−.281 (.048)	−.530 (.045)
Nonwords				
3 letters	.163 (.039)	.284 (.035)	.101 (.049)	.164 (.045)
4 letters	.395 (.034)	.331 (.037)	.395 (.034)	.331 (.037)
5 letters	.515 (.035)	.348 (.033)	.582 (.048)	.476 (.045)
N size				
Words	—	—	−.056* (.033)	−.157*** (.035)
Nonwords	—	—	.066** (.031)	.126*** (.029)

Note: Standard errors in parentheses.
*$p < .10$. **$p < .05$. ***$p < .001$.

specified: per grade, one dummy for the effect of neighbourhood size on words and one for the effect of neighbourhood size on nonwords. These four variables were added to the model described in the previous section, including a dummy variable for each length by lexicality by grade combination.

Errors. The effect of neighbourhood size on words was not significant in Grade 2, nor in Grade 5. For nonwords, however, both second and fifth graders made more errors in responding to nonwords with larger neighbourhood sizes than in responding to nonwords with smaller neighbourhood sizes, $\chi^2(1) = 24.98$, $p < .001$, and $\chi^2(1) = 28.59$, $p < .001$, respectively. The effects of neighbourhood size on the error rates of words, nonwords, or words and nonwords combined did not differ between second and fifth grade. As in the previous analyses, the length effects in error rates to nonwords were significant, both in Grade 2, $\chi^2(2) = 9.20$, $p < .05$, and in Grade 5, $\chi^2(2) = 31.08$, $p < .001$. In Grade 2, shorter nonwords yielded higher error rates than longer nonwords, whereas in Grade 5, longer nonwords yielded higher error rates than shorter nonwords.

Response latencies. Because the main interest was to compare the parameter estimates of neighbourhood size over stimulus types and grades, analyses were based on the standardized reaction times. The third and fourth columns of values of Table 4 contain the parameter estimates for the effects in this model. Due to standardization, the estimates for neighbourhood size can be interpreted as beta-coefficients.

The effect of neighbourhood size on words approached significance in Grade 2, $\chi^2(1) = 2.78$, $p = .095$, and was significant in Grade 5, $\chi^2(1) = 20.41$, $p < .001$. For both second and fifth graders, words with a larger neighbourhood size yielded shorter reaction times than words with a smaller neighbourhood size. However, this effect was larger for fifth graders than for second graders, $\chi^2(1) = 4.40$, $p < .05$. With neighbourhood size in the model, the length effect in Grade 5 remained significant, $\chi^2(2) = 32.74$, $p < .001$. However, the length effect in Grade 2 was no longer significant, indicating that the length effect was overestimated when neighbourhood size was not controlled for.

Nonwords with a larger neighbourhood yielded longer reaction times than nonwords with a smaller neighbourhood in both second, $\chi^2(1) = 4.37$, $p < .05$, and fifth grade, $\chi^2(1) = 18.28$, $p < .001$. The magnitude of this effect did not differ significantly between second and fifth graders. With neighbourhood size in the model, the length effect in Grade 2 was again significant, $\chi^2(2) = 35.96$, $p < .001$. The length effect in Grade 5 was not significant in the previous analysis, but reached significance in this model, $\chi^2(2) = 17.04$, $p < .001$. For both second and fifth graders, longer nonwords yielded longer reaction times. The length effects did not differ significantly between Grade 2 and Grade 5. These results indicate that length effects in nonwords are underestimated when neighbourhood size is not controlled for.

The results showed that length effects changed when neighbourhood size was included in the model. However, the correlation between length and neighbourhood size is large ($r = -.84$, $p < .001$, for words, and $r = -.80$, $p < .001$, for nonwords). Such a strong relation increases the risk that results are affected by multicollinearity. However, we have two reasons to believe that this is not the case for our results. First, our correlations did not reach the .90 criterion suggested by Tabachnick and Fidell (2001). Second, multicollinearity can be detected by standard errors that increase up to three times, resulting in nonsignificant parameter estimates (Tabachnick & Fidell, 2001). In our models, however, standard errors were similar across models including only length and those including both length and neighbourhood size (see Table 4).

Discussion

In the current study, we investigated how length effects in children's lexical decisions can be interpreted. First, we examined whether length effects were present. In line with previous studies, we initially did not take neighbourhood size into account. We found that beginning readers show longer response latencies to longer items for both words and nonwords (Acha & Perea, 2008; Filippo et al., 2006; Martens & de Jong, 2006). Similar to adults, more advanced readers did not show a length effect for nonwords (Juphard et al., 2004). For words, however, in contrast to previous results, response latencies decreased when the number of letters increased.

Next, we controlled for neighbourhood size to test the possibility that length effects should actually be ascribed to differences in neighbourhood size. With neighbourhood size in the model, for words no length effect was found for beginning readers, whereas the decrease in response latencies in advanced readers became stronger. For nonwords, both beginning and advanced readers were affected by length, with longer reaction times to longer nonwords. These results corroborate the importance of accounting for neighbourhood size when examining length effects.

The length effect for words in our advanced readers is in line with the effect reported for adults by New et al. (2006), who also controlled for neighbourhood size. This suggests that fifth-grade readers and adults approach the lexical decision task similarly. New et al. focused solely on words. How neighbourhood size might affect length effects for nonwords has not previously been studied. The present results indicate that in advanced readers, the length effect is obscured due to differences in neighbourhood size. However, when neighbourhood size is controlled, a length effect is found. Interestingly, these results are in line with the different results reported by Juphard et al. (2004), who did not take neighbourhood size into account and did not find a length effect in nonwords, and the results of Balota et al. (2004), who did control for neighbourhood size and did find a length effect in nonwords.

The main issue remains how to interpret length effects in lexical decisions. We examined the possibility that length effects emerge because beginning readers approach the lexical decision task as a naming task. However, this possibility was not supported by our findings on neighbourhood size. An important finding of the current study is that words with a larger neighbourhood yielded shorter latencies, whereas nonwords with a larger neighbourhood yielded longer latencies. These effects of neighbourhood size mirror those found in adult lexical decisions (e.g., Carreiras et al., 1997), rather than in children's naming latencies (e.g., Marinus & de Jong, 2010a).

The effect of neighbourhood size on words was stronger in advanced than in beginning readers.

Although this difference was not significant for nonwords, it pointed in the same direction. Because advanced readers arguably have developed a larger orthographic lexicon than beginning readers, it is not surprising that the effect of neighbourhood size increases. However, it is surprising that the difference is larger for words than for nonwords. This pattern of results might indicate that few neighbours are sufficient to hinder nonword processing, whereas a larger neighbourhood is needed to facilitate word processing. Alternatively, the difference in the effect of neighbourhood size could be related to the relative frequency of stimuli and their neighbours. Marinus and de Jong (2010a) found that neighbours of higher frequency than the stimulus affected performance more than neighbours of lower frequency. For the nonwords, all neighbours were by default of higher frequency than the stimulus. For the words, however, two thirds of the stimuli had a neighbour of higher frequency (Baayen et al., 1993; Schrooten & Vermeer, 1994). Beginning readers might not yet know and, therefore, might not be affected by low-frequency neighbours of words.

In all, the effects of neighbourhood size on children's lexical decisions strongly suggest that children adopt a lexical search, rather than a phonological recoding strategy. This conclusion is further supported by the absence of an effect of articulatory suppression. Both beginning and advanced readers who were asked to repeatedly utter a nonword while performing the lexical decision task showed the same mean reaction times and length effect patterns as children performing either the standard or the tapping condition of the task. An effect of articulatory suppression would be expected if children approach lexical decision as a naming task, because then the nonlexical route might be involved. However, if lexical decisions are based on lexical search, in line with the DRC model, the nonlexical route is not involved, and, accordingly, an effect of articulatory suppression is not to be expected.

Our result is only partly in line with the reduced reaction times that Arthur et al. (1994) reported for the articulatory suppression condition. Both Arthur et al.'s and our results indicate that articulatory suppression does not hamper performance in a lexical decision task. However, in contrast to our results, Arthur et al. found that articulatory suppression even led to faster decision times. The authors suggested that children used a redundant phonological check in a standard lexical decision task, which disappeared in the articulatory suppression condition. Although it is difficult to guess as to the nature of these different outcomes, one possibility might be that the high-frequency words in our study, different from the average to high-frequency words used by Arthur et al., were known sufficiently well not to require a phonological check.

Because we did not have a task that is known to be affected by articulatory suppression, we cannot rule out the possibility that our articulatory suppression condition was not successfully implemented nor that our between subjects-design was not powerful enough to capture the effects. These possibilities, however, seem unlikely. In a previous study by our group, using a between-subjects design for the articulatory suppression and tapping conditions, with an equally large group of children of a similar age, the same articulatory suppression condition was found to have an effect on the orthographic learning of unfamiliar words (de Jong, Bitter, van Setten, & Marinus, 2009).

Interestingly, the current results suggest a slightly different interpretation for the findings of de Jong et al. (2009) and similar studies on orthographic learning (Kyte & Johnson, 2006; Share, 1999). In these studies, children performed a lexical decision task. The nonwords in this task were repeated. Orthographic learning was examined in a subsequent naming task including these nonwords as well as control nonwords that children had not previously been exposed to. Children read nonwords included in the lexical decision task faster than new nonwords. Based on the finding that homophone spellings of the target nonwords were also read faster than new nonwords (e.g., BLAIN versus BLANE), de Jong et al. argued that children had acquired a phonological representation of the nonwords through phonological recoding during the lexical decision task. However, this argument presumed that children approach the lexical decision task as a naming task. Given the results of the current study, that children's lexical decisions are based on lexical search, the results of de Jong et al. suggest that phonological codes are activated during lexical decisions enabling, through repeated exposure, the acquisition of a phonological representation of a novel word.

Support for a lexical search strategy in combination with length effects for nonwords poses a challenge for the parallel distributed processing model as well as for the dual-route model. The PDP model predicts effects of neighbourhood size (e.g., Plaut et al., 1996; Seidenberg & McClelland, 1989). However, neighbourhood size did not fully explain length effects. When neighbourhood size was controlled for, length effects were found in nonwords. The dual-route model predicts length effects when the nonlexical route dominates processing (e.g., Coltheart et al., 2000; Jackson & Coltheart, 2001). However, when lexical decisions are based on lexical search, length effects should not be found. Neither naming-like performance via the nonlexical route nor lexical search via the lexical route can fully explain the findings.

Evidence in favour of lexical search rather than phonological recoding might indicate that the origin of the length effect in lexical decision should be found in other aspects of the word recognition process, such as the initial stage of visual feature analysis of the letter string. The connectionist dual-process (CDP+) model, although not designed to account for lexical decision processes, might be interesting in this respect, because it assumes a larger role for the visual stage by means of a graphemic buffer in which a letter string is serially parsed into graphemes before phonemes come to play a role (Perry, Ziegler, & Zorzi, 2007). Since lexical decisions mainly depend on visual word recognition, a closer look at the role of the purely visual stage in processing a (non)word might result in new insights concerning length effects. It could be that additional letters in a stimulus yield longer processing times in a graphemic buffer while a visual representation is put together, whereas the link to phonological information is made equally fast for all items. A similar focus on letter identification was recently suggested by Risko, Lanthier, and Besner (2011), who found that interletter spacing elicited serial, but not sublexical, processing.

Taken together, the current study has shown that length effects are present in the lexical decisions of children. Neighbourhood size affected lexical decisions in children in the same way as in adults. No independent evidence was found for phonological recoding. Therefore, we conclude that children do not approach the lexical decision task as a naming task but, rather, like adults, adopt a lexical search strategy. However, because length effects were present, we suggest that in future studies (serial) visual processing might be considered as an alternative or additional factor in explaining length effects.

Original manuscript received 10 March 2011
Accepted revision received 24 November 2011
First published online 12 April 2012

REFERENCES

Acha, J., & Perea, M. (2008). The effects of length and transposed-letter similarity in lexical decision: Evidence with beginning, intermediate, and adult readers. *British Journal of Psychology*, *99*, 245–264.

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

Arthur, T. A. A., Hitch, G. J., & Halliday, M. (1994). Articulatory loop and children's reading. *British Journal of Psychology*, *85*, 283–300.

Baayen, R. H., Piepenbrock, R., & Van Rijn, H. (1993). *The CELEX lexical database [CD-ROM]*. Philadelphia, PA: University of Pennsylvania, Linguistic Data Consortium.

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

Bates, E., Burani, C., d'Amico, S., & Barca, L. (2001). Word reading and picture naming in Italian. *Memory & Cognition*, *29*, 986–999.

Besner, D. (1987). Phonology, lexical access in reading, and articulatory suppression: A critical review. *The Quarterly Journal of Experimental Psychology*, *39*, 467–478.

Besner, D., Davies, J., & Daniels, S. (1981). Reading for meaning: The effects of concurrent articulation. *The Quarterly Journal of Experimental Psychology*, *33,* 415–437.

Brus, B., & Voeten, B. (1995). *Eén minuut test vorm A en B. Verantwoording en handleiding [One-minute-test versions A and B. Accountability and manual]*. Lisse, The Netherlands: Swets & Zeitlinger.

Carreiras, M., Perea, M., & Grainger, J. (1997). Effects of orthographic neighbourhood in visual word recognition: Cross-task comparisons. *Journal of Experimental Psychology: Learning, Memory & Cognition*, *23,* 857–871.

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 Associates.

Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. (2000). DRC: A dual route cascaded model of visual word recognition and reading aloud. *Psychological Review*, *108,* 204–256.

de Jong, P. F., Bitter, D. J. L., van Setten, M., & Marinus, E. (2009). Does phonological recoding occur during silent reading, and is it necessary for orthographic learning? *Journal of Experimental Child Psychology*, *104,* 267–282.

Fiebach, C. J., Ricker, B., Friederici, A. D., & Jacobs, A. M. (2007). Inhibition and facilitation in visual word recognition: Prefrontal contribution to the orthographic neighbourhood size effect. *NeuroImage*, *36,* 901–911.

Filippo, G. D., de Luca, M., Judica, A., Spinelli, D., & Zoccolotti, P. (2006). Lexicality and stimulus length effects in Italian dyslexics: Role of the overadditivity effect. *Child Neuropsychology*, *12,* 141–149.

Frauenfelder, U., Baayen, R., & Hellwig, F. (1993). Neighbourhood density and frequency across languages and modalities. *Journal of Memory and Language*, *32,* 781–804.

Jackson, N. E., & Coltheart, M. (2001). *Routes to reading success and failure: Toward an integrated cognitive psychology of atypical reading*. New York, NY: Taylor & Francis.

Juphard, A., Carbonnel, S., & Valdois, S. (2004). Length effect in reading and lexical decision: Evidence from skilled readers and a developmental dyslexic participant. *Brain and Cognition*, *55,* 332–340.

Kyte, C. S., & Johnson, C. J. (2006). The role of phonological recoding in orthographic learning. *Journal of Experimental Child Psychology*, *93,* 166–185.

Marinus, E., & de Jong, P. F. (2010a). Size does not matter, frequency does: Sensitivity to orthographic neighbours in normal and dyslexic readers. *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.

Martens, V. E. G., & de Jong, P. F. (2006). The effect of word length on lexical decision in dyslexic and normal reading children. *Brain and Language*, *98,* 140–149.

Monaghan, P., & Ellis, A. W. (2010). Modeling reading development: Cumulative, incremental learning in a computational model of word naming. *Journal of Memory and Language*, *63,* 506–525.

New, B., Ferrand, L., Pallier, C., & Brysbaert, M. (2006). Reexamining the word length effect in visual word recognition: New evidence from the English lexicon project. *Psychonomic Bulletin & Review*, *13,* 45–52.

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.

Plaut, D. C., McClelland, J. L., Seidenberg, M. S., & Patterson, K. (1996). Understanding normal and impaired word reading: Computational principles in quasi-regular domains. *Psychological Review*, *103,* 56–115.

Quené, H., & van den Bergh, 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. J., & Goldstein, H. (2008). *A user's guide to MLwiN Version 2.10*. Bristol, UK: University of Bristol, Centre for Multilevel Modelling.

Risko, E. F., Lanthier, S. N., & Besner, D. (2011). Basic processes in reading: The effect of interletter spacing. *Journal of Experimental Psychology: Learning, Memory & Cognition*, *37*, 1449–1457.

Schneider, W., Eschman, A., & Zuccolotto, A. (2002). *E-prime: User's guide*. Pittsburgh, PA: Psychology Software.

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.

Seidenberg, M. S., & McClelland, J. L. (1989). A distributed, developmental model of word recognition and naming. *Psychological Review*, *96,* 523–568.

Seidenberg, M. S., & Plaut, D. C. (1998). Evaluating word-reading models at the item level: Matching the grain of theory and data. *Psychological Science*, *9,* 234–237.

Share, D. L. (1999). Phonological recoding and orthographic learning: A direct test of the self-teaching hypothesis. *Journal of Experimental Child Psychology*, *72,* 95–129.

Spinelli, D., de Luca, M., Filippo, G. D., 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.

Tabachnick, B. G., & Fidell, L. S. (2001). *Using multivariate statistics* (4th ed.). Boston, MA: Allyn & Bacon.

Tenjović, L., & Lalović, D. (2005). The effects of articulatory suppression on word recognition in Serbian. *Journal of Psycholinguistic Research*, *34,* 541–553.

Weekes, B. S. (1997). Differential effects of number of letters on word and nonword naming latency. *The Quarterly Journal of Experimental Psychology*, *50,* 439–456.

Yarkoni, T., Balota, D., & Yap, M. (2008). Moving beyond Coltheart's N. A new measure of orthographic similarity. *Psychonomic Bulletin & Review*, *15,* 971–979.

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.