Cognitive Psychology

(This is a sample cover image for this issue. The actual cover is not yet available at this time.)

Old and new ways to study characteristics of reading disability: The case of the nonword-reading deficit

Wim Van den Broeck [a,*], Astrid Geudens [b]

[a] Faculty of Psychology and Educational Sciences, Vrije Universiteit Brussel Pleinlaan 2, BE-1050 Brussels, Belgium
[b] Lessius University College (Catholic University Leuven), Jozef de Bomstraat 11 BE-2018 Antwerpen, Belgium

ARTICLE INFO

ABSTRACT

Theoretical and computational models of reading have traditionally been informed by specific characteristics of disabled readers. One of the most frequently studied marker effects of developmental dyslexia is the nonword-reading deficit. Disabled readers are generally believed to show a specific problem in reading nonwords. This study presents a survey of frequently cited methods used to examine this effect by controlling general reading ability in various ways. An extensive analysis, however, shows that the majority of these methods (grade equivalents scores, the reading-level match design, and interactions in a chronological-age match design) actually fail to account for confounding variables such as age and general slowing, potentially affecting the conclusions reached. To alleviate this problem, an alternative method is presented: i.e. state trace analysis. Applying this method in a sample of Dutch disabled and typical readers, the results revealed an absence of a nonword-reading deficit in the disabled readers. Furthermore, after controlling for their decoding ability, disabled readers showed inferior word reading performance, which strongly suggests that the fundamental problem of disabled readers does not relate to the reading of nonwords but concerns their (dis)ability to acquire orthographic (word-specific) knowledge. Further, predictions for disabled readers in an inconsistent orthography like English are formulated. Finally, based on a review of neurobiological studies, implications for theories of reading disability are discussed.

* Corresponding author. Fax: +32 0 2 629 2532.
  E-mail addresses: wivdbroe@vub.ac.be (W. Van den Broeck), Astrid.Geudens@lessius.eu (A. Geudens).

1. Introduction

One of the most frequently investigated marker effects of developmental dyslexia is the nonword-reading deficit (for reviews see Herrmann, Matyas, & Pratt, 2006; Rack, Snowling, & Olson, 1992; Van den Broeck, Geudens, & van den Bos, 2010; van IJzendoorn & Bus, 1994). Dyslexic readers would experience a specific problem in reading nonwords (these are wordlike nonsense words such as nup, trel, futpil). As nonwords have no lexical representation in long-term memory, and can only be read by transposing letters or letter clusters into phonemes or larger phonological units, many researchers interpret dyslexic readers' relative disability in reading nonwords as an unambiguous indication of a phonological disorder, in line with the currently dominant phonological deficit hypothesis of developmental dyslexia (cf. Liberman, 1992; Vellutino, Fletcher, Snowling, & Scanlon, 2004; Ziegler & Goswami, 2005). The reported nonword-reading deficit in dyslexic readers has been considered to be one of the core phenomena impacting on the construction of theoretical and computational models of reading (for dual route accounts of reading, see Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001; Perry, Ziegler, & Zorzi, 2007, 2010; and for single process connectionist accounts, see Harm & Seidenberg, 1999; Manis, Seidenberg, Doi, McBride-Chang, & Petersen, 1996; Plaut, McClelland, Seidenberg, & Patterson, 1996).

Evidently, basic effects that motivate theories of reading, such as the nonword-reading deficit, should be established unambiguously. This is a prerequisite in order to avoid the problem of "mutual confirmation ..(of).. method, data and theory" (Van Orden, Pennington, & Stone, 1990, p. 497). So far, it remains unclear, however, whether the reported nonword-reading deficit in disabled readers is indeed an unequivocal finding. In this paper, we search for unconfounded demonstrations of the nonword reading deficit and analyze frequently cited methods used to examine this effect by controlling general reading ability in various ways. Subsequently, we propose an alternative method, state trace analysis, to effectively control for confounding variables. In the following, we set out with a discussion of the most frequently adopted method in NRD studies, i.e. the reading-level match design (RLM). We will argue that the interpretation of NRD findings in these RLM studies is not without problems.

A key element in the demonstration of the nonword-reading deficit (NRD) is the finding that disabled readers show a specific problem in reading nonwords, implying that these difficulties exist over and beyond their general reading difficulties. Hence, general reading level is a key factor that should been taken into account while studying the NRD. To satisfy this condition, researchers have generally used the reading-level match design as a method to control for general reading level (Backman, Mamen, & Ferguson, 1984; Bryant & Goswami, 1986). In a RLM design, individuals with reading disabilities are matched with younger typical readers on a measure of reading ability with a real-word reading test. After this match on real-word reading, both groups are compared on their nonword-reading. The RLM design is usually preferred to the traditional chronological-age match design as many information processing differences between disabled readers and typical readers of the same age could as well be the result of both groups' different reading experiences (cf. Geudens, Sandra, & Van den Broeck, 2004). Given the experimental control on reading level, the RLM design is generally considered a more selective device in isolating critical processing differences between disabled and typical readers (cf. Stanovich & Siegel, 1994).

Despite the popularity of the RLM design, this design is, however, not without methodological problems and interpretational ambiguities (Goswami & Bryant, 1989; Jackson & Butterfield, 1989). The central problem is that the RLM design inevitably confounds reading level with age: the disabled readers necessarily being older than the control subjects. This observation has recently given rise to a new, developmental interpretation of nonword-reading deficit findings in the context of the RLM design as reported in our previous paper (Van den Broeck et al., 2010). In that paper, we provided empirical evidence for the hypothesis that the observed NRD in a RLM design does not reflect a genuine deficit, but is an artifact of normal developmental differences in word-specific knowledge between both reading groups. We hypothesized that the group of older disabled readers reached the same reading score in the word reading task (used to match both groups) as the younger typical readers as a result of their superior experience with the words to be read in this task, simply because they are older. Alternatively, we suggested that the younger typical readers needed to rely more on their decoding

ability in order to reach the same performance level as the older disabled readers on the word reading task. Following this reasoning, we concluded that a matching procedure with a RLM design would create an imbalance in decoding ability between both groups and hence could produce a spurious nonword-reading deficit. It is important to note that our interpretation of the nonword-reading deficit in terms of normal developmental processes does not entail that disabled readers' orthographic and semantic knowledge is at a normal level, however still is at a higher level than that of the *younger* typical readers. Clearly, this developmental interpretation raises doubts about the existence of an unambiguous NRD. Hence, it is essential to scrutinize the existing methods that have been used to study the NRD for their capacity to reveal a NRD without being compromised by confounding variables. Before discussing these methods, we consider the NRD findings in the broader context of the processes of phonological decoding and orthographic learning.

2. Phonological decoding and orthographic learning in disabled readers

The ability to recognize written words accurately and rapidly is generally considered as the hallmark of skilled automatic word reading (Perfetti, 1985; Share, 1995). This ability is based on the encoding of word-specific spelling patterns (unique sequences of letter strings) in long-term memory representations. Each successful match between the presented letter string and its corresponding lexical memory representation automatically activates the associated phonological, semantic and syntactic properties of the word (e.g. Coltheart, 1978; Forster, 1976; Morton, 1969; for a review see Carr & Pollatsek, 1985). Thus, regardless of variations in size, case, font and handwriting, this system of abstract orthographic representations, called "*the orthographic lexicon*", mediates lexical access (Grainger & Jacobs, 1996). However, contrary to fluent readers, disabled or dyslexic readers are characterized by either erratic or laborious reading. According to the self-teaching hypothesis (Jorm & Share, 1983; Share, 1995, 2008) orthographic learning,[1] the process through which orthographic representations are formed, consists of two independent processes. First, in the phase of phonological recoding (or simply "decoding") the reader assembles the phonological code of an unfamiliar written word without the help from an external teacher, based on his or her knowledge of grapheme–phoneme associations. Second, if this first step succeeds, the phonological code of the word can be mapped onto its orthographic counterpart, establishing word-specific knowledge of the spelling. Hence, for the acquisition of orthographic knowledge, phonological recoding is a necessary condition or a sine qua non (Share, 1995). The precise mechanism by which orthographic representations develop in typical reading as well as in disabled reading is a central issue in current reading research.

Without any doubt, Share's self-teaching model is the most dominant account of the developmental process toward fully specified orthographic representations (Share, 1995, 1999). The self-teaching hypothesis was supported in several studies with typical beginning readers using an experimental paradigm adapted from Reitsma (1983). In these studies, target words were presented either four or six times in a natural text (Share, 1999). These targets were novel letter strings (pseudowords) representing a fictitious place, animal or fruit. Every pseudoword (e.g., yait) had an alternative homophone spelling (e.g. yate). In each case only one spelling, the target spelling of the pseudowords, was presented to the child. The general outcome of studies based on this paradigm was that 3 days after independently reading the stories, target spellings were recognized more often, named faster and spelled more accurately than their alternate homophone spellings (Bowey & Muller, 2005; de Jong & Share, 2007; Share, 2004), hence indicating orthographic learning ability in young readers.

The reason why we bring up this discussion on orthographic learning is that it is crucial to our understanding of which components in disabled reading are relatively intact and which are deficient, which is precisely also a key issue in studies using a RLM design. Although research on orthographic learning in disabled readers is rare, particularly research that makes use of the self-teaching paradigm, the current studies seem to indicate that orthographic learning also in disabled readers is indeed

---

[1] The construct of orthographic learning refers to two related, yet different aspects: first, the acquisition of word-specific representations, and second, getting knowledge about general aspects of the writing system such as sequential dependencies, structural redundancies, sensitivity to bigram frequencies, etc. (Castles & Nation, 2006; Geudens & Sandra, 2002; Vellutino, Scanlon, & Tanzman, 1994).

possible. Importantly, disabled readers would need more exposures in building up orthographic knowledge than typical readers (Ehri & Saltmarsh, 1995; Reitsma, 1983; Share & Shalev, 2004). The fact that these disabled readers show poorer orthographic learning than typical readers can be explained by the idea of the self-teaching account, i.e. the degree of orthographic learning depends on the overall level of decoding success (Share, 1999; Share & Shalev, 2004). Now, as we will argue, Share's idea of orthographic learning commensurate with target decoding success, is in fact in contrast with the standard conclusion in most RLM-studies that disabled readers have a specific problem in reading nonwords, i.e. a problem over and beyond their word reading ability. How should we interpret this discrepancy? Characteristic of the RLM design is that disabled readers and typical readers are matched on real-word reading level. To explain why disabled readers attain the same reading level as controls in such design, in spite of their weak phonological processing, an *orthographic compensation* process has been put forward as a necessary corollary (Rack et al., 1992). This orthographic compensation process, as an alternative to Share's commensurability thesis, can be interpreted in various ways (cf. Stanovich & Siegel, 1994). First, it could be thought of as a consequence of orthographic processing ability being less impaired in disabled readers. Second, it could be interpreted as a strategic preference attempting to compensate for a so-called phonological deficit. Third, a final interpretation could be that disabled readers reach the same word-reading level as their younger nondisabled controls, because they have had more exposure to written language or print (Stanovich & Cunningham, 1993; Stanovich & West, 1989). Importantly, although this third interpretation of the orthographic compensation hypothesis bears some resemblance with our developmental account as formulated in Van den Broeck et al. (2010) also stressing the importance of word reading experiences, there is a clear difference. In line with the above account of the orthographic compensation hypothesis in terms of differences in print exposure, disabled readers' familiarity with the words in the matching task compensates for the lower decoding ability in reading real words. Nevertheless, the poor performance on the nonword-reading test indicates a genuine decoding deficit. Our developmental account, on the contrary, implies that the matching procedure automatically selects a group of younger typical readers who have better decoding abilities, not by nature but by design.

As Share and Shalev (2004) argue, the self-teaching paradigm offers a strong test of the orthographic compensation hypothesis by examining the on-line acquisition of orthographic information when print exposure is held constant. Presenting an identical number of target exposures for both typical and disabled readers, disabled readers showed impaired orthographic learning proportionate to levels of target decoding success. This finding was observed both in the post-test spelling and orthographic choice measurements and falsifies the orthographic compensation hypothesis whether it is seen as an intrinsic processing ability (argument 1), a strategic preference (argument 2) or a result of print experience (argument 3). On the contrary, Share and Shalev's demonstration that disabled readers do not show evidence of orthographic compensation is consistent with our developmental interpretation of nonword-reading deficit findings in a RLM-design as put forward in Van den Broeck et al. (2010), where we showed that the differences found in such RLM-studies may actually reflect a normal developmental trend: with increasing age, word reading is determined by developmentally higher levels of word-specific knowledge. Hence, apart from the methodological problems already discussed, the self-teaching paradigm raises further questions about the unequivocal interpretation of the NRD findings in RLM studies.

3. Varieties of reading disability

Before we continue our search for an *un*confounded nonword reading deficit as a general characteristic of reading disability, and analyze other frequently adopted *methods* in NRD studies, there is one argument that needs our further attention. A possible counterargument could be that a general characteristic of disabled reading does not exist. Castles and Coltheart (1993, 1996; see also Jackson & Coltheart, 2001) indeed argued that the group of disabled readers may not be a homogenous group but instead reveal remarkably different patterns of impairment. In their studies, some children showed a 'phonological subtype' of impairment with poor nonword reading relative to the reading of exception words (55% of the poor readers), others showed a 'surface subtype' of impairment with

poor exception word reading relative to nonword reading (30% of the poor readers). Nevertheless, these findings were criticized on methodological grounds (Snowling, Bryant, & Hulme, 1996; Stanovich, Siegel, & Gottardo, 1997; Ziegler & Goswami, 2005): comparing disabled readers to typical readers of the same age, as in Castles and Coltheart (1993), may yield processing trade-offs between sublexical and lexical paths of word recognition that depend on the overall level of word reading skill. To overcome this comment, a new series of subtyping studies was published, adopting a reading-level match design (Manis et al., 1996; Murphy & Pollatsek, 1994; Stanovich et al., 1997). Essentially, these studies demonstrated that once general reading level was taken into account, 20–25% of the children with reading disability could be identified as phonological dyslexics, whereas the surface subtype almost disappeared (1% or 2%). The rest of the children showed a mixed profile. The prevailing interpretation of these results is that surface dyslexia would arise from a milder form of the phonological deficit leading to a delayed pattern of reading development, whereas phonological dyslexia would represent a more severe reading disorder (Griffithts & Snowling, 2002; Stanovich et al., 1997; but see Jackson and Coltheart (2001) for an alternative view).

Two conclusions follow from this argumentation. First, although there is an unquestionable awareness in the literature that the expression of reading disability is a diverse matter, the hypothesis of a phonological deficit as a common underlying disorder together with the prediction of a NRD remains largely uncontested. Second, because the second wave of subtyping studies made use of the RLM methodology, confounding reading level with age, the extremely skewed proportions of phonological dyslexics vs. surface dyslexics could possibly be the consequence of a confound with age. Indeed, if the normal development of reading irregular words differs (is faster) from reading nonwords, then – as argued before – lower nonword reading performance may simply be the result of normal developmental trends instead of revealing a genuine subtype. As the literature shows, it is far from clear how many subtypes can be identified and what the precise proportion of each subtype would be. For these reasons, we believe that a sound research strategy to study the NRD would be to first establish whether a NRD can be unambiguously demonstrated in the entire group of disabled readers. Regardless of the outcome, the subtyping issue deserves further attention using methods that avoid the confounding factors just mentioned (see Van den Broeck, Geudens, & Staels, 2011). The present publication focuses only on this first step: the study of nonword reading ability in the entire group of disabled readers.

4. Methods to examine a nonword reading deficit

This part of the article provides a conceptual discussion of the existing methods that are adopted in the study of the NRD and analyzes the way in which they control for general reading level (technical details are relegated to the appendices). The analysis reveals that all four current methods (grade equivalent scores, the RLM design, standard scores, and interaction effects in a chronological-age match design) are either uninformative with respect to the study of a processing deficit (standard scores) or result in biased interpretations (the other three methods). After discussing the four methods, we will introduce a new way of approaching the NRD question by adopting state trace analysis. Central to this method is that it avoids potential confounding variables. Based on state trace analysis, we present data of a Dutch sample of disabled and typical readers and provide support for the hypothesis that the disabled readers do not exhibit a nonword reading deficit (see Section 5).

4.1. Age- or grade-equivalent scores

The first method builds on age- or grade-equivalent scores in order to establish whether disabled readers have a problem in reading nonwords over and beyond their general reading difficulties. Although grade-equivalent scores are very popular measures in school practice to evaluate the performance of a child on an academic task, they are hardly used in reading disability research. Nevertheless, it is useful to discuss the principles of this method, because, as we will demonstrate in one of the next sections on the RLM design, the procedure to use grade-equivalent scores is mathematically equivalent to the RLM procedure. Grade-equivalent scores express a child's performance in terms of the grade or age level in the norming sample at which the average score is the same as the child's score.

For example, when a 9-year-old child has a raw score of 30 on a reading test and 30 is the mean raw score of 7-year-old children, then this child has a reading age of 7 years, showing a 2-year delay. An important reason for why grade-equivalent scores are an attractive measure in a school context is the perceived ease to compare scores between children, tasks and moments in terms of an easily understood concept, i.e. delay or lead in number of years (and months) of academic instruction. To illustrate the use of grade-equivalent scores in the study of the NRD in Dutch children, Table 1 shows the mean raw scores on a word reading test and a nonword reading test of disabled readers and of a norming sample (taken from Braams (2002)).

As can be seen in Table 1, reading disabled children from grade 6 read words (58.1) at a level that is average for children between grade 2 (48.3) and grade 4 (66.3). Thus, they show a word reading delay of approximately three grades. This delay represents the baseline in this method to which the delay in reading nonwords can be referenced. Inspection of the nonword reading score of these disabled readers (36.8) makes clear that they have a nonword reading delay of at least four grades: their score is even lower than for average grade 2 children (39.2). Apparently, the nonword reading delay of the disabled readers is larger than their word reading delay, which seems to be a clear demonstration of their nonword-reading deficit.

However, it is well known from the psychometric literature that the method of age- or grade-equivalent scores is faced with difficulties (cf. Anastasi, 1968; Bracken, 1988; Gulliksen, 1950; Reynolds, 1981; Thurstone, 1926). The reported problems are exactly the same as those surrounding the concept and measurement of mental age (cf. Stanley, 1964) and motivated, some 50 years ago, the replacement of a ratio IQ (mental age divided by chronological age) by the currently used deviation IQ which is essentially a transformed z-score or standard score. A standard score reflects the relative position of an individual on a given task, compared with individuals of the same age. The key issue is that in order to validly compare age- or grade-equivalent scores between tasks, it is required that the growth curves of both tasks' means are parallel and that the standard deviations of both tasks are the same for each point in time (Evers & Resing, 2007). Since both requirements are not easily fulfilled, the use of age- or grade-equivalent scores could lead to a seriously biased assessment of a child's relative proficiency in each task.

To exemplify this bias, we present two fictitious situations in Fig. 1a and b in which the relative position of a child on a word reading test and a nonword reading test is exactly the same, viz. two standard deviations under the mean ($z = -2$). In the situation depicted in Fig. 1a, the standard deviation of the nonword reading test is twice as large as the standard deviation of the word reading test. As can be seen, despite the child's identical standing expressed in standard scores on both tasks, the grade-equivalent scores erroneously suggest that this child's ability to read nonwords is more impaired than his/her word reading ability. In Fig. 1b, the same effect is observed when the slope of the growth curve for the mean scores of the nonword reading test is less steep than that of the word reading test (keeping standard deviations constant). In both situations, the bias results from the fact that grade-equivalent scores do not take into account the specific development of both tasks' distributions. As illustrated in Fig. 1a, the larger delay on the nonword reading task is not typical for the disabled readers, because the average readers show an equal nonword-reading deficit when they are compared with the excellent readers (reading two deviations above the mean; $z = 2$). In reality, many

Table 1
Mean raw scores on a word reading test and a nonword reading test.

	Disabled readers		Typical readers (reference group)	
	Word reading	Nonword reading	Word reading	Nonword reading
Grade 2	21.7 (−1.9)	14.0 (−1.4)	48.3	39.2
Grade 4	43.3 (−1.7)	29.2 (−1.4)	66.3	56.3
Grade 6	58.1 (−1.8)	36.8 (−1.9)	78.3	72.1

Note: The numbers in parentheses are *z*-scores. From "De zin van onzinwoorden: het gebruik van pseudowoorden bij de signalering, de diagnostiek en de behandeling van dyslexie. (The meaning of nonsense words: the use of pseudowords in screening, diagnosis and treatment of dyslexia)" by Braams (2002), Tijdschrift voor Remedial Teaching, 10 (2), p. 6. Copyright 2002 by LBRT. Adapted with permission.
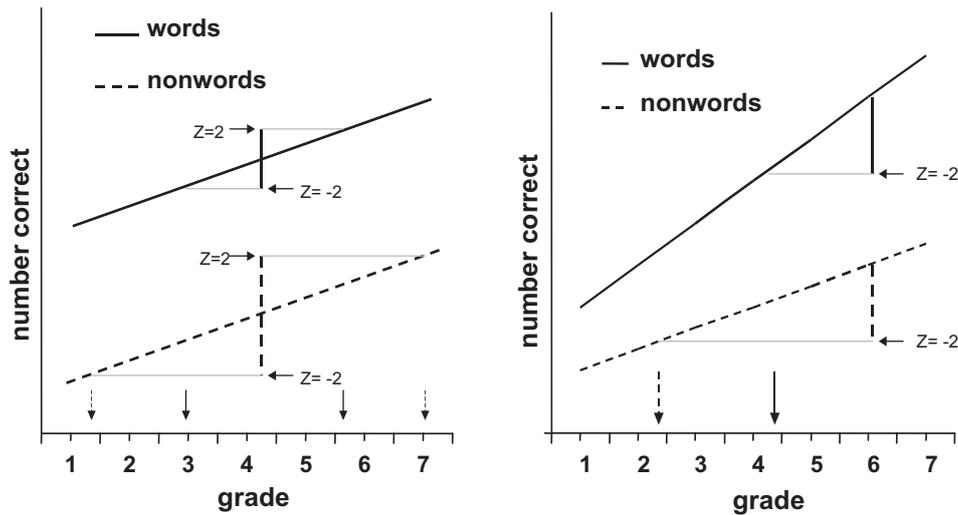
Fig. 1. Grade-equivalent scores of disabled readers on a word- and a nonword reading test. Slopes are equal and standard deviations are not equal in (a), and slopes are not equal and standard deviations are equal in (b). Arrows pointing to the *X*-axis indicate grade-equivalent scores (full arrows for word reading scores and dotted arrows for nonword reading scores).

nonword reading tests show a somewhat slower developmental curve and also have a systematically larger standard deviation, compared with word reading tests (Section 4.3). Therefore, larger nonword reading delays are frequently observed when employing grade-equivalent scores.

4.2. Standard scores

A second method to establish a baseline of general word reading ability as a reference point for nonword reading performance is the well-kwown method of comparing standard scores (usually *z*-scores). By first centering the raw score (by subtraction of the mean) and then dividing this result by the standard deviation, the *z*-score expresses the performance of an individual in terms of the number of standard deviations the raw score is distanced from the mean. The advantage of this measure of relative position is that it effectively deals with the problems inherent in grade-equivalent scores (unequal standard deviations and unequal differences between means). This method also enables a valid comparison between reading groups of the same age. When standard scores (*z*-scores) for the data in Table 1 are calculated, it is obvious (see Table 1) that the disabled readers are not weaker in reading nonwords than in reading words. On the contrary, disabled readers of grades 2 and 4 are somewhat better in reading nonwords than in reading words. This nonword reading advantage of disabled readers is demonstrated most clearly when we compare the *z*-scores of the 10% poorest readers on a word

Table 2
z-Scores of disabled readers (<Pct 10) on a word- and a nonword reading test.

	Word reading test	Nonword reading test
Grade1 (*n* = 29)	−1.18	−.96
Grade 2 (*n* = 31)	−1.40	−1.22
Grade 3 (*n* = 16)	−1.92	−1.45
Grade 4 (*n* = 28)	−1.69	−1.34
Grade 5 (*n* = 25)	−1.75	−1.55
Grade 6 (*n* = 27)	−1.64	−1.29
Grade 7 (*n* = 28)	−1.76	−1.20
Secondary (*n* = 19)	−1.83	−1.60
Adults (*n* = 8)	−2.08	−1.62

reading task for each grade level with the *z*-scores on a nonword reading task (see Table 2), taken from a large-scale norm-study presented in Van den Broeck et al. (2010). For each grade level, the *z*-scores on the nonword reading task are less extreme than the *z*-scores on the word reading task. Although this outcome may sound a little surprising at first sight, at least in the light of the NRD research, this clear and systematic nonword reading advantage of the disabled readers is merely an illustration of the ubiquitous regression effect. When a group is selected at the lowest end of a specific variable (word reading), than the mean standard score of this group on any other variable that holds an imperfect correlation with the selection variable, will regress toward the mean (cf. Rogosa & Willett, 1985). It is important to note that this regression effect and the lack of a perfect correlation between the two tasks are both manifestations of the empirical fact that the nonword reading performance is determined by somewhat different factors than those on which the children were selected in the first place. For example, if semantic knowledge would indeed play a role in reading words (see Nation & Cocksey, 2009), the poor readers are partly selected on their semantic knowledge. This factor could, however, not play any selective role in reading nonwords.

Although the nonword reading advantage of disabled readers in terms of standard scores is, as far as we know, never mentioned in the research literature, it should remind reading researchers that defining reading disability (or developmental dyslexia) as poor word reading is a decision that largely determines the relative weaknesses of disabled readers in terms of their relative position in the entire population of readers.

Nonetheless, this nonword reading advantage is a foregone conclusion once reading disability is defined as poor reading, and hence does not add any new information about the reading process. Moreover, a *z*-score is a dimensionless quantity and for that reason does not directly refer to a real process. What the reading researcher really wants to know is whether and why nonwords put an extra burden on the processing load of disabled readers, e.g. in terms of longer reaction times or an enhanced error rate. Importantly, this regression effect of relative standing and ensuing nonword reading advantage is not incompatible with a nonword reading deficit in terms of an absolute measure of the reading process. For instance, the decline in reading times from words to nonwords can be much larger for disabled readers than for typical readers, even when the relative position of the disabled readers is better for nonwords than for words.

4.3. The reading-level match design

In Section 1, we explained that the frequently used RLM design confounds reading level with age, and therefore enables interpretations of nonword-reading deficit findings in terms of normal developmental differences between typical and disabled reading groups. In this section, we further delve into the precise conditions for obtaining a nonword-reading deficit using a RLM design. We demonstrate that the RLM method is mathematically equivalent to the procedure of using grade-equivalent scores. Implications for reading research are subsequently discussed.

We start with the observation that each RLM-study operationalizes reading disability, sometimes aside from other criteria (e.g. average intelligence), at least on the basis of a predetermined level of reading ability that should not be exceeded (e.g. below percentile 10 on a word reading test). This cut-off procedure as such dichotomizes a variable that is continuous in reality. As Van den Broeck et al. (2010, p. 721) have clarified, the statistically most preferable RLM design is one in which the matching variable is the same as the selection variable, because it avoids regression artifacts. Conceptually, it is also the most logical choice to use the same reading task as an operationalization of reading ability in selecting as well as in matching the participants. Based on a prediction equation, it is possible to determine exactly in which circumstances a nonword-reading deficit will emerge and those in which it will not. Further in this section, we extend our analysis to the situation in which the selection variable is different from the matching variable.

When the selection variable is the same as the matching variable (a word reading test), and the bivariate distribution of word reading and nonword reading in the population is known, the mean nonword reading score of the disabled readers and the younger typical readers can be calculated exactly from their mean word reading score (which is the same for both groups). In practice, this bivariate distribution has to be estimated from the sample means, standard deviations and correlations of
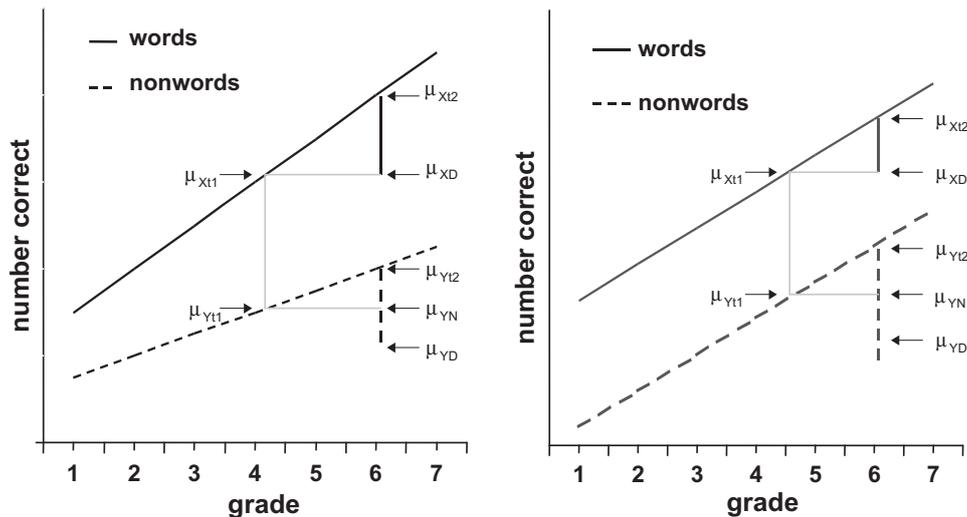
Fig. 2. Predicted nonword-reading scores of disabled and reading-level matched normal readers in case the mean word-score of the matched control group (μ_{XD}) equals the mean word-score of that age group (μ_{Xt1}). Additionally for (a) the standard deviations of both tests are equal at the age of the disabled readers, and the correlation between both tests at that age equals one (hence, the absence of a regression effect). For (b) the slopes are equal, and the ratio of the standard deviation of the word reading test and the standard deviation of the nonword-reading test at the age of the disabled readers ($1/\lambda$) is smaller than the correlation of both tests at that age.

both variables for the ages of the two groups. Let us further assume that the two variables are binormally distributed.[2] Now we can predict precisely when a nonword reading deficit will emerge. In Appendix A, the general prediction equation is derived and can be applied to all kinds of situations. Here we illustrate this analysis for two specific situations.

First, when the mean word-score of the matched control group exactly equals the mean word-score of that age group (see Fig. 2a) and the standard deviations of the word reading task and nonword reading task at the age of the disabled readers are equal, then a nonword reading deficit depends entirely on the slope of both tasks and on the correlation between them at the age of the disabled readers. In Fig. 2a, this situation is visualized showing that the slope of the word reading variable has to be steeper than the slope of the nonword reading variable in order to go beyond the regression effect (due to a lack of a perfect correlation; Fig. 2a assumes a perfect correlation), thus to obtain a nonword-reading deficit. In contrast, when the slopes are equal, then a nonword-reading deficit depends completely on the ratio of the standard deviations of both tasks and their correlations. In this situation as depicted in Fig. 2b a nonword-reading deficit will only emerge when the ratio of the standard deviation of the word reading test and the standard deviation of the nonword reading test ($1/\lambda$) at the age of the disabled readers is smaller than the correlation coefficient at that age (see Appendix A for a mathematical proof).

From this analysis, three conclusions follow. First, when reading disability or developmental dyslexia is operationally defined as the cut-off score on a continuous variable, a nonword-reading deficit is entirely determined by characteristics of the total continuous population and an arbitrary cut-off score. Hence, it cannot be regarded as a specific characteristic of disabled readers. Second, when the selection variable and the matching variable are one and the same, and reading disability is exclusively operationalized by performance on this variable, the reading disabled group and the matched typical reading group can only differ from each other in age. Hence, any difference observed between the two groups on the criterion variable should be attributed to factors related to the age-difference. In other words, the question that can be answered in a RLM-design study is one that relates to why and how the empirical distributions of the word- and nonword reading tests change with age. The

---

[2] It follows from this that the regression of X on Y (and of Y on X) is linear and homoscedastic (cf. Stuart & Ord, 1994). If this condition would not be met, the prediction equations would become more complex, but our substantive conclusions would be unaffected.
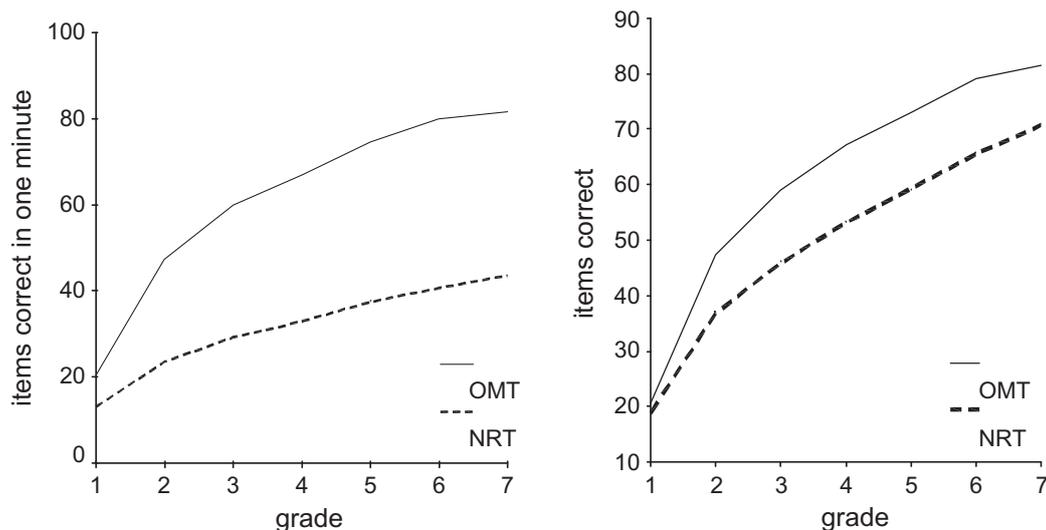
Fig. 3. (a) Development of mean OMT- and Klepel-scores when both are 1-min tests. The diminishing slopes with grade are the consequence of the increasing difficulty of the items presented. (b) Development of mean OMT- and Klepel-scores when OMT is a 1-min test and Klepel is a 2-min test.

interpretation of the nonword-reading deficit in terms of changing word-specific knowledge, put forward by Van den Broeck et al. (2010), is an attempt to capture this point. Third, from this hypothesis in terms of developing word-specific knowledge, it follows directly that word reading is more susceptible to the influence of age-related word-specific knowledge than is nonword reading. Hence, the growth curve of the word reading test will be somewhat steeper than the growth curve of the nonword reading test.

Empirical evidence for this prediction was found in the norm-study by Van den Bos and lutje Spelberg (1994). In this study the nonwords of the nonword reading test (Klepel) were matched with the words of the word reading test (One Minute Test) on pronounceability, word length and word structure, enabling a valid comparison between both tests' growth curves of the means. The only exception was that the Klepel is a 2-min test. To this end – to adequately compare the development of word reading and nonword reading skills – the authors also administered nonword reading scores on the Klepel after 1 min of reading. Fig. 3a shows that the development of nonword reading is substantially slower than the development of word reading. When, however, the Klepel-scores (after 2 min of reading) are compared with the One Minute Test scores, the two developmental curves are almost parallel (see Fig. 3b). Importantly, this artificial increase in the slope of the nonword reading scores is accompanied by a substantial increase in the standard deviations of these scores, so that the standard deviation of the Klepel-scores is larger than the standard deviation of the OMT-scores for each grade-level. To summarize, the hypothesized differential word-specific knowledge of groups differing in age creates the two formal conditions to observe a nonword-reading deficit: unequal slopes and/or unequal standard deviations.

At this point of our discussion, the similarity between the RLM design and the method of comparing age- or grade-equivalent scores to measure a nonword-reading deficit, as discussed before, is becoming clear: different slopes and different standard deviations are those conditions to detect a spurious NRD. Thus, it should not come as a surprise that both methods prove to be mathematically equivalent. Appendix B shows that the size of the delay in nonword reading can be expressed as a function of the magnitude of the NRD in a RLM-design. This result has far-reaching implications. It means that the well-known psychometric problems associated with the use of grade- or age-equivalent scores are also inherent in the utilization of a RLM-design as a means to detect and measure a NRD. Obviously, both methods do not seem appropriate to examine whether disabled readers have a specific problem in reading nonwords, because neither of the methods take into account the normal developmental processes that evolve in time. In other words, both methods make inappropriate comparisons. If one is interested in the relative proficiency of disabled readers in reading nonwords, one should compare disabled readers and typical readers of the same age (see below).
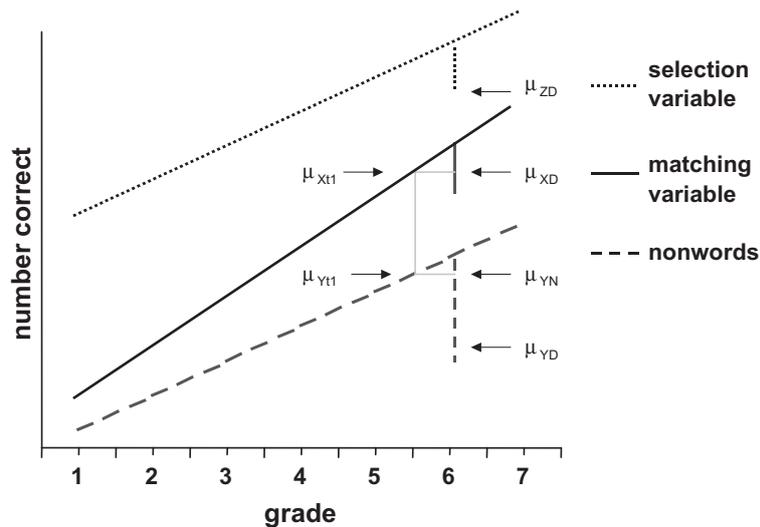
Fig. 4. Illustration of the predicted nonword-reading deficit when the selection variable (Z) correlates higher with nonword reading (Y) than the matching variable (X) does. μ_{ZD} is the mean score of the disabled readers on the selection task.

Finally, we will consider what happens with the prediction of the nonword-reading deficit when the selection variable does not equal the matching variable. When the selection variable is known, it takes a multiple regression equation with the selection variable and the matching variable as independent variables to exactly compute the mean nonword reading score of the disabled and the non-disabled readers. However, in many RLM studies the selection variable is not known since the selection of the disabled group is based on clinical criteria. In that case, the two groups can differ on a multitude of variables and their nonword reading scores will not exclusively depend on their different ages. It is precisely this most problematic design that is employed frequently in RLM studies. However, it is reasonable to assume that the clinical selection criteria are closely related to the reading of books and texts appropriate for the children's age. When these clinically selected children are subsequently matched with younger typical readers on a difficult word reading test requiring a high level of word-specific knowledge, the correlation of the matching task with the nonword reading task is probably lower than the correlation of the selection variable(s) with nonword reading. In this case, which is common in English language studies, a large nonword-reading deficit is expected (see Fig. 4). Fig. 4 shows that the predicted mean score (μ_{XD}) of the disabled readers on the matching variable (X), based on their score (μ_{ZD}) on the selection variable (Z), regresses considerably toward the mean, which results in a matched control group that is situated at a relatively high level on the selection variable and on the criterion variable (Y) as well. The nonword reading scores of the disabled readers (μ_{YD}), which are strongly correlated with the scores on the selection variable, do regress only slightly from their extreme position on the selection variable. As can be seen from Fig. 4, this procedure results in a large nonword-reading deficit.

4.4. Interaction effects in a chronological-age match design

A fourth classic method to investigate a processing deficit is the demonstration of an interaction effect in a chronological-age match design (see Martens & de Jong, 2006; van der Leij & van Daal, 1999; Yapp & van der Leij, 1993). To illustrate this method, we plotted the word reading and nonword reading data of Table 1 for grade six readers (see Fig. 5), and observed a clear interaction effect, indicating that the decline from the reading of words to the reading of nonwords is much larger for disabled readers than for typical readers. The decline of the typical readers is taken here as the baseline.

This method, based on the interaction effect, finds its roots in linear theory which presumes that the dependent variable (reading performance) can be expressed as the sum of some main effects (lexicality of the items and reading group) and interactions among these independent variables (lexicality by group). Although linear theory has been proven to be a reliable model in many experimental
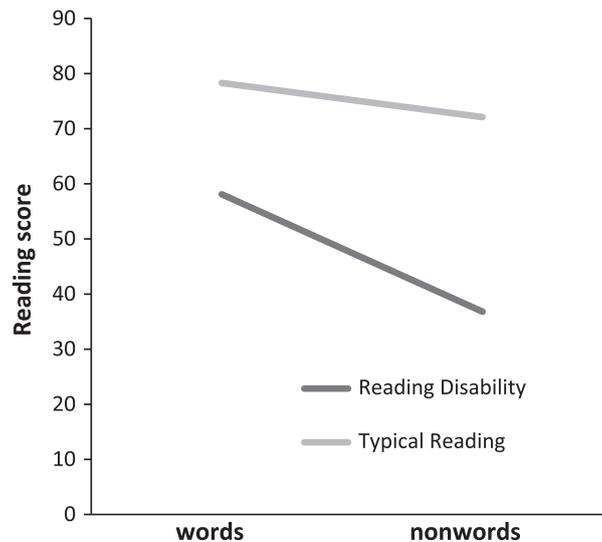
Fig. 5. Interaction effect between item lexicality and reading group.

studies, it has been considered to be too restrictive in others, a comment that has been linked to the observation that researchers are usually unaware of the precise form of the relationship between the observed measures and the underlying constructs (Dunn & James, 2003). Therefore, is has been argued that relying on the presence of a statistical interaction as evidence for a qualitative group-related difference is not without problems (Loftus, 1985; Loftus, Oberg, & Dillon, 2004; Loftus, Truax, & Nelson, 1987). Even nonordinal interaction effects can be made to disappear or reverse by applying a suitable monotonic nonlinear transformation to the dependent variable (Bogartz, 1976; Loftus, 1978). This scale-dependency problem is still exacerbated in research where nonexperimental variables such as age or pathology are involved because in such situations it is likely that an unspecific general factor influences performance (Kliegl, Mayr, & Krampe, 1994).

We will illustrate this problem by considering the following conditional cases. Assume a general slowing problem in disabled readers. In this case, their reading times can be expressed as a fixed ratio of the reading times of typical readers, once peripheral components (i.e. sensory and motor processes) have been subtracted out.[3] Next, assume a word reading task in which typical readers need 700 ms on average to read each word, and peripheral processes take 200 ms. Suppose the group-related slowing factor is 2.0. Then, the expected average reaction time for disabled readers is $200 + 2.0 \times 500 = 1200$ ms. (Note that no slowing of the peripheral component is supposed). Now introduce a nonword reading task, increasing the average reaction time of the typical readers by 500 ms, resulting in a total time of 1200 ms. We then expect an average reaction time for the disabled readers of $200 + 2.0 \times 1000 = 2200$ ms.

What this illustration shows is that we can observe a group difference of 500 ms. in the word reading condition and of 1000 ms. in the nonword reading condition, even though no other deficit is associated with the specific processes underlying nonword reading. Thus, performing an ANOVA including an interaction term would lead to a spurious interaction effect. Moreover, this scale-dependency problem also arises when floor or ceiling effects, leading to heteroscedasticity, are present in the data (Loftus, 1985). As is generally acknowledged, reaction times are notoriously susceptible for these problems. For example, a measure as used in Table 1 and Fig. 5, i.e. the number of words read correctly within a given time period, can also lead to floor or ceiling effects (depending on the difficulty range of the items). Apart from the possibility of general slowing, the interaction effect in Fig. 5 could be the result of a ceiling effect in the word reading condition for the typical readers.

To solve this problem of misinterpreting overadditive interaction effects, other methods have been proposed in the literature and some of them have also been applied in reading research. Common to all these methods is the assumption of a specific relationship between the performance of the target

---

[3] This illustration is adapted from Verhaeghen and Cerella (2002) who give a similar illustration for aging research.

group and the performance of the control group. For instance, with the aim of neutralizing the effect of a general slowing factor, one could transform raw reaction times into logarithms or ratios (Salthouse & Hedden, 2002, for an application in reading research, see Martens & de Jong, 2006). However, the implied multiplicative relationship between the performance (including input and output processes) of groups and between experimental conditions is an assumption that cannot always be substantiated (cf. Verhaeghen & Cerella, 2002). An alternative solution proposed by Faust, Balota, Spieler, and Ferraro (1999) with the purpose of removing the influence of individual differences in overall performance level, is to use a *z*-score transformation obtained by taking each individual's condition means, subtracting their overall mean, and dividing these by the standard deviation of their condition means (for applications in reading research, see Di Filippo, de Luca, Judica, Spinelli, & Zoccolotti, 2006; Paizi, De Luca, Zoccolotti, & Burani, in press; Zoccolotti, De Luca, Di Filippo, Judica, & Martelli, 2009; Zoccolotti, De Luca, Judica, & Spinelli, 2008). Again, linearity of group and experimental effects (i.e. additive or multiplicative) is assumed (De Brauwer, Verguts, & Fias, 2006). While these methods certainly lead to a substantial improvement, we believe they still impose unnecessary restrictions that can be relaxed using the more general framework of state trace analysis.

5. State trace analysis

5.1. Rationale of state trace analysis

So far, we discussed four methods that have frequently been adopted in reading research aiming to detect a specific nonword reading deficit in disabled readers by controlling for general reading level. Our analysis shows that the general application of these methods entails methodological and interpretational problems that jeopardize unequivocal conclusions. Even though ameliorations of some of these methods could be proposed, we opt for a more fundamental solution that avoids any potential confound. Before explaining the general outline of this method, i.e. state trace analysis, we illustrate its rationale by means of a concrete example applied to the nonword reading deficit. The key issue is how to match typical and disabled readers simultaneously on their reading level and their age. The only way to satisfy this condition is to operationalize reading performance at the item level instead of the participants' level. Suppose, for instance, that a disabled reader needs one second to read the word 'bag' correctly. Then, we can look for a word that takes a typical reader of the same age also exactly one second to read, e.g. 'important'. We term this procedure a functional matching. Next, we construct a nonword for each of both words so that the same structural features (number of letters and syllables, syllable structure and consonant structure) are represented in both words and nonwords, e.g. 'nat' (structurally matched with 'bag') and 'umpalters' (structurally matched with 'important'). The crucial question now is whether it takes the disabled reader more time to read 'nat' than it takes the typical reader to read 'umpalters'. Because the nonword pair 'nat–umpalters' is structurally equivalent with the word pair 'bag–important', any difference in reading performance between disabled readers' reading of 'nat' items and typical controls' reading of 'umpalters' items may represent a direct and unambiguous indication of a potential nonword reading deficit.

However, it is essential for this method to be valid that group differences in reading performance between 'nat' and 'umpalters' items are not confounded with any other psycholinguistic characteristics of the items, such as letter frequency, bigram frequency, orthographic neighborhood or any other feature. This means that any difference in difficulty between 'bag' and 'important' items should be the same as for 'nat–umpalters' items. To fulfill this condition, the word–nonword pairs should be additionally matched on one or more of these item features (or checking that the differences in difficulty are in fact the same). After appropriately applying this method, any group difference can unequivocally be attributed to the lexicality factor. Thus, this design represents a truly empirical issue and the outcome is by no means a foregone conclusion.

Applying this method, three alternative hypotheses could be formulated. The first hypothesis proposes that disabled readers have a specific nonword reading problem. If this is indeed true, we expect them to take more time to read 'their' nonwords compared to the typical readers. Second, an alternative hypothesis does not assume a specific nonword reading problem for disabled readers but in

contrast proposes a smaller difference between word and nonword reading for disabled readers than for typical readers, based on their varying word-specific knowledge base. Following this hypothesis, the difference between words and nonwords would be relatively insignificant ('bag' and 'nat') for disabled readers as their limited word-specific knowledge would force them to approach both kinds of items with a decoding strategy. We name this hypothesis the 'decoding stagnation' hypothesis after the suggestion first formulated by Frost, Katz, and Bentin (1987, p. 104) that inexperienced readers rely more on phonological codes. For typical readers on the other hand, the difference between words and nonwords would be more substantial ('important' vs. 'umpalters') because of their larger word-specific knowledge. Hence, if typical readers, who need one second to read a word, are able to exploit more word-specific knowledge than the disabled readers do, their reading of structurally matched nonwords would be considerably poorer. Consequently, one would expect a nonword reading advantage in disabled readers as they are relatively unhampered by the difference between both kinds of items. The third and final hypothesis advances the idea that reading processes at word level could be qualitatively quite similar for both groups. In that case, no differences between both groups are expected in their nonword reading times.

In the general framework of state trace analysis, the idea of functional matching is extended by systematically varying reading time through the manipulation of item difficulty. State trace analysis was developed originally by Bamber (1979) and has more recently become an important instrument in detecting underlying causality in cognitive research through the seminal work of Geoffrey Loftus (Loftus et al., 2004, see also Dunn & James, 2003; Newell & Dunn, 2008; Prince, Brown, & Heathcote, 2012). State trace analysis has been shown to be a successful method in memory research (Busey, Tunnicliff, Loftus, & Loftus, 2000), in perception research (Loftus & Irwin, 1998), in aging research (Verhaeghen & Cerella, 2002; Verhaeghen, Steitz, Sliwinski, & Cerella, 2003) and in developmental research (De Brauwer et al., 2006). The fundamental question state trace analysis seeks to answer is whether the relationship between a set of causal variables and a set of indicator variables is mediated by one or more latent variables (Prince et al., 2012). Pertaining to the issue of the NRD, the central question is whether the relationship between a set of indicator variables (speed and accuracy of reading words and nonwords) and a set of causal variables (lexicality of items and reading group) is determined by the same underlying latent processing mechanism or by different mechanisms (see Fig. 6).

Prince et al. (2012) describe the manipulated factor corresponding to the indicator variables as the state factor. In our study, this state factor is lexicality. The factor used to influence the mediating latent process is termed the dimension factor. Whereas in many studies this dimension factor is another manipulated experimental variable, we include a natural variable (typical vs. disabled reading group) in this study in order to differentially impact on the latent process. This is an important distinction. Generally the dimension factor represents a direct manipulation of the underlying configuration of latent processes. However, in our study, the role of the dimension factor (typical vs. disabled reading) constitutes the central empirical question. In other words, our use of state trace analysis as an equivalence technique (matching disabled and typical readers on word reading performance) is not intended to resolve the question whether words and nonwords are processed using different reading mechanisms in typical readers. Instead, our aim is to discover whether the reading processes that mediate between the causal variables and the indicator variables in Fig. 6, whatever their nature is,
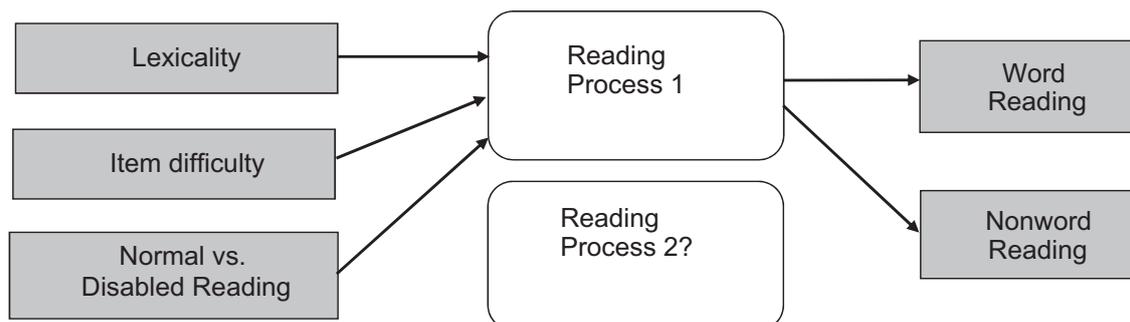

Fig. 6. Latent dimensionality underlying the relationship between a set of causal variables and a set of indicator variables.

are equal or different in the two reading groups. This implies that no conclusions will be drawn about the dispute in the literature concerning dual route vs. single route theories of word recognition.

The third causal variable (item difficulty) is included in order to induce variation in the overall level of performance (the trace factor). The aforementioned problem of scale-dependency (see 2.4) is circumvented in state trace analysis by adopting a nonparametric approach. The only requirement in this approach is the broad assumption that latent variables have a monotonic effect on the indicator variables. Hence, when proficiency at the underlying reading process increases, performance on word and nonword reading should also increase. The specific form this monotonically increasing (or decreasing) relationship takes for each indicator is irrelevant. From this assumption, a crucial prediction can be derived (Bamber, 1979): if only one latent variable is required to explain the relationship between indicator variables and causal variables, the plot of one indicator against the other must also be monotonically increasing (or decreasing).

The outcome of a state trace analysis is a state trace plot. The axes of this plot, representing the indicator variables, define a state space for the examined processes (see Fig. 7). The trace factor (item difficulty in this study) in combination with the other causal variables induces variation among the points in the plot, which can be considered "as providing a 'trace' of the behavior of the system in the corresponding state space" (Prince et al., 2012, p. 80). If only one latent process underlies the behavior of that system, variation of the latent factor creates a curve that is monotonically in(de)creasing. When we translate this condition to our nonword reading deficit experiment, and hypothesize that the same reading process is responsible for both the reading behavior in reading disabled and typical readers, all data points should be situated on one monotonically increasing curve. Alternatively, if both reading groups' underlying reading mechanisms differ, each group will have its own curve. Inversely, going from data to theory, if the curve were non-monotonic the data points could not have been produced by a one-dimensional model (cf. Fig. 7). As Prince et al. (2012) point out, the converse conclusion – monotonicity implying a one-dimensional model – is only valid if there is overlap of the sets of data points for both groups on at least one of both axes. We elaborate on this last point when discussing the empirical study. A state trace plot can be visually interpreted as follows (see Fig. 7).

At each point at the *X*-axis typical readers and disabled readers are equated on reading time. Although this equation is based on different word items (easier words for disabled readers), the difference in structural and psycholinguistic characteristics within a word pair is exactly the same as the difference within a corresponding nonword pair. Consequently, the observed difference in reading times of the nonword pairs at the *Y*-axis offers a valid indication of a potential nonword reading problem of one of both groups. Finally, with respect to the rationale of state trace analysis, we implicitly assume that the data are measured without error. Of course, in a real experiment, data are not error-less and this poses the statistical question which model, a one-dimensional or a multidimensional, fits the data best.
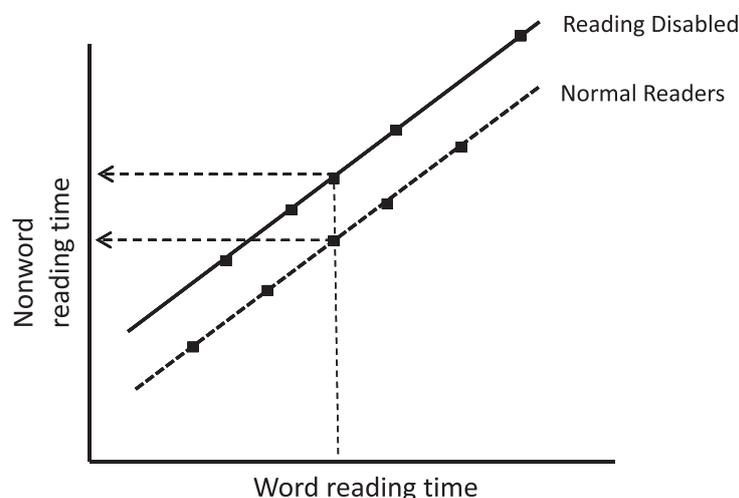

Fig. 7. State-trace plot illustrating non-monotonic data points, indicating multi-dimensionality.

5.2. State trace experiment

Attempting to avoid possible confounds discussed in the previous sections, this experiment was designed to answer the question whether disabled readers show a specific problem in reading non-words after adequately controlling for general reading level. The three formulated hypotheses – a non-word reading deficit, a nonword reading advantage, or a null effect – are treated as a priori equally probable possibilities because, as we extensively discussed, most of the existent studies may not offer a valid empirical basis to make predictions given methodological shortcomings.

5.2.1. Method
5.2.1.1. Participants. 171 children, 94 boys and 77 girls, from third and fourth grade participated in this study. Their mean age was 9 years and 2 months. Ages ranged from 8 years and 5 months to 11 years and 4 months. All children attended regular elementary schools, located in several regions in Flanders (Dutch speaking part of Belgium) and in urban and rural areas. Most children were from indigenous families (79%) and children from foreign origin (21%) were mainly from Moroccan or Turkish descent. All children were checked to have a sufficient command of the Dutch language to be able to attend the Dutch curricula.

The participants were classified into one of two groups. The disabled readers scored ⩽percentile 10 on the widely employed age-normed Dutch One-Minute-Test (OMT, Brus & Voeten, 1973) compared to their specific age group (third or fourth grade). Importantly, this criterion is much stricter than the criteria used in many published NRD studies, in which percentile 20 or 25 generally represents the upper limit. The OMT is highly reliable (parallel test reliabilities around .90) and has a high discriminative power because it is a combination of a power test and a speed test. None of the individuals participating in this study were identified as having serious language, emotional, or behavioral disorders. Estimated IQ-scores of the reading disabled children, based on the vocabulary and block designs subtests of the WISC-R were in the normal range (mean estimated IQ was 93). Although most children would meet the classic discrepancy criterion of developmental dyslexia (reading level significantly below IQ level), this criterion was not explicitly considered. Due to recent developments in dyslexia research, the discrepancy criterion has been largely abandoned (cf. Lyon, Shaywitz, & Shaywitz, 2003; Van den Broeck, 2002). The percentile scores of the nondisabled readers on the OMT were ⩾50. This procedure resulted in 73 disabled readers and 98 nondisabled readers. Although our choice of percentile cut offs, at first sight, gives rise to larger variability in reading performance in the nondisabled group than in the disabled group, our data showed that variability in reading times was actually larger for the disabled group.

5.2.1.2. Design and material. As we explained in Section 5.1, a state trace experiment requires at least three factors to be included in the design. In our study, the dimension factor was Reading Group and the state factor was Lexicality of the presented items. Thus, words and nonwords were presented to typical and disabled readers of the same age. Note that we considered (for practical reasons) all our participants as one age-group because no comparisons between ages were made. The trace factor, which was included to influence the overall level of performance, was Item difficulty. The difficulty of the items was systematically manipulated by varying word length (one, two or three syllables), by varying the number of consonant clusters within a syllable (zero, one, or two clusters), and by varying body-neighborhood (low or high). Body-neighbors are words that share the same rime (e.g. cat, hat, rat) (Ziegler & Perry, 1998). This feature was used to affect item difficulty because research demonstrates that words or pseudowords with many orthographic neighbors are processed more fluently than words with few orthographic neighbors (in naming tasks there are only facilitatory neighbourhood effects for nonwords and low frequency words; see Marinus & de Jong, 2010; Ziegler, Perry, Ma-Wyatt, Ladner, & Schulte-Körne, 2003). Items with low body frequency had a logarithmic body frequency of 1.58 and items with high body frequency had a logarithmic body frequency of 3.02. These counts were based on an electronic version of an official list of Dutch words (Elektronisch Groene Boekje, 1996). By crossing the four variables (lexicality, word length, number of clusters, and body frequency) a $2 \times 3 \times 3 \times 2$ within subjects design was created. In each cell of the design, corresponding to one of the 36 conditions, four items were presented (totaling 144 items, see Appendix C). Although

this number would be too few per cell in a classic experimental design, it is important to acknowledge that word length, number of clusters and body frequency were only included in this design to manipulate item difficulty and hence were not factors of interest.

To make sure that the difference in item difficulty between words and nonwords was exclusively related to lexicality, a structurally matched nonword was constructed for each word. All nonwords were phonotactically legal letter strings (pseudowords): i.e. they could have been real words, although they are not. The structural match within the word–nonword pairs was accomplished by strictly matching on word length in terms of letters and syllables, on syllable structure and consonant structure, and matching as closely as possible on positional redundancy (i.e. the frequency that a letter occurs in a given position in a word, see Appendix C). In order to check that differences in well known psycholinguistic features within functionally matched word pairs ('bag–important') were the same as in the corresponding nonword pairs ('nat–umpalters'), we calculated correlations for the means of summed bigram frequency, orthographic neighbors (N), and positional redundancy for all pairs of word–nonword conditions (see Appendix C). The correlation for corresponding word–nonword conditions for summed bigram frequency was .970, for orthographic neighbors .993 and for positional redundancy .999. This implies that an increment in difficulty from one word condition to another due to any of these characteristics is almost perfectly associated with a corresponding increment for the nonword conditions. Thus, even in a context where positional redundancy, bigram frequency and size of the orthographic neighborhood is systematically larger for the words than for the nonwords (see Appendix C) this would not pose any problems – this situation being actually typical for most nonwords – at least as long as the psycholinguistic difference between word conditions is the same as between the corresponding nonword conditions. To illustrate, imagine the nonword 'nat' with a mean bigram frequency of four, and the nonword 'umpalters' with a mean bigram frequency of eight. Now, if 'nat' is read more slowly by the disabled readers than 'umpalters' is read by the typical readers, then this performance difference cannot be attributed to the difference in bigram frequency at least if the proportional difference in mean bigram frequency of the corresponding word conditions is the same (e.g. the mean bigram frequency of 'bag' being six and that of 'important' being 12).

All items were presented to all participants on a computer screen using E-Prime 1.2 (Schneider, Eschman, & Zuccolotto, 2002). Participants were instructed to read all items correctly and as fast as possible. After presenting six warming-up items, items were presented in blocks of four items (all items of a specific condition). The sequence of the blocks was randomized and the sequence of the items within each block was fixed. Each block was preceded by a red warning signal. Reading times for each item were registered automatically between the presentation of the item and the moment that the participant pressed the space bar. By pressing the space bar the next item of a block was automatically presented. In this way, the pace of the presentation of the items was determined by the participant and the entire processing time for each item from presentation until utterance was recorded. This procedure, including an aspect of fluency, resulted in a smooth progression of the experiment. By using this procedure, the reading times consisted of three components: time to onset of articulation, articulation time, and rest time (time between articulation offset and button press). One may wonder whether all these three components are relevant to our hypotheses.[4] Let us assume that all three components contain 'central' reading processes as well as 'peripheral' processes (e.g. sensory input and motor output processes). As a result of the structural match between the word–nonword pairs ('bag–nat' and 'important–umpalters'), processing time due to peripheral processes can be considered as largely the same for these word–nonword pairs. So, the differences in reading times between words and nonwords must be due to central reading processes. Now it can be proven mathematically that the combination of the structural word–nonword match and the functional word-reading match between reading groups ensures that the crucial difference in reading times between disabled readers' nonwords ('nat') and typical readers' nonwords ('umpalters') must also be the result of central reading processes (see Appendix D for formal proof). Importantly, this procedure captures all central processes starting from the presentation of the stimulus to the moment the participant indicates that she/he is ready for the next stimulus.

---

[4] We are grateful to the action editor for drawing our attention to this potential problem.

Therefore, it can be considered as an ecologically valid procedure, compared to procedures that only capture parts of the entire reading process.

Further, the experimenter recorded whether the items were read correctly or not. Because some of the nonwords could be pronounced in two ways (in some cases the letter 'e' could have two possible pronunciations: /ɛ/ or /ə/), a dominant pronunciation and a less occurring pronunciation, all experimenters were instructed to score both versions as correct. This scoring method did not result in any ambiguities.

5.2.2. Data-analytical procedure and results

The presentation of the results is organized into three sections. The first two sections deal with the issue of the nonword reading deficit. First, we present state trace analysis results for the reading times of correctly read items. Second, we discuss the results for accuracy scores. Third, we present an 'inverse' analysis. After functionally equating typical and disabled readers on nonword reading ability, we can answer the question whether disabled readers suffer from a specific problem in acquiring word-specific knowledge by comparing their real word reading to that of typical readers.

5.2.2.1. Analysis for reading times. Since preliminary inspection of our data revealed that body frequency only exerts a small effect on reading performance, we decided to collapse the data for body frequency so that each condition contained eight items instead of four items. This entailed a reduction of the number of trace factor levels from 18 to 9 and was beneficial for measurement precision and stability while leaving enough trace factor levels to detect potential non-monotonicity. To avoid the detrimental effect of outliers on measurement reliability, all analyses were based on median reading times taken over the eight items of each condition. In a typical state trace experiment, data are averaged over participants in order to inspect the state trace plot for monotonicity (see Fig. 8). Loftus et al. (2004) suggested using Spearman's rank correlation coefficient (Rho) as a non-parametric measure of statistical dependence. If monotonicity is present, rho should equal one. Rho for the data in Fig. 8 was .98. This outcome raises a statistical question: is rho large enough to conclude that the relationship between word reading and nonword reading (based on data points for both groups) is monotonic
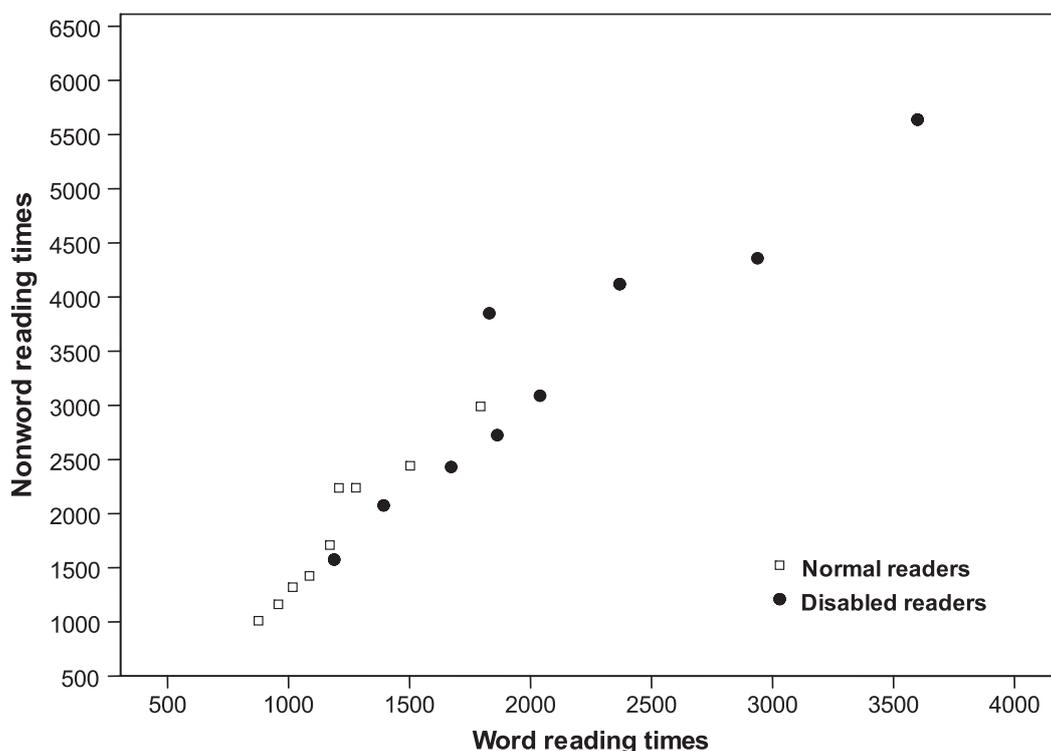

Fig. 8. State-trace plot with data averaged over subjects for normal and disabled readers.

and that the small departure from one is due to random variation? Or, is this departure an indication of a systematic, though small, effect?

Before explaining our adopted statistical procedure, an important problem should be discussed in depth. Although averaging over participants reduces error variance, this method could be misleading if certain conditions are not fulfilled. As Prince et al. (2012) made clear, neither monotonicity nor non-monotonicity is necessarily preserved under averaging when going from the individual level to the overall level. Hence, it is possible that an average monotonic curve is the result of non-monotonic individual curves, and vice versa, that an average non-monotonic curve results from monotonic individual curves. Moreover, by averaging the data individual differences are made invisible. Given these problems, we closely inspected all individual state trace plots as a precautionary measure.

Although more erratic than the average plot, visible inspection revealed that in general, a linear relationship predominated. This visual impression was substantiated by statistically testing in all individuals whether the likelihood of a linear model exceeded the likelihood of a model where quadratic and cubic trends are additionally included. For this comparison Bayesian Information Criteria (BIC) were calculated for both models. The BIC has been proposed by Raftery (1996) as an index to assess the overall fit of a model and allows a comparison of models (see also Long, 1997). Given that BIC assesses whether the model fits the data sufficiently well to justify the number of parameters that are used, the model with the lowest BIC is the best fitting model that is yet parsimonious. These analyses revealed that the linear model was the one with the lowest BIC for all individuals. Given the plausibility of a linear curve for the entire group of individuals, the averaging operation could safely be undertaken. Under these conditions, the curve of the averages (over participants for each condition) equals the average of the curves (based on the individual linear parameters, i.e. intercepts and slopes) (Keats, 1983). Now it would seem natural to perform a hierarchical linear regression analysis on these averaged data points to test whether they are situated on one or two regression lines (one for each reading group). However, as Kliegl et al. (1994) indicated, there are several problems associated with this procedure; more specifically the discarding of the lack of independence of the data, and the discarding of the within-person variance and covariance. In fact, due to the use of a within-subjects design, our data are statistically dependent, violating a central assumption of linear regression analysis. Moreover, as a result of the nested nature of the design (within person conditions are nested under the between person level), within-person variance and between-person variance should be modeled appropriately. For these reasons, we decided to carry out a multilevel analysis adequately taking into account these problems. Note that although the central predictions were formulated within a non-parametric framework, the linearity of our data enabled multilevel parametric statistical methods with high statistical power or sensitivity.

As we mentioned before, state trace analysis is only informative when there is overlap of the sets of data points for both reading groups. Because the aim of this study was to match subjects on word reading ability, the multilevel data analysis was exclusively based on the overlapping data points on word reading performance. Thus, data points corresponding to fast reading times of typical readers were discarded from the analysis if there were no disabled readers with the same fast reading times. And vice versa, data points corresponding to slow reading times of the disabled readers were discarded if typical readers did not show the same slow reading times. This data reduction procedure resulted in a slight reduction of the number of data points (from $171 * 9 = 1539$ to 1520 data points), so that the remaining number guaranteed a powerful analysis. In order to understand why the number of nonoverlapping data points was quite small, two points should be considered. First, as we explained, the match on reading times is not based on the same item material. Second, when taking nine data points for each individual, the overlap is much larger than when the data are averaged over participants (see Fig. 8). Multilevel modeling (Raudenbush & Bryk, 2002) was used to obtain state trace functions for each individual, to obtain average parameters (and their variances) across participants, and to compare the parameters from these functions in the reading disabled and typical reading samples. This analytic approach required the specification of a condition-level model, which represents the median reading time for each nonword reading condition as a function of the reading time from a structurally matched word reading condition for each individual:

$$N_{ij} = \beta_{0j} + \beta_{1j}W_{ij} + r_{ij},$$

where N_{ij} is the predicted median nonword reading time from participant j in condition i. W_{ij} is the median word reading time from participant j in condition i. β_{0j} is the intercept, β_{1j} is the slope relating word to nonword reading times for participant j, and r_{ij} is the residual for participant j in condition i. The person-level model then represents each regression parameter from the condition-level model as a function of the overall mean and each individual's unique effect as follows:

$$\beta_{0j} = \gamma_{00} + \mu_{0j},$$

$$\beta_{1j} = \gamma_{10} + \mu_{1j},$$

where γ_{00} is the average intercept across all individuals, γ_{10} is the average slope across all individuals, and μ_{0j} and μ_{1j} are the increments to intercept and slope associated with individual j. The fixed effects, γ_{00} and γ_{10}, provide precision weighted estimates of the average within-subject intercept and slope, whereas the random effects μ_{0j} and μ_{1j} provide estimates of the within-subject regression parameter variance. This multilevel random coefficients regression model served as the null model. In order to test whether a one-dimensional model or a two-dimensional model provided the best fitting model, the null model was compared with a full model in which a dummy variable was introduced at the participants level to encode for reading group (group = 0 if typical readers, and group = 1 if disabled readers) as follows:

$$\beta_{0j} = \gamma_{00} + \gamma_{01} Group + \mu_{0j},$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11} Group + \mu_{1j},$$

where γ_{01} represents the difference in intercepts and γ_{11} represents the difference in slopes between the reading groups. Word reading scores were centered at the grandmean, which was 1.59 s. Fitting the full model using MPlus 6 (Muthén and Muthén, 1998/2010) produced the following parameter estimates: for the average intercept of the typical readers, γ_{00} = 2.512 (SE = .042, p = .000), for the average slope of the typical readers, γ_{10} = 1.546 (SE = .058, p = .000), for the difference in intercepts between the reading groups, γ_{01} = −.050 (SE = .064, p = .431), and for the difference in slopes between the reading groups, γ_{11} = −.076 (SE = .083, p = .359).

These results show that disabled readers demonstrated a slight nonword reading advantage (a lower intercept). However, the null hypothesis – that the state trace curves for the typical and for the disabled readers do not differ – could not be rejected. As evidenced by a likelihood ratio test, the full model was not significantly different from the null model ($\chi^2[2]$ = .326, p = .425). This means that if no difference between the curves of both groups (if H_0 is true) arose in reality, the probability of finding a difference as large as or even larger than in our sample is .42. As state trace analysis entails that the null hypothesis is in fact a substantive hypothesis, this number does not seem quite convincing. Note that a nonsignificant result due to a lack of power must not be confused with support for the null hypothesis. What we really want to know is the probability that the null hypothesis is true given the observed data (the inverse probability).

To this end, a Bayesian analysis was performed in which the probability of the null model was compared to the probability of the full model given the data and given the assumption that no model was preferred above the other. The BIC-values indicated that the null model fitted the data best (BIC-$_{null}$ = 3704.31 and BIC$_{full}$ = 3717.66). Based on the difference between these BIC-values the 'Bayesian factor' could be calculated. (Kass & Raftery, 1995). The Bayesian factor is equal to the posterior odds in favor of the most likely hypothesis. The posterior odds of the null model (M_N) relative to the full model (M_F) equal:

$$\frac{Pr(M_N / Observed\ Data)}{Pr(M_F / Observed\ Data)}$$

The Bayesian factor favoring the null model was 793, implying that the null model was almost 800 times as probable as the full model, a number that can be interpreted as 'very strong' or 'decisive' evidence (Kass & Raftery, 1995). The parameter estimate for the average intercept of this null model was

γ_{00} = 2.487 (SE = .032, *p* = .000), and for the average slope γ_{10} = 1.506 (SE = .042, *p* = .000). Thus, for all readers an increase in reading time of one second for words was associated with an increase in reading time for nonwords of 1.5 s.

5.2.2.2. Analysis for accuracy scores. Reading skill is the combination of reading accuracy and reading speed. Although reading accuracy was held constant in the analysis of reading times (we only considered items read correctly), it could not be excluded that some participants, especially poor readers, managed to read many items fast and accurately at the expense of making errors on other items. This possible trade-off between accuracy and speed is usually checked in reading research by inspecting the correlation between reading times and number of errors made across participants (or across conditions in a within subjects design). If this correlation indicates that fast reading is generally associated with accurate reading, the trade-off problem is either discarded as harmless or the error analysis is considered as independent from the analysis of reading times. However, the absence of a trade-off at the between-subjects (or conditions) level does not exclude a trade-off at the within-subjects level (van der Linden, 2009). As reading ability is basically constant within an individual, increased speed for some items or conditions may be associated with decreased accuracy. If that is the case, discarding accuracy scores or treating them independently in a separate analysis is not an appropriate method, the reason being that items that are excluded from the analysis, being read incorrectly, can make a unique contribution to the assessment of reading ability.

For these reasons, we started with a multi-level analysis of the correlation between reading times and reading accuracy. This analysis confirmed the absence of a trade-off at the within-subjects level for both word reading and nonword reading (the Pearson correlation between reading times and reading errors for words was .286, and for nonwords .885), as well at the between conditions level averaged over participants (correlation between reading times and reading errors for words was .154, and for nonwords .523). Moreover, as depicted in Fig. 9 the time-accuracy function (TAF), relating reading accuracy and reading speed, averaged over participants for each condition, was nearly linear indicating that reading times and reading errors originate from the same underlying process (Kliegl et al., 1994). Further, TAF's for disabled readers and typical readers were virtually identical demonstrating
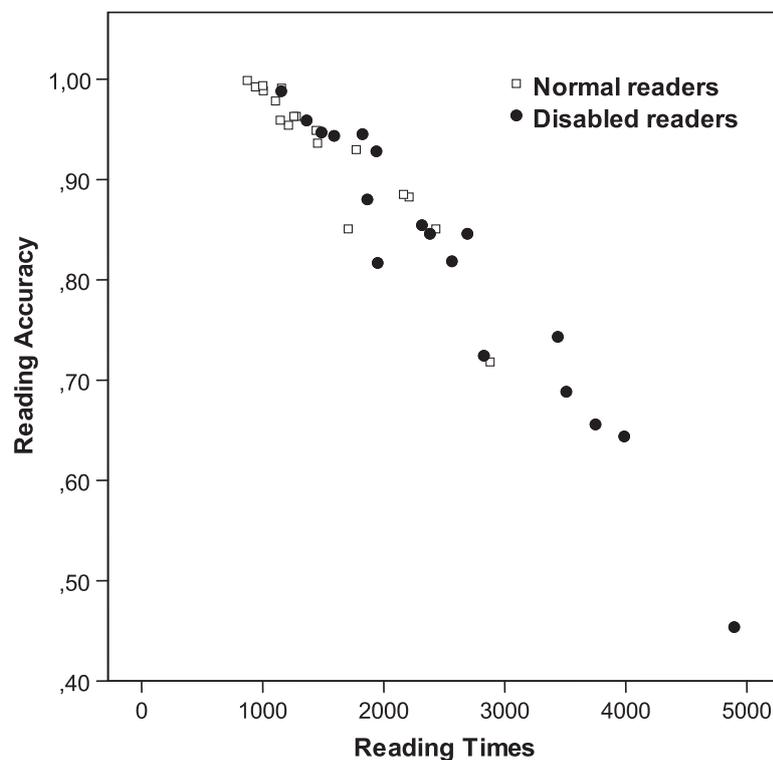

Fig. 9. Time-accuracy function relating reading accuracy and reading speed for normal and disabled readers.

that reading disabled subjects did not suffer from a specific speed or accuracy deficit. Apparently, when disabled readers were matched with typical readers on reading time, no difference in the expected number of errors could be observed.

Given the absence of a trade-off both at the within-subjects level and at the between-subjects level, we proceeded with a separate multilevel analysis of reading errors. Since reading errors are really count data, it would be inappropriate to apply the ordinary linear regression model as this would result in inefficient, inconsistent and biased estimates (Long, 1997, p. 217). A variety of models deals explicitly with characteristics of count outcomes.

The basic model is the Poisson regression model, but it has proven to be vulnerable to the problem of overdispersion (Long, 1997). Overdispersion arises when the variance of the counts exceeds their mean, violating a defining characteristic of the Poisson model, i.e. the equality of mean and variance. As is the case in our error data, overdispersion is often the result of a preponderance of zeros (most items are read correctly). The Negative Binomial Regression Model (NBRM) is a finite mixture model that offers an adequate solution to this problem by explicitly including a dispersion parameter (α) that allows the conditional variance of the counts to exceed the conditional mean (conditional on covariates) (see Hilbe, 2007). Using MPlus 6 a multilevel NBRM was applied to the reading error data. At the condition level the amount of nonword errors (λ_{ij}) was modeled as a function of the amount of word errors for each individual in each condition ($\ln \lambda_{ij} = \beta_{0j} + \beta_{1j}W_{ij} + \varepsilon_{ij}$), and at the participants level the intercepts and slopes of this equation were modeled as a function of reading group (for the full model). As we did earlier, the full model was compared to a null model in order to test for a nonword reading deficit. Fitting the full model, the following parameter estimates were obtained: for the average intercept of the typical readers, $\gamma_{00} = -.143$ (SE = .070, $p = .040$), for the average slope of the typical readers, $\gamma_{10} = .378$ (SE = .050, $p = .000$), for the difference in intercepts between the reading groups, $\gamma_{01} = .096$ (SE = .113, $p = .393$), and for the difference in slopes between the reading groups, $\gamma_{11} = .010$ (SE = .059, $p = .861$).

To interpret these parameters, they should be exponentiated. For instance, the average slope of the typical readers was exp(.378) = 1.46, implying that an increase of one error for words is associated with an increase of about 1.5 errors for nonwords (almost 50%). Also, the average slope of the disabled readers was hardly different from that of the typical readers (exp[.378 + .010] = 1.47): a 1% increase compared to the typical readers that is not significant. The comparison of the full model with the null model based on BIC-values revealed that the null model fitted the data considerably better than the full model ($BIC_{null} = 7254.41$ and $BIC_{full} = 7267.87$). The Bayesian factor favoring the null model was 835, implying that the null model is certainly to be preferred over the full model. The parameter estimate for the average intercept of this null model was $\gamma_{00} = -.104$ (SE = .056, $p = .062$), and for the average slope $\gamma_{10} = .387$ (SE = .034, $p = .000$). Clearly, as for the reading time data no evidence at all was found for the existence of a nonword reading deficit.

5.2.2.3. Analyses to detect a processing deficit of word-specific knowledge. This data-analytical procedure adopted in the previous section could as well be used to examine whether disabled readers have a specific problem in processing word-specific knowledge. If typical and disabled readers are functionally equated on their nonword reading ability (e.g. disabled readers reading 'fim' in exactly the same amount of time as typical readers reading the nonword 'bapeltrup'), the question is whether disabled readers will need more time to read the structurally matched words (e.g. 'cat') compared to the typical readers (e.g. 'paperclip'). Similar to the previous analyses, any difference in word reading times between both groups has to be attributed to the lexicality factor. Fitting a full model, including group as a predictor variable in the person level model, and after centering nonword reading times at the grandmean, MPlus 6 produced the following parameter estimates: for the average intercept of the typical readers, $\gamma_{00} = 1.515$ (SE = .021, $p = .000$), for the average slope of the typical readers, $\gamma_{10} = .433$ (SE = .018, $p = .000$), for the difference in intercepts between the reading groups, $\gamma_{01} = .143$ (SE = .032, $p = .000$), and for the difference in slopes between the reading groups, $\gamma_{11} = .024$ (SE = .026, $p = .353$). These results show that disabled readers demonstrated a significant word reading deficit (a higher intercept). As indicated by a likelihood ratio test the full model was significantly different from the null model ($\chi^2[2] = 4.782$, $p = .046$). The Bayesian factor favoring the full model was 237, providing very strong evidence against the null hypothesis.

Analyzing the errors with a Negative Binomial Regression Model, fitting the full model, the following parameter estimates were obtained: for the average intercept of the typical readers, $\gamma_{00} = -1.862$ (SE = .097, p = .000), for the average slope of the typical readers, γ_{10} = .422 (SE = .030, p = .000), for the difference in intercepts between the reading groups, γ_{01} = .369 (SE = .149, p = .013), and for the difference in slopes between the reading groups, γ_{11} = −.048 (SE = .037, p = .192). Although disabled readers made significantly more errors on the real words, the full model was not significantly different from the null model ($\chi^2[2]$ = 1.917, p = .19). Also the Bayesian factor hardly favored one model over the other (1.81).

Taken together, the analyses of reading times and reading errors offer substantial evidence for a genuine problem of disabled readers in their processing of word-specific knowledge. Also, the combined raw effect size seems to be substantial. The disabled readers made on average 1.45 (exp[.369] = 1.45) more errors than the typical readers (out of eight items), and they took 143 ms. longer to read words correctly.

6. Discussion

In this article, we presented an extensive and critical survey of the methods and designs used in the literature to study specific processing deficits in reading disability. In one way or the other, the objective in these studies is similar: to control for general reading level in order to examine the existence of a nonword reading deficit. Successively, we discussed the technique of age- or grade-equivalent scores, standard scores, the reading-level match design (RLM), and the use of interaction effects in a chronological-age match design. For all these methods, with the exception of standard scores, we reported methodological shortcomings. For the method based on age- or grade equivalent scores and the reading-level match design, which proved to be mathematically equivalent, we argued that their main limitation lies in the fact that they fail to account for normal developmental trends. In Van den Broeck et al. (2010), we reasoned that an increase in word-specific knowledge with age provides a plausible alternative explanation for nonword reading deficit findings in a RLM-design. At this point in the discussion, we take this argument one step further. When disabled readers are compared with typical readers in a RLM design and the basic design principles are satisfied (entailing that the selection variable equals the matching variable; and reading disability is exclusively operationalized by performance on word reading ability), both groups can only differ from each other in age. Hence, any difference observed between the two groups on the criterion variable, i.e. nonword reading, can in fact only be attributed to factors related to the *age-difference*. In other words, in such circumstances, a RLM design is a pure developmental design suited to detect age-related shifts in the measured reading abilities. Therefore, we suggest this design to be used in future research aimed at detecting developmental differences. In our view, the standard interpretation of the RLM design as a quasi-experimental design is problematic and for that reason should be abandoned.

The next method we discussed, the standard score method, was described as an established and valid method to assess the relative position in same-aged participants' word and nonword reading ability. Yet, we pointed out that at least two related elements linked to the use of standard scores seem to be less familiar, both in practice and research: first, the fact that a regression effect is unavoidable and second, the observation that disabled readers will show a nonword reading *advantage* when they are diagnosed with a real-word reading test. A practical diagnostic implication of these observations is that a reassessment of a disabled readers' reading ability using a *nonword* reading test will generally result in a higher standard score compared to the initial measurement of real-word reading ability. It should be emphasized, however, that this standardized measure of relative standing provides no information about the absolute difference in processing load between words and nonwords in terms of a substantive measurement (e.g. extra time needed).

It is this latter issue, the measurement of the different processing load in typical and disabled readers in their word versus nonword reading ability, which in fact is the objective of the subsequently discussed method: the study of interaction effects in a chronological-age match design. Also, for this method, we reported methodological problems. The central concern was that interaction effects in a chronological age match design may give rise to scale dependency problems. These could be caused by

a general slowing factor in disabled readers, for instance in cases where no adequate controls for general reading level are assured. We reported a number of improvements to this method, but also proposed an alternative: state trace analysis.

In our search for an unconfounded nonword reading deficit, a method was required that would take into account the aforementioned methodological and interpretational problems inherent to the frequently adopted methods that we analyzed. State trace analysis came about as a valuable solution. State trace analysis can be described as a case of factor analysis in an experimental design. By using this method, it is possible to test individuals of the same age, hence avoiding developmental confounds. State trace analysis examines the relationship between two indicators of the reading system by manipulating at least two independent variables. Furthermore, the behavior of this reading system is observed over a range of items varying in difficulty. According to Prince et al. (2012), state trace analysis has the advantage of separating the issue of dimensionality (the number of underlying dimensions and their nature) from specific parameter model assumptions.

6.1. Interpretation of the empirical results in terms of different hypotheses

Given these advantages of the state trace method on methodological and interpretational grounds, we decided to use this method to find out whether Dutch-speaking reading disabled children show a specific problem in reading nonwords, as has been proposed in the literature. Applying state trace analysis in our sample of Dutch-speaking children and using Bayesian data-analytical methods, our data supported the substantive null hypothesis that disabled readers do not differ from typical readers on nonword reading, given that general word reading level is adequately controlled for. This conclusion was reached both on the basis of reading times and accuracy scores. Given the established character of the standard NRD interpretation in the literature, this conclusion asks for a further discussion of its implications for theories of reading disability. Strictly, one could limit these implications to theories of reading in the linguistic context of the investigated participants, i.e. Dutch. Since Dutch is a largely consistent language with a shallow connection between orthography and phonology (Patel, Snowling, & De Jong, 2004; Seymour, Aro, & Erskine, 2003), our results may be safely extrapolated to other consistent languages. Moreover, in Italian, another shallow orthography, similar reports showing the absence of a NRD have been made (Di Filippo et al., 2006; Zoccolotti et al., 2008). But what about disabled readers learning to read in an inconsistent or deep orthography like English (Share, 2008)? This question is of course empirical in nature and should be answered by a replication of this study in an English sample. For now, our perspective is theoretical. In what follows, we consider which predictions could be derived from our study with Dutch participants for disabled readers in an inconsistent orthography. We successively discuss predictions based on the general interpretation of the phonological deficit, psycholinguistic grain size (PGS) theory, an alternative developmental view, and the decoding stagnation hypothesis.

To start with, based on the phonological deficit hypothesis (Rack et al., 1992; Vellutino et al., 2004) and the idea that nonword reading builds on phonological skills to a greater extent than word reading, researchers have assumed that disabled readers are characterized by a genuine nonword reading deficit, *regardless* the depth of the orthography involved (e.g. van der Leij & Morfidi, 2006; but see Van den Broeck et al., 2010). However, reasoning from the currently influential psycholinguistic grain size (PGS) theory, proposed by Ziegler and Goswami (2005, 2006), different predictions should be made for different orthographies. According to PGS theory, the phonological representations, which are activated in the reading process, are the reflection of the grain size that is required in each language to attain orthographic-phonological consistency. In English, larger orthographic units such as rimes are generally more consistent than single graphemes (e.g. to arrive at a correct pronunciation of the grapheme 'a', the context of the rime unit is relevant – ave is pronounced consistent as in 'gave', 'save' and 'rave') (Treiman, Mullennix, Bijeljac-Babic, & Richmond-Welty, 1995). In consistent languages like Dutch, German, Italian etc., there is no need to rely on these larger units, in most of cases, to arrive at a correct pronunciation of graphemes (Martensen, Maris, & Dijkstra, 2000). Based on these observations, Ziegler and Goswami proposed that reading in consistent orthographies involves small linguistic units, whereas reading in inconsistent orthographies requires the use of larger units as well. We discuss predictions for both consistent and inconsistent orthographies respectively. First, if reading words and

nonwords in a consistent orthography would indeed rely on the same small linguistic units (grapheme–phoneme correspondences) as hypothesized by PGS theory, then no, or at least a less sizeable nonword-reading deficit should be expected, which is in accord with our state trace results. If, on the other hand, the processing of words would differ from nonwords (use of mainly larger units vs. mainly small units) in an inconsistent orthography, this could lead to the prediction of a more substantial nonword reading deficit, at least if one adheres to the classic assumption formulated by the Haskins group in the 1970s that the phonological deficit of disabled readers primarily affects the reading of small units (corresponding to phonemes) (Liberman, Shankweiler, Liberman, Fowler, & Fischer, 1977; see also Share & Stanovich, 1995, p. 10).

Still according to an alternative developmental hypothesis, phonological processes are equally involved in word and nonword reading (see Van den Broeck et al., 2010, pp. 729–730). In fact, all recent theoretical accounts of written word recognition agree that apart from a sequential more laborious grapheme–phoneme mode of processing, a faster interactive mode of processing is needed to read real words fluently and almost automatically (cf. Coltheart et al., 2001; Perry et al., 2007). Importantly, this interactive mode of processing written words is at least partly phonological in nature. From this line of reasoning, a phonological deficit would equally affect word and nonword reading. Hence, no nonword reading deficit is expected, neither in consistent orthographies nor in inconsistent orthographies, like English.

Even though research indeed indicates that phonological processes are involved both in word and nonword reading (Ashby, 2010; Diependaele, Ziegler, & Grainger, 2010), the relative impact of these phonological processes could still be different. Based on this latter idea, the decoding stagnation hypothesis is proposed. This hypothesis makes different predictions for different orthographies, yet in the opposite direction as the hypothesis we derived from PGS theory. The decoding stagnation hypothesis suggests that the processing difference between words and nonwords in disabled readers is smaller than in typical readers because typical readers have a larger word-specific knowledge base (Van den Broeck et al., 2010). Lacking such an extensive word-specific knowledge base, disabled readers have to rely on their ability to decode smaller grapheme–phoneme units. As a consequence, disabled readers are relatively less affected by the processing difference between words and nonwords. Relating this hypothesis to the issue of orthographic consistency, one could argue that the role of word-specific knowledge is relatively larger in inconsistent orthographies, thus implying a larger processing difference between words and nonwords in typical readers. Unable to profit fully from their word-specific knowledge, these typical readers' performance on the nonwords will decline more considerably than that of disabled readers. Thus, the decoding stagnation hypothesis predicts that disabled readers show a nonword reading *advantage* compared to typical readers.

Noticeably, we found partial support for this latter prediction in our state trace analysis findings. As we reported in Section 5.2.2.1, no statistically significant nonword reading effect was observed in our data for the analysis of reading times. However, when eliminating word reading times above 2.5 s. from the analysis, removing the most erratic part of the data and retaining 89.2% of the original data points, a multilevel state trace analysis revealed a statistical trend toward a nonword reading advantage for the disabled readers. The difference in intercepts between the reading groups was −.103 (SE = .054, *p* = .059) and the Bayesian factor supporting the full model was 13.1, indicating 'positive evidence' according to Kass and Raftery (1995). If this trend would indeed reflect a genuine processing difference between the reading groups, as proposed by the decoding stagnation hypothesis, possibly a more substantial nonword reading advantage could be expected in a sample, learning to read in an inconsistent orthography like English. Thus, the decoding stagnation hypothesis would predict an interaction effect between depth of orthography and size of the nonword reading effect. As the current data from the state trace analysis reveal a nonword reading advantage trend for disabled readers in a consistent orthography, i.e. Dutch, one could argue that they mildly favor this prediction. Further cross-linguistic research should, however, decide on this matter.

Another important finding in our data, using state trace analysis, was that disabled readers revealed inferior word reading performance after they were functionally equated with the typical readers on decoding ability (nonword reading). This new empirical evidence strongly suggests that one fundamental problem of disabled readers concerns their (dis)ability to acquire word-specific knowledge. This result is in line with a substantial body of empirical evidence from behavioral studies

showing that the build-up of orthographic knowledge is not absent, yet impaired in disabled readers (Bergmann & Wimmer, 2008; Ehri & Saltmarsh, 1995; Lachmann & van Leeuwen, 2008; Reitsma, 1983; Share & Shalev, 2004; Staels & Van den Broeck, submitted for publication). This deficit also explains why disabled readers have to rely to a greater extent on decoding reading strategies compared to typical readers.

Importantly, the aforementioned different perspectives and their implications do not affect the main outcome of this paper: disabled readers do not show a genuine and specific nonword reading deficit in a consistent orthography.

6.2. Suggestions for further research based on state trace analysis

Some further suggestions for future research can be made, based on the technique of state trace analysis. As the adopted state trace method proved a reliable and improved device to match the participants on the critical control variable word reading, this technique may prove valuable as well to study other key research questions in the reading domain as well as in other research areas. Concerning reading disability research, four useful applications can be distinguished. First, other marker effects such as the length effect (a specific problem disabled readers have been reported with the reading of longer words; Martens & de Jong, 2006; Spinelli et al., 2005) could be studied in an analogous way as conducted in this article. Second, the state trace method offers a unique approach to study several marker effects cross-linguistically. Adopting this technique, reading disabled and typical reading groups in different languages could be simultaneously matched on the critical control variable.

Third, an important issue to reflect on in this discussion is the implication of our findings for theories that support the existence of subtypes in reading disability. In the introduction we referred to these theories and argued in favor of a stepwise research strategy in which the first aim is to establish whether the entire group of disabled readers can be characterized unambiguously as having a nonword reading deficit or not. As we found no evidence for a nonword reading deficit in the examined population, future research which makes use of state trace methodology could be aimed at studying heterogeneity within the group of disabled readers (cf. Van den Broeck et al., 2011). After all, one should not ignore that the size of the nonword reading advantage in our sample of disabled readers was modest. On the one hand, this could suggest that not every individual with reading disability would demonstrate these effects. On the other hand, given the high correlation between word reading ability and nonword reading ability in our sample, the magnitude of word–nonword reading dissociations could also be modest. In this respect, the existence of so-called 'hard subtypes' (Stanovich et al., 1997, p. 115) could be doubted in particular. Hard subtypes are reading disabled participants who score at a normal level on a measure of sublexical processing (e.g. nonword reading) but outside the normal range on a measure of lexical processing (e.g. exception word reading), or vice versa. Milder or relative dissociations between both types of processing could be observed in the so-called 'soft subtypes'. Given the constraints of the observed strong relationship between word reading and nonword reading ability, such mild dissociations between lexical and sublexical processing would imply a relative impairment of one of the processing modes. To study these relative impairments, state trace analysis could be used to construct individual state traces. To determine subtypes, parameters of individual state trace curves could be compared to the overall distribution of parameters (see van der Mark et al., 2011).

Fourth, it is also of vital interest in neuroimaging research on reading disability to disentangle poor task performance in the scanner, decreased reading ability, and a dyslexia specific deficit (see also Palmer, Brown, Petersen, & Schlaggar, 2004). In this research domain, recently the first performance matching attempts have been made in trying to unravel these effects by making use of a double control group, a reading-level control group and a chronological age control group (Hoeft et al., 2006, 2007). However, as we have explained in this paper, the juxtaposition of two different confounds, one between reading group and age and one between reading group and reading level, is not a desirable solution because the first confound does not necessarily compensate for the second and vice versa. Therefore, a better solution in many neuroimaging studies would be the application of the here presented functional matching technique, based on state trace analysis.

In general, state trace analysis as a matching technique could be effectively adopted whenever a group showing a particular disorder has to be matched with a typical group, in order to test for a hypothesized specific deficit.

Before discussing the implications of our findings for theories of reading disability, so as to place our findings in a broader theoretical context, we proceed in the following section with an overview of recent neurobiological studies pertaining to reading processes in typical and disabled reading.

6.3. Relation to neurobiological studies in typical and disabled reading

Recently, there has been an ever increasing number of neurobiological studies that shed light on typical and disabled reading development (for reviews see, Dehaene & Cohen, 2011; Price & Devlin, 2011; Schlaggar & McCandliss, 2007). As we will show in the discussion that follows, these neurobiological studies largely deal with the same issues and interpretational problems as are present in behavioral studies. Concerning the question of what neuroimaging studies contribute to the study of cognition, differing opinions have been put forward. Although some researchers have claimed that so far functional neuroimaging research has not been able to distinguish between competing psychological theories (Coltheart, 2006), and that even in principle such studies cannot provide evidence about the nature of cognition (Van Orden & Paap, 1997; see also Uttal, 2001), many researchers argue that neurobiological measures, though not intrinsically explanatory, simply comprise another dependent variable (Henson, 2005) or description (Frost et al., 2009), along with behavioral data (see also Poldrack (2006, 2008) on 'reverse inference'). In the following, we explore which, if any, additional constraints neurobiological data impose on theories of typical and disabled reading.

6.3.1. The cortical substrate of word identification

Evidence from neuroimaging studies reveals that the cortical substrate of skilled word identification involves the tuning of a broadly tripartite left hemisphere brain circuit (see Frost et al. (2009) and Pugh et al. (2001) for reviews). Frost et al. (2009) distinguish an anterior subsystem and two posterior subsystems: an occipitotemporal (ventral) and a temporoparietal (dorsal) system. The anterior system, located in the inferior frontal gyrus (IFG), is involved in phonological coding and in semantic processing. According to Poldrack et al. (1999), the ventral part of the IFG is associated with semantic retrieval while the more dorsal region is involved in phonological processing (articulation and subvocal rehearsal). The temporoparietal subsystem, including Wernicke's area, the angular gyrus and the supramarginal gyrus, functions as an association zone between visually presented letter strings and their phonological and semantic counterparts (Price, Winterburn, Giraud, Moore, & Noppeney, 2003). This dorsal pathway can be viewed as a serial attention spotlight applied to the letter-by-letter analysis of low frequency or pseudowords (Cohen, Dehaene, Vinckier, Jobert, & Montavont, 2008). The ventral pathway is located in the left ventral occipitotemporal cortex (VOT) near the fusiform gyrus and responds automatically and rapidly to visual words and their abstract orthographic properties (Binder, Medler, Westbury, Liebenthal, & Buchanan, 2006; Glezer, Jiang, & Riesenhuber, 2009; Vinckier et al., 2007), that is, orthographic properties that are invariant for case, size, font (or handwriting) and location (e.g. a and A are visually quite different, yet they are coded as the same orthographic sign).

This VOT includes the so called Visual Word Form Area (VWFA; Cohen et al., 2000), the functioning of which has instigated by far the majority of recent neuroimaging studies on word reading. Rather than being a discrete location, the visual word form system extends along the VOT surface, showing a hierarchical organization with a posterior to anterior gradient of increasing print specificity to larger orthographical and more complex units (Brem et al., 2006; Dehaene, Cohen, Sigman, & Vinckier, 2005; Van der Marck et al., 2009; Vinckier et al., 2007). Along this gradient, neurons are tuned to increasingly complex word features, viz. from oriented bars, to letters, bigrams (ordered pairs of letters), morphemes and small words (Dehaene et al., 2005; Glezer et al., 2009). Although the VWFA was initially thought of as a prelexical, bottom-up processor (Dehaene, Le Clec, Poline, Le Bihan, & Cohen, 2002) based on the idea that single-word reading can be seen as a specialized form of object processing, a series of recent studies have provided empirical evidence that this brain region is sensitive to the processing of orthographic information at all grain-size levels, from the single grapheme up to the whole word level (2007; Bruno, Zumberge, Manis, Lu, & Goldman, 2008; Glezer et al., 2009; Kronbichler et al.,

2004, 2007; Schurz et al., 2010; but see Braet, Wagemans, & Op de Beeck, 2012). In these studies, the VWFA was shown to be sensitive to the familiarity of the orthographic forms, in such a way that increasing familiarity led to a reduced activation. Moreover, these studies indicate that VWFA would also play a role in serial sublexical processing of unfamiliar orthographic strings, hence showing a double function. Schurz et al. (2010) demonstrated that activity in VWFA was dependent on a length by lexicality interaction, that is a small or absent effect of length for familiar orthographic forms, and a substantial length effect for pseudowords. On the one hand, this double function of VWFA relates easily to dual process models of visual word recognition (cf. Coltheart et al., 2001). On the other hand, both the behavioral and the brain studies seem to indicate that word level and subword level processing is far more interdependent than assumed in dual route accounts of reading.

6.3.2. Interactive learning in the Visual Word Form Area (VWFA)

Because the human brain is initially not prepared or specialized for the reading task (McGuinness, 1997), it requires an intensive and protracted learning process before it (more specifically VWFA) can function as an efficient orthographic input lexicon (the recycling hypothesis of Dehaene and Cohen (2007)), capable of fast and correct identification of most printed words. On theoretical grounds, it would be difficult to regard this learning process as a pure bottom-up perceptual encoding process, since learning to read involves establishing the contact between arbitrary visual signs and already stored semantic and phonological information. But also empirically, there is now ample evidence that higher cortical language zones, including the dorsal system, serve to train the ventral visual word form system to recognize the relevant orthographic patterns via an interactive process (see also Pugh et al., 2001). First, although a few reading sessions are sufficient to demonstrate an enhanced VWFA activation in adults learning to read a new script, mere visual familiarity by itself is not enough (Hashimoto & Sakai, 2004; Song, Hu, Li, Li, & Liu, 2010; Yoncheva, Blau, Maurer, & McCandliss, 2010). Apparently, the development of VWFA activation requires systematic attention to the letter–sound correspondences. The association between VWFA and language processing is further sustained by the observation that left laterality of VWFA activation goes hand in hand with left lateralization of the frontal language areas (Cai, Paulignan, Brysbaert, Ibarrola, & Nazir, 2010). There is also direct evidence that VWFA receives top-down influences from higher language areas during spoken language tasks. During an active listening task (Yoncheva, Zevin, Maurer, & McCandliss, 2010) and during a spoken lexical decision task (Dehaene et al., 2010) increased activity in VWFA was observed relative to surrounding regions, whereas passive listening to sentences did not lead to VWFA activation. So it seems that the top-down recruitment of VWFA is not mandatory, and is mainly observable when the task requires an active strategy to retrieve orthographic information (but see Kherif, Josse, & Price, 2011).

This issue relates to the discussion in behavioral studies on the potential influence of feedback from phonology to orthography, brought about in the context of feedback consistency effects (Stone, Vanhoy, & Van Orden, 1997). Words are feedback inconsistent when their shared phonological units could be spelled in different ways (e.g. /-ip/ in deep and heap). Although inhibitory effects of feedback inconsistency in naming tasks as well as in lexical decision tasks were initially reported (Pecher, 2001; Pexman, Lupker, & Reggin, 2002; Stone et al., 1997; Ziegler, Montant, & Jacobs, 1997), other studies have questioned the validity of these results, based on methodological grounds (Norris, McQueen, & Cutler, 2000; Peereman, Content, & Bonin, 1998). In a large-scale regression study conducted by Balota, Cortese, Segernt-Marshall, Spieler, and Yap (2004) feedback inconsistency effects at the rime-body level were shown to be restricted to the slowest subjects, and to low frequency words (see McKague, 2005). In summary, based on both the neurobiological studies as well as the behavioral studies, it seems that even though the reading system has the potential for full interactivity (or resonance in the terminology of Stone et al. (1997)), this interactivity is usually restricted by task demands and the novelty of the stimulus material.

6.3.3. Implications for theories of visual word recognition

These restrictions on a fully interactive reading system are not without theoretical implications. With this respect McKague (2005; see also McKague, Davis, Pratt, & Johnston, 2008) has offered a significant argumentation and integrative theoretical proposal to explain how the reading system succeeds in dealing with the conflicting requirements of fast reliable word recognition on the one

hand, and the need to learn novel words on the other hand. This latter conflict was identified by Grossberg (1987) as the *stability–plasticity* dilemma. Fast and reliable identification of already stored lexical items is well captured by the traditional modular theories of word recognition (Coltheart, 1978; Forster, 1976; Morton, 1979). Besides these models also most of the current models of visual word recognition are described by McKague as stationary models because they consist of a predetermined lexicon with no mechanism specified to enable the acquisition of new representations. Neither the indirect sublexical route of the dual route model (Coltheart et al., 2001; Perry et al., 2007), nor the back propagation learning algorithm of the triangle parallel distributed processing (PDP) model (Harm & Seidenberg, 2004; Plaut et al., 1996; Seidenberg & McClelland, 1989) provide a mechanism for ongoing orthographic learning, capable of learning beyond the initial training set or the prewired local representations (see McKague (2005) for further discussion).

In contrast, the subsymbolic resonance model of Van Orden and colleagues (Stone & Van Orden, 1994; Van Orden & Goldinger, 1994; Van Orden et al., 1990) is capable of orthographic learning by generalizing its knowledge to new inputs (although the model was never implemented). Embracing a fully interactive architecture that does not depend on explicit feedback external to the system (as in the triangle model), this model is pre-eminently suited to learn the systematic relationship between orthography and phonology via covariant learning principles (e.g. Hebbian learning). However, this interactive connectivity entails a concomitant instability in the system, because of the preponderance of one-to-many mappings that exist between phonology and orthography at the lexical (homophones) and at the sublexical level (feedback inconsistency), at least in English, but also to a lesser degree in many other languages with a shallower orthography. According to McKague (2005), it is precisely the lack of localist representations and hierarchical structures in this resonance model that prevents it from developing data structures that are impervious to the potentially contradictory effects of mandatory feedback, and that precludes the stable encoding of word-specific spelling patterns.

Based on these theoretical considerations, and on a series of empirical studies on orthographic learning manipulating feedback consistency, McKague (2005) proposes an item-specific learning mechanism "*capable of shifting between its specialist input-driven recognition mode and a more flexible learning mode in which the entire system of connections provides feedback to enable the output of orthographic, phonological and semantic processes to become associated. The ability to shift modes suggests that the literate mind must be sensitive to novelty in such a way that attentional processes can be captured by a novel input when it fails to activate a stored lexical representation, thus triggering the more flexible learning mode.*" (McKague, 2005; p. 26). An essential mechanism during this learning mode is sublexical feedback from phonology, or "orthographic recoding" (McKague, 2005), by which orthographic units (e.g. letter clusters) get unitized in such a way that they correspond with their phonological counterparts. This orthographic recoding mechanism would play a previously unrecognized yet vital role in orthographic learning, and is viewed as a necessary complement to the item-based developmental theories of Ehri (1992), Perfetti (1992), and Share (cf. self-teaching mechanism, 1995), all theories which only assume a feedforward flow of information from orthography to phonology.

6.3.4. Plasticity of the Visual Word Form Area (VWFA)

Could the VWFA be that cortical structure that is capable of efficiently retrieving stored sublexical and word-specific (lexical) orthographic forms and at the same time enabling the learning of new orthographic forms? At least, there is converging evidence from neurobiological studies suggesting that the VWFA is wired up by long-term experience since it responds specifically to word forms and letter strings of an orthography known by the participant. The 'skill zone' interpretation of VWFA finds support in several studies (using fMRI or ERP) pointing to a gradual tuning of the activation pattern with increasing reading ability toward an adult-like activation pattern (Brem et al., 2006; Maurer, Brem, Bucher, & Brandeis, 2005; Maurer et al., 2006). There is also correlational support for VWFA as a skill zone: Bruno et al. (2008) found a positive correlation between word reading efficiency and left VOT activation, whereas a negative correlation was reported between word reading ability and activation in the inferior frontal gyrus (IFG) and the left posterior superior temporal gyrus (Wernicke's area), which is associated with phonological access. Further, increased left VOT (including VWFA)

activity correlated more strongly with reading ability than with age (Sandak, Mencl, Frost, & Pugh, 2004).

The development of VWFA print sensitivity was shown in a longitudinal training study to emerge rapidly during acquisition of grapheme–phoneme correspondences (GPC's) in young pre-reading children who practiced these GPC's for only 4 h (Brem et al., 2010). Another study demonstrated that short-term language experiences in adult readers who learned to associate novel stimuli with English letters, resulted in an altered VWFA response (Song, Bu, Hu, Luo, & Liu, 2010). Further analysis revealed that this learning-induced representation engaged the same neural population underlying the representation for language forms constructed through long-term experience. In a notable longitudinal fMRI study, Ben-Shachar, Dougherty, Deutsch, and Wandell (2011) followed 7–12-year-old children for four consecutive years, and showed age-related increases in VWFA sensitivity to written words. Moreover, they reported that increased sensitivity was associated with the rate of improvement in sight word reading. Importantly, during reading skill acquisition, a transition takes place from bilateral (left and right hemisphere) extrastriate region activation to a predominance of left VOT involvement (Brown et al., 2005; Schlaggar et al., 2002; Turkeltaub, Gareau, Flowers, Zeffiro, & Eden, 2003).

In summary, due to the double function of the VWFA serving as a modular orthographic lexicon, and as an interactive learning mechanism, it seems that VWFA is an excellent candidate cortical region to solve the stability–plasticity dilemma with which the reading brain is confronted.

6.3.5. Brain functioning in reading disability

As already reported, VWFA activity correlates positively with word reading skill. Indeed, VWFA activation patterns explain about 50% of the variance in reading speed across individuals (Dehaene et al., 2010). Hence, it is not surprising that reading disabled children showed evidence of an impaired VWFA specialization for print processing and orthographic familiarity (Shaywitz et al., 2007; Van der Marck et al., 2009). Similarly, adults with dyslexia demonstrated diminished engagement of posterior cortical structures (VWFA, angular gyrus and Wernicke's area) and showed a tendency to over-engage anterior structures (e.g. inferior frontal gyrus) (Paulesu et al., 2001; Pugh et al., 2000; Shaywitz et al., 1998). Another reported finding is that reading disability is associated with increased activity in posterior (temporoparietal) right hemisphere regions (Shaywitz et al., 2002; Simos, Breier, Fletcher, Bergman, & Papanicolaou, 2000; but see Richlan, Kronbichler, & Wimmer, 2009). According to the typical interpretation of these anomalies, reading disabled individuals would rely more on an elaborate letter-by-letter processing with a larger investment on working memory capacity in order to compensate for their lack of automatic word recognition.

Several studies have shown that reading intervention programs aiming at improving the reading skills of disabled readers resulted in both improved reading measures and changed brain activity patterns (Sandak et al., 2004; Shaywitz et al., 2004; Simos et al., 2002; Temple et al., 2003). Typically, these interventions induced changes in brain activation patterns that resemble those of typical readers, i.e. an increase in the engagement of the left temporoparietal region, accompanied by a reduction in the activation of the right temporoparietal areas (see Frost et al. (2009) for a review). Although these intervention programs were typically phonics-based, it should be noted that they were never purely phonological (or auditory), because the relationship with printed words was always present. For that reason, they do not allow a causal interpretation of the role of phonology. Similar intervention studies with adolescent or adult disabled readers also revealed improvements in left reading-related brain regions, but also compensatory activity in right perisylvian cortex (Eden et al., 2004; Pugh et al., 2008). The relative ease with which these changes in brain function came about, give rise to a fundamental question (see also Frost et al., 2009). When reading interventions can lead to improved brain functions relatively easily, why are these disabled readers still fairly poor readers? Apparently, the improved brain scans are not telling the whole story, though they seem to indicate that the functional brain regions are not permanently disabled. We concur with Frost et al. (2009) that these readers appear to fail to consolidate the learning experiences into longer term changes in neural functioning.

Recently, empirical evidence is accruing that the core problem of disabled reading concerns a specific connectivity disorder in binding orthographic and phonological information. In a recent neuroimaging study van der Mark et al. (2011) showed that during a phonological lexical decision task the VWFA was functionally connected with typical left frontal and parietal language areas in control children, whereas a significant disruption between these areas was observed in children with reading disability. In another fMRI-study of Booth and coworkers (Desroches et al., 2010), who examined the differences in brain activation between reading disabled children and age-matched normally achieving children during an auditory rhyming task (with orthographically conflicting endings, pint–mint or jazz–has, and orthographically nonconflicting endings, gate–hate or press–list), both groups showed similar activation in bilateral superior temporal gyri, a region associated with phonological processing, while only typically achieving children showed activation of left fusiform cortex, involved in orthographic processing. These findings suggest that typical reading children activate orthographic representations during spoken language processing, while those with reading difficulties do not. This last group appears to suffer from impairment in integrating phonological and orthographic representations.

Given the fact that reading engages several distant cortical areas, proficient reading development not only involves sufficient recruitment of the relevant cortical regions, but also adequate interactivity between these regions. Paulesu and coworkers (1996) were the first to hypothesize that a dysfunction in cortical connectivity would lie at the root of dyslexia. The anatomical substrate enabling efficient signal transmission between distal cortical regions consists of long axonal connections (white matter tracts). Using diffusion tensor imaging (DTI) methodology, measuring the integrity of white matter tracts, indexed by fractional anisotropy (FA), several studies revealed strong positive correlations between standardized reading measures and FA in a left temporoparietal region, while other studies demonstrated that FA differed in this region between disabled and typical readers (for reviews see Schlaggar & McCandliss, 2007; Vandermosten, Boets, Wouters, & Ghesquière, in press). Moreover, intensive remedial instruction in word-level decoding skills was shown by Keller and Just (2009) to lead to a positive change in left hemisphere white matter tracts of poor reading children (due to an increased myelination).

Finally, the research group of Blomert and colleagues has collected empirical evidence, based on electrophysiological and neuroimaging studies, for their position that the orthographic-phonological binding deficit in dyslexic readers manifests itself in a specific problem to integrate letters and speech sounds early in reading development (see Blomert (2011) for a review). Although dyslexic readers learn to associate letters with speech sounds, they seem to be unable to integrate them into new audiovisual objects (Froyen, Willems, & Blomert, in press). Remarkably, letter–speech sound association strength proved to be independent from performance in speech sound discrimination and phonological awareness. The formation of these integrated letter–sound audiovisual objects develops slowly, and is considered as a necessary milestone towards the tuning of larger orthographic-phonological units.

6.4. Implications for theories of reading disability

The picture that emerges from this overview of neurobiological studies deviates considerably from what can be derived from the standard phonological deficit hypothesis and may be interpreted as a call to revise this hypothesis. Our analysis of the neurobiological studies suggests that the most striking problem for disabled readers is their failure in orthographic learning, associated with a disability to train the left VOT system (including VWFA). One could argue that the underlying cause of this disability may in fact not be a phonological deficit as such, but rather a disorder in connecting orthographic and phonological information (Paulesu et al., 1996). From both the behavioral studies as the neurobiological studies it transpires that disabled readers predominantly make use of more explicit phonological and semantic processing routes (see also Bergmann & Wimmer, 2008), which is line with the phonological stagnation hypothesis we introduced in the previous section.

Because phonology is a universal and essential constituent of word recognition (see Perfetti, Liu, & Tan, 2005), separating and demonstrating a unique causal role of phonology in the reading process

proves extremely difficult. A wealth of evidence shows that reading-disabled children have an impairment in representing, storing, and retrieving phonological information (for reviews, see Ramus, 2003; Vellutino et al., 2004). Yet, also questions have been reported about an unequivocal causal role of phonological awareness as one domain of phonological skills (Castles & Coltheart, 2004; Hulme, Snowling, Caravolas, & Carroll, 2005). Based on a detailed analysis of the performance of dyslexic adolescents on a range of phonological tasks, Ramus and Szenkovits (2008) came to the conclusion that the phonological representations of people with dyslexia are normal, but that the phonological problems they experience are only demonstrable when task requirements enhance the processing load in terms of short-term memory, conscious awareness, or time constraints. These authors hypothesize that the basic problem of individuals with dyslexia is a deficit in accessing the phonological representations, not the nature of these representations. Based on the evidence discussed previously, we would speculate that this difficulty of disabled readers in accessing the phonological representations may present a downstream consequence of their fundamental orthographic-phonological disconnection disorder. Learning to read in an alphabetic orthography results in an everlasting re-description of phonological representations (Alegria & Morais, 1991; Morais, Cary, Algeria, & Bertelson, 1979), facilitating the access of these representations.

Another fundamental problem the phonological deficit hypothesis is faced with is why the phonological processing deficit fails to manifest itself, or is much less striking outside the domain of the reading process. One would indeed expect that the extremely complex phonological processes in speech perception and production are even more vulnerable to a phonological deficit than the phonological processes involved in reading relatively simple words at the start of reading development. One may wonder how it is possible that children who prove to be experts in speech phonology (by producing and understanding the most complex utterances), and sometimes even have an above average verbal–phonological development, still demonstrate a phonological deficit that would explain their poor reading? Although there have been attempts to answer this question (see Fowler, 1991; Liberman et al., 1977; Metsala & Walley, 1998), the subtlety of the phonological disorder is an issue that deserves further attention.

Finally, as the existence of a nonword-reading deficit in disabled readers has been considered as one of the cornerstones of the phonological deficit hypothesis, our finding of the absence of a genuine nonword-reading deficit additionally urges for an explanation. Alternatively, the specific disability to acquire word-specific knowledge that we observed in disabled readers can be taken as a direct consequence of the disconnection disorder between orthographic and phonological processing areas in the brain. From this line of reasoning, we would expect that disabled readers also exhibit a specific problem in processing orthographic knowledge at the subword level, a hypothesis that was not tested in our study.

In our view, to maintain as a viable and productive causal theory, the phonological deficit hypothesis has to be analyzed and adjusted in such way that it can envisage the fundamental problem of reading disability as a genuine 'learning' disturbance in terms of an orthographic-phonological binding deficit. Although such disconnection hypothesis could fit with both theoretical and empirical evidence, it remains a challenge to identify the precise systemic characteristics of neuronal functioning that cause this specific disconnection.

7. Conclusions

In this paper, using state trace analysis as an improved matching or equivalence technique, we obtained compelling empirical evidence to conclude that Dutch-speaking reading disabled children do not demonstrate a genuine nonword reading deficit relative to their word reading performance. As a second conclusion, based on an 'inverse' state trace analysis, matching disabled and typical readers on decoding ability, disabled readers were shown to exhibit a clear deficit in real word reading. Both results are taken to imply that the fundamental problem of disabled readers is not a specific problem in reading nonwords, but a difficulty in acquiring orthographic (word-specific) knowledge. We interpreted these results in line with the recently formulated hypothesis that disabled readers show a

disorder in orthographic-phonolocal binding. Future cross-linguistic research, using state trace analysis, should clarify whether these conclusions still hold in other orthographies, especially in inconsistent orthographies like English.

Acknowledgments

Appendix A

Using a simple regression equation, the mean for each group on the nonword reading test can be written as

$$\mu_{YN} = \rho_{t1} \frac{\sigma_{Yt1}}{\sigma_{Xt1}} (\mu_{XD} - \mu_{Xt1}) + \mu_{Yt1} \tag{A.1}$$

$$\mu_{YD} = \rho_{t2} \frac{\sigma_{Yt2}}{\sigma_{Xt2}} (\mu_{XD} - \mu_{Xt2}) + \mu_{Yt2} \tag{A.2}$$

where μ_{YN} and μ_{YD} are the predicted nonword reading scores of the nondisabled readers and the disabled readers, respectively; ρ_{t1} and ρ_{t2} are the correlations between word reading and nonword reading at the age of the nondisabled readers and the disabled readers, respectively; σ_{Yt1} and σ_{Yt2} are the standard deviations of the nonword reading test for the two ages, respectively; μ_{Xt1} and μ_{Xt2} are the standard deviations of the word reading test for the two ages, respectively; μ_{XD} is the mean word reading score of the disabled readers; μ_{Xt1} and μ_{Xt2} are the means of the word reading test for the two ages, respectively; μ_{Yt1} and μ_{Yt2} are the means of the nonword reading test for the two ages, respectively.

Thus, a nonword-reading deficit is present when μ_{YN} is larger than μ_{YD}. After substitution of $\sigma_{Yt1}/\sigma_{Xt1}$ by κ and $\sigma_{Yt2}/\sigma_{Xt2}$ by λ, and after reorganizing the terms in (A.1) and (A.2), the formula expressing the condition when a nonword-reading deficit exists, can be written as

$$\rho_{t2}.\lambda(\mu_{Xt2} - \mu_{XD}) + \rho_{t1}.\kappa(\mu_{XD} - \mu_{Xt1}) > \mu_{Yt2} - \mu_{Yt1} \tag{A.3}$$

As can be seen from the formula, when the mean word-score of the matched control group (μ_{XD}) exactly equals the mean word-score of that age group (μ_{Xt1}), the second term of the inequality is zero. In that case a nonword-reading deficit is present when the adjusted difference between the mean word reading score of the disabled readers (μ_{XD}) and the average word reading score of all same-aged readers (μ_{Xt2}) is larger than the difference between the mean nonword reading scores of all readers at the two ages (see Fig. 2a and b). In general, when the correlations equal the inverse ratio of the standard deviations ($1/\lambda$ and $1/\kappa$), the slope of the word reading variable has to be steeper than the slope of the nonword reading variable in order to obtain a nonword-reading deficit (see Fig. 2a).

Appendix B

When we calculate the magnitude of the nonword-reading deficit by transferring the term at the right side of inequality (A.3) to the left side (see formula (B.1)), and divide this magnitude by the slope of the nonword reading curve (t_1 is the age of the nondisabled readers and t_2 is the age of the disabled readers), we arrive at the nonword reading delay expressed in number of years (formula (B.2)).

$$NRD = \rho_{t2}.\lambda(\mu_{Xt2} - \mu_{XD}) + \rho_{t1}.\kappa(\mu_{XD} - \mu_{Xt1}) - (\mu_{Yt2} - \mu_{Yt1}) \tag{B.1}$$

$$delay = \frac{NRD}{\left(\frac{\mu_{Yt2} - \mu_{Yt1}}{t_2 - t_1}\right)} \tag{B.2}$$

Appendix C

Word Items

1 Syllable	PR BF N	2 Syllables	PR BF N	3 Syllables	PR BF N
No consonant cluster					
High body frequency					
Bal		Gitaar		Oranje	
Mat	22,525	Kanjer	8659	Aandelen	4342
Bel	16,220	Open	48,682	Familie	82,786
Tas	23.5	Huizen	3	Iedereen	1.25
Low body frequency					
Nooit		Lawaai		Babbelbox	
Juf	10,085	Laadruim	2500	Zomereik	2292
Poes	11,979	Cognac	25,930	Radio	66,584
Web	11	Bosduif	0	Dierenbeul	0.5
One consonant cluster					
High body frequency					
Vlot		Wereld		Kwebbelen	
Stem	7744	Twinkel	4837	Tandpasta	2281
Strak	29,658	Monster	72,287	Verveling	93,777
Snel	8.25	Iemand	0.5	Zeshonderd	1
Low body frequency					
Kalk		Gebluf		Mopperkont	
Sneeuw	5896	Beenmerg	3060	Boekenwurm	1756
Vork	16,656	Baljurk	42,114	Labyrint	63,085
Krab	7.5	Viswalm	0.25	Uitgedund	0.25
Two consonant clusters					
High body frequency					
Spons		Woordspel		Baanbrekend	
Start	3724	Spelling	2458	Droomwereld	1892
Kring	38,029	Knotsgek	60,824	Drenkeling	11,471
Brand	2.25	Halssnoer	0.5	Straktrekken	0
Low body frequency					
Schurk		Wenkbrauw		Tegendraads	
Streng	3711	Krijtrots	2363	Navelstreng	1620
Dwerg	49,359	Handpalm	46,247	Nagedreund	921,330
Klont	1	Geflirt	0	Uitgaansbuurt	

Nonword items

1 Syllable	PR BF N	2 Syllables	PR BF N	3 Syllables	PR BF N
No consonant cluster					
High body frequency					
Nag		Hizaar		Onarle	
Zan	22,835	Bunzer	7802	Oefbezen	4292
Kek	11,712	Owet	38,080	Nagidie	58,139
Nas	15	Buiver	1.75	Iegenees	0
Low body frequency					
Roois		Guvaai		Kaspeltux	
Gup	9375	Naakbuim	2380	Jinedeig	2272
Joes	13,208	Folwac	27,499	Jagio	46,610
Zeb	8.25	Tinluif	0	Ziegelkeul	0
One consonant cluster					
High body frequency					
Dror		Heberd		Fluddelen	
Breg	7333	Twongel	4430	Warplubka	2008
Stris	27,182	Kelstar	58,318	Herbalist	78,348
Gral	6	Ievand	0.5	Bilrasterd	0.25
Low body frequency					
Golk		Bestuf		Tapperdont	
Kreeuw	6156	Geekzorg	2799	Toukenfulm	1712
Dorf	16,209	Golfurk	19,266	Bedozint	56,334
Glab	4.75	Kifkolm	0	Eutgebund	0
Two consonant clusters					
High body frequency					
Sperd		Koostbrel		Zeukgrudend	
Glats	3278	Grolbant	2589	Bluifdekeld	1855
Klang	38,420	Platsler	67,755	Blisgedent	87,065
Krond	2.25	Hepsdroer	0.25	Stroesfrikken	0
Low body frequency					
Splurf		Densvrauw		Bedengruuds	
Schert	3599	Zweumkits	2386	Buzelsperg	1457
Klerg	39,699	Waldzolm	36,717	Schozurfelp	73,672
Spoft	2.25	Bevrirt	0	Uitbeelswuurt	0

Note: PR is summed positional redundancy of the letters in an item, BF is summed bigram frequency, and *N* is the number of orthographic neighbors, averaged for each condition.

Appendix D

Assuming that all three components, onset time (o), articulation time (a) and rest time (r), contain 'central' reading processes (c) as well as 'peripheral' processes (p), we can write the reaction time for each kind of item as the sum of six terms, and after summing the central and the peripheral processing times, as the sum of two terms:

$$RT_{nat} = (o_{cn} + a_{cn} + r_{cn}) + (o_{pn} + a_{pn} + r_{pn}) = c_n + p_n$$

$$RT_{umpalters} = (o_{cu} + a_{cu} + r_{cu}) + (o_{pu} + a_{pu} + r_{pu}) = c_u + p_u$$

$$RT_{bag} = (o_{cb} + a_{cb} + r_{cb}) + (o_{pb} + a_{pb} + r_{pb}) = c_b + p_b$$

$$RT_{important} = (o_{ci} + a_{ci} + r_{ci}) + (o_{pi} + a_{pi} + r_{pi}) = c_i + p_i$$

As a result of the structural match between the word–nonword pairs ('bag–nat' and 'important–umpalters'), processing time due to peripheral processes can be considered as the same for these word–nonword pairs. Thus,

$$p_n = p_b \quad \text{and} \quad p_u = p_i. \tag{D.1}$$

Due to the functional match,

$$c_b + p_b = c_i + p_i \quad \text{thus} \quad p_b - p_i = c_i - c_b \tag{D.2}$$

Now, using elementary algebra, it can be shown that the difference in response time between disabled readers' nonwords ('nat') and typical readers' nonwords ('umpalters') reflects the difference in central processes of nonword pairs after the difference in central processes of word pairs ('bag–important') is taken into account.

$$
\begin{aligned}
RT_{nat} - RT_{umpalters} &= (c_n + p_n) - (c_u + p_u \\
&= (c_n + p_b) - (c_u + p_i) \quad \text{(using D.1)} \\
&= (c_n - c_u) + (p_b - p_i) \\
&= (c_n - c_u) - (c_b - c_i) \quad \text{(using D.2)}
\end{aligned}
$$

References

Alegria, J., & Morais, J. (1991). Segmental analysis and reading acquisition. In L. Rieben & C. A. Perfetti (Eds.), *Learning to read: Basic research and its implications* (pp. 135–148). Hillsdale, NJ: Erlbaum Associates.

Anastasi, A. (1968). *Psychological testing*. New York: Macmillan.

Ashby, J. (2010). Phonology is fundamental in skilled reading: Evidence from ERPs. *Psychological Bulletin & Review, 17*(1), 95–100. http://dx.doi.org/10.3758/PBR.17.1.95.

Backman, J. E., Mamen, M., & Ferguson, H. (1984). Reading-level design: Conceptual and methodological issues in reading research. *Psychological Bulletin, 96*, 560–568. http://dx.doi.org/10.1037//0033-2909.96.3.560.

Balota, D. A., Cortese, M. J., Segernt-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. http://dx.doi.org/10.1037/0096-3445.133.2.283.

Bamber, D. (1979). State-trace analysis: A method of testing simple theories of causation. *Journal of Mathematical Psychology, 19*, 137–181. http://dx.doi.org/10.1016/0022-2496(79)90016-6.

Ben-Shachar, M., Dougherty, R. F., Deutsch, G. K., & Wandell, B. A. (2011). The development of cortical sensitivity to visual word forms. *Journal of Cognitive Neuroscience, 23*(9), 2387–2399. http://dx.doi.org/10.1162/jocn.2011.21615.

Bergmann, J., & Wimmer, H. (2008). A dual-route perspective on poor reading in a regular orthography: Evidence from phonological and orthographic lexical decisions. *Cognitive Neuropsychology, 25*, 653–676. http://dx.doi.org/10.1080/02643290802221404.

Binder, J. R., Medler, D. A., Westbury, C. F., Liebenthal, E., & Buchanan, L. (2006). Tuning of the human left fusiform gyrus to sublexical orthographic structure. *Neuroimage, 33*(2), 739–748. http://dx.doi.org/10.1016/j.neuroimage.2006.06.053.

Blomert, L. (2011). The neural signature of orthographic–phonological binding in successful and failing reading development. *Neuroimage, 57*, 695–703. http://dx.doi.org/10.1016/j.neuroimage.2010.11.003.

Bogartz, R. S. (1976). On the meaning of statistical interactions. *Journal of Experimental Child Psychology, 22*, 178–183. http://dx.doi.org/10.1016/0022-0965(76)90099-0.

Bowey, J. A., & Muller, D. (2005). Phonological recoding and rapid orthographic learning in t hird-graders' silent reading: A critical test of the self-teaching hypothesis. *Journal of Experimental Child Psychology, 92*, 203–219. http://dx.doi.org/10.1016/j.jecp. 2005.06.005.

Braams, T. (2002). De zin van onzinwoorden: het gebruik van pseudowoorden bij de signalering, de diagnostiek en de behandeling van dyslexie (The meaning of nonsense words: The use of pseudowords in screening, diagnosis and treatment of dyslexia). *Tijdschrift voor Remedial Teaching, 10*(2), 4–9.

Bracken, B. A. (1988). Ten psychometric reasons why similar tests produce dissimilar results. *Journal of School Psychology, 26*, 155–166. http://dx.doi.org/10.1016/0022-4405(88)90017-9.

Braet, W., Wagemans, J., & Op de Beeck, H. P. (2012). The visual word form area is organized according to orthography. *Neuroimage, 59*(3), 2751–2759. http://dx.doi.org/10.1016/j.neuroimage.2011.10.032.

Brem, S., Bach, S., Kucian, K., Guttorm, T. K., Martin, E., Lyytinen, H., et al (2010). Brain sensitivity to print emerges when children learn letter–speech sound correspondences. *Proceedings of the National Academy of Sciences, 107*(17), 7939–7944. http://dx.doi.org/10.1073/pnas.0904402107.

Brem, S., Bucher, K., Halder, P., Summers, P., Dietrich, T., Martin, E., et al (2006). Evidence for developmental changes in the visual word processing network beyond adolescence. *Neuroimage, 29*, 822–837. http://dx.doi.org/10.1016/j.neuroimage.2005.09.023.

Brown, T. T., Lugar, H. M., Coalson, R. S., Miezin, F. M., Petersen, S. E., & Schlaggar, B. L. (2005). Developmental changes in human cerebral functional organization for word generation. *Cerebral Cortex, 15*, 275–290. http://dx.doi.org/10.1093/cercor/bhh129.

Bruno, J. L., Zumberge, A., Manis, F. R., Lu, Z., & Goldman, J. G. (2008). Sensitivity to orthographic familiarity in the occipito-temporal region. *Neuroimage, 39*, 1988–2001. http://dx.doi.org/10.1016/j.neuroimage.2007.10.044.

Brus, B. T., & Voeten, M. J. M. (1973). *Eén minuut-test*. Nijmegen: Berkhout.

Bryant, P. E., & Goswami, U. (1986). Strengths and weaknesses of the reading level design: A comment on Backman, Mamen, and Ferguson. *Psychological Bulletin, 100*, 101–103. http://dx.doi.org/10.1037//0033-2909.100.1.101.

Busey, T. A., Tunnicliff, J., Loftus, G. R., & Loftus, E. F. (2000). Accounts of the confidence–accuracy relation in recognition memory. *Psychonomic Bulletin & Review, 7*, 26–48. http://dx.doi.org/10.3758/BF03210724.

Cai, Q., Paulignan, Y., Brysbaert, M., Ibarrola, D., & Nazir, T. A. (2010). The left ventral occipito-temporal response to words depends on language lateralization but not on visual familiarity. *Cerebral Cortex, 20*(5), 1153–1163. http://dx.doi.org/10.1093/cercor/bhp175.

Carr, T. H., & Pollatsek, A. (1985). Recognizing printed words: A look at current theories. In D. Besner, T. G. Waller, & G. E. MacKinnon (Eds.), *Reading research: Advances in theory and practice*. Orlando: Academic Press.

Castles, A., & Coltheart, M. (1993). Varieties of developmental dyslexia. *Cognition, 47*, 149–180. http://dx.doi.org/10.1016/0010-0277(93)90003-E.

Castles, A., & Coltheart, M. (1996). Cognitive correlates of developmental surface dyslexia: A single case study. *Cognitive Neuropsychology, 13*, 25–50. http://dx.doi.org/10.1080/026432996382051.

Castles, A., & Coltheart, M. (2004). Is there a causal link from phonological awareness to success in learning to read? *Cognition, 91*, 77–111. http://dx.doi.org/10.1016/S0010-0277(03)00164-1.

Castles, A., & Nation, K. (2006). How does orthographic learning happen? In S. Andrews (Ed.), *From inkmarks to ideas: Challenges and controversies about word recognition and reading* (pp. 151–179). London: Psychology Press.

Cohen, L., Dehaene, S., Naccache, L., Lehéricy, S., Dehaene-Lambertz, G., Hénaff, M., et al (2000). The visual word form area: Spatial and temporal characterization of an initial stage of reading in normal subjects and posterior split-brain patients. *Brain, 123*, 291–307. http://dx.doi.org/10.1093/brain/123.2.291.

Cohen, L., Dehaene, S., Vinckier, F., Jobert, A., & Montavont, A. (2008). Reading normal and degraded words: Contribution of the dorsal and ventral visual pathways. *NeuroImage, 40*, 353–366. http://dx.doi.org/10.1016/j.neuroimage.2007.11.036.

Coltheart, M. (2006). What has functional neuroimaging told us about the mind (so far)? *Cortex, 42*, 323–331. http://dx.doi.org/10.1016/S0010-9452(08)70358-7.

Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. (2001). DRC: A dual route cascaded model of visual word recognition and reading aloud. *Psychological Review, 108*(1), 204–256. http://dx.doi.org/10.1037//0033-295X.108.1.204.

Coltheart, M. (1978). Lexical access in simple reading tasks. In G. Underwood (Ed.), *Strategies of information processing*. New York: Academic Press.

De Brauwer, J., Verguts, T., & Fias, W. (2006). The representation of multiplication facts: Developmental changes in the problem size, five, and tie effects. *Journal of Experimental Child Psychology, 94*, 43–56. http://dx.doi.org/10.1016/j.jecp.2005.11.004.

de Jong, P. F., & Share, D. L. (2007). Orthographic learning during oral and silent reading. *Scientific Studies of Reading, 11*, 55–71. http://dx.doi.org/10.1207/s1532799xssr1101_4.

Dehaene, S., & Cohen, L. (2007). Cultural recycling of cortical maps. *Neuron, 56*, 384–398. http://dx.doi.org/10.1016/j.neuron.2007.10.004.

Dehaene, S., & Cohen, L. (2011). The unique role of the visual word form area in reading. *Trends in Cognitive Sciences, 15*(6), 254–262. http://dx.doi.org/10.1016/j.tics.2011.04.003.

Dehaene, S., Cohen, L., Sigman, M., & Vinckier, F. (2005). The neural code for written words: A proposal. *Trends in Cognitive Sciences, 9*(7), 335–341. http://dx.doi.org/10.1016/j.tics.2005.05.004.

Dehaene, S., Le Clec, H. G., Poline, J. B., Le Bihan, D., & Cohen, L. (2002). The visual word form area: A prelexical representation of visual words in the fusiform gyrus. *NeuroReport, 13*(3), 321–325. http://dx.doi.org/10.1097/00001756-200203040-00015.

Dehaene, S., Pegado, F., Braga, L. W., Ventura, P., Filho, G. N., Jobert, A., et al (2010). How learning to read changes the cortical networks for vision and language. *Science, 330*, 1359–1364. http://dx.doi.org/10.1126/science.1194140.

Desroches, A. S., Cone, N. E., Bolger, D. J., Bitan, T., Burman, D., & Booth, J. R. (2010). Children with reading difficulties show differences in brain regions associated with orthographic processing during spoken language processing. *Brain Research, 1356*, 73–84. http://dx.doi.org/10.1016/j.brainres.2010.07.097.

Di Filippo, G., 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*(2), 141–149. http://dx.doi.org/10.1080/09297040500346571.

Diependaele, K., Ziegler, J., & Grainger, J. (2010). Fast phonology and the bi-modal interactive activation model. *European Journal of Cognitive Psychology, 22*, 764–778. http://dx.doi.org/10.1080/09541440902834782.

Dunn, J. C., & James, R. N. (2003). Signed difference analysis: Theory and application. *Journal of Mathematical Psychology, 47*, 389–416. http://dx.doi.org/10.1016/S0022-2496(03)00049-X.

Eden, G. F., Jones, K. M., Cappell, K., Gareau, L., Wood, F. B., Zeffiro, T. A., et al (2004). Neural changes following remediation in adult developmental dyslexia. *Neuron, 44*, 411–422. http://dx.doi.org/10.1016/j.neuron.2004.10.019.

Ehri (1992). Reconceptualising sight word reading. In P. B. Gough, L. C. Ehri, & R. Treiman (Eds.), *Reading acquisition* (pp. 107–143). Hillsdale, NJ: Lawrence Erlbaum.

Ehri, L. C., & Saltmarsh, J. (1995). Beginning readers outperform older disabled readers in learning to read words by sight. *Reading and Writing: An Interdisciplinary Journal, 7*, 295–326. http://dx.doi.org/10.1007/BF03162082.

Elektronisch Groene Boekje (1996). *Sdu Uitgevers den Haag.* Rotterdam: AND Electronic Publishing BV.

Evers, A., & Resing, W. (2007). Het drijfzand van didactische leeftijdsequivalenten (The quicksand of grade equivalent scores). *De Psycholoog, 42*, 466–472.

Faust, M. E., Balota, D. A., Spieler, D. H., & Ferraro, F. R. (1999). Individual differences in information-processing rate and amount: Implications for group differences in response latency. *Psychological Bulletin, 125*, 777–799. http://dx.doi.org/10.1037//0033-2909.125.6.777.

Forster, K. I. (1976). Accessing the mental lexicon. In R. J. Wales & E. Walker (Eds.), *New approaches to language mechanisms* (pp. 57–287). Amsterdam: North Holland.

Fowler, A. (1991). How early phonological development might set the stage for phoneme awareness. In S. A. Brady & D. P. Shankweiler (Eds.), *Phonological processes in literacy* (pp. 97–117). Hillsdale, NJ: Erlbaum.

Frost, R., Katz, L., & Bentin, S. (1987). Strategies for visual word recognition and orthographical depth: A multilingual comparison. *Journal of Experimental Psychology: Human Perception and Performance, 13*, 104–115. http://dx.doi.org/10.1037//0096-1523.13.1.104.

Frost, S. J., Sandak, R., Mencl, W. E., Landi, N., Rueckl, J. G., Katz, L., et al (2009). Mapping the word reading circuitry in skilled and disabled readers. In P. McCardle & K. Pugh (Eds.), *Helping children learn to read: Current issues and new directions in the integration of cognition. Neurobiology and genetics of reading and dyslexia research and practice* (pp. 3–19). Psychology Press.

Froyen, D., Willems, G., & Blomert, L. (in press). Evidence for a specific cross-modal binding deficit in dyslexia: An MMN-study of letter–speech sound processing. *Developmental Science.* http://dx.doi.org/10.1111/j.1467-7687.2010.01007.x.

Geudens, A., Sandra, D., & Van den Broeck, W. (2004). Segmenting two-phoneme syllables: Developmental differences in relation to early reading skills. *Brain and Language, 90*, 338–352. http://dx.doi.org/10.1016/S0093-934X(03)00446-2.

Geudens, A., & Sandra, D. (2002). The role of orthographic onset-rime units in Dutch beginning readers. In L. Verhoeven, P. Reitsma, & C. Elbro (Eds.), *Precursors of functional literacy* (pp. 49–67). Amsterdam: John Benjamins.

Glezer, L. S., Jiang, X., & Riesenhuber, M. (2009). Evidence for highly selective neuronal tuning to whole words in the "visual word form area". *Neuron, 62*, 199–204. http://dx.doi.org/10.1016/j.neuron.2009.03.017.

Goswami, U., & Bryant, P. E. (1989). The interpretation of studies using the reading level design. *Journal of Reading Behavior, 21*, 413–424. http://dx.doi.org/10.1080/10862968909547687.

Grainger, J., & Jacobs, A. M. (1996). Orthographic processing in visual word recognition: A multiple read-out model. *Psychological Review, 103*, 518–565. http://dx.doi.org/10.1037/0033-295X.103.3.518.

Griffithts, Y. M., & Snowling, M. J. (2002). Predictors of exception word and nonword reading in dyslexic children: The severity hypothesis. *Journal of Educational Psychology, 94*(1), 34–43. http://dx.doi.org/10.1037//0022-0663.94.1.34.

Grossberg, S. (1987). Competitive learning: From interactive activation to adaptive resonance. *Cognitive Science, 11*, 23–63. http://dx.doi.org/10.1016/S0364-0213(87)80025-3.

Gulliksen, H. (1950). *Theory of mental tests.* New York: Wiley.

Harm, M. W., & Seidenberg, M. S. (1999). Phonology, reading acquisition, and dyslexia: Insights from connectionist models. *Psychological Review, 106*, 491–528. http://dx.doi.org/10.1037//0033-295X.106.3.491.

Harm, M., & Seidenberg, M. S. (2004). Computing the meanings of words in reading: cooperative division of labor between visual and phonological processes. *Psychological Review, 111*, 662–720. http://dx.doi.org/10.1037/0033-295X.111.3.662.

Hashimoto, R., & Sakai, K. L. (2004). Learning letters in adulthood: Direct visualization of cortical plasticity for forming a new link between orthography and phonology. *Neuron, 42*, 311–322. http://dx.doi.org/10.1016/S0896-6273(04)00196-5.

Henson, R. (2005). What can functional neuroimaging tell the experimental psychologist? *Quarterly Journal of Experimental Psychology, 58A*, 193–233. http://dx.doi.org/10.1080/02724980443000502.

Herrmann, J. A., Matyas, T., & Pratt, C. (2006). Meta-analysis of the nonword-reading deficit in specific reading disorder. *Dyslexia, 12*, 195–221. http://dx.doi.org/10.1002/dys.324.

Hilbe, J. M. (2007). *Negative binomial regression.* Cambridge, UK: Cambridge University Press.

Hoeft, F., Hernandez, A., McMillon, G., Taylor-Hill, H., Martindale, J. L., Meyler, A., et al (2006). Neural basis of dyslexia: A comparison between dyslexic and nondyslexic children equated for reading ability. *The Journal of Neuroscience, 26*(42), 10700–10708. http://dx.doi.org/10.1523/JNEUROSCI.4931-05.2006.

Hoeft, F., Meyler, A., Hernandez, A., Juel, C., Taylor-Hill, H., Martindale, J. L., et al (2007). Functional and morphometric brain dissociation between dyslexia and reading ability. *Proceedings of the National Academy of Sciences, 104*(10), 4234–4239. http://dx.doi.org/10.1073/pnas.0609399104.

Hulme, C., Snowling, M., Caravolas, M., & Carroll, J. (2005). Phonological skills are (probably) one cause of success in learning to read: A comment on Castles and Coltheart. *Scientific Studies of Reading, 9*, 351–365. http://dx.doi.org/10.1207/s1532799xssr0904_2.

Jackson, N. E., & Butterfield, E. C. (1989). Reading-level-matched designs: Myths and realities. *Journal of Reading Behavior, 21*, 387–412. http://dx.doi.org/10.1080/10862968909547686.

Jackson, N. E., & Coltheart, M. (2001). *Routes to reading success and failure. Toward an integrated cognitive psychology of atypical reading. Macquarie Monographs in Cognitive Science.* Taylor & Francis.

Jorm, A., & Share, D. (1983). Phonological recoding and reading acquisition. *Applied Psycholinguistics, 4*, 103–147. http://dx.doi.org/10.1017/S0142716400004380.

Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association, 90*(430), 773–795. http://dx.doi.org/10.2307/2291091.

Keats, J. A. (1983). Ability measures and theories of cognitive development. In H. Wainer & S. Messick (Eds.), *Principles of modern psychological measurement: A festschrift for Frederic M. Lord* (pp. 81–101). Hillsdale, NJ: Lawrence Erlbaum.

Keller, T. A., & Just, M. A. (2009). Altering cortical connectivity: Remediation-induced changes in the white matter of poor readers. *Neuron, 64*(5), 624–631. http://dx.doi.org/10.1016/j.neuron.2009.10.018.

Kherif, H., Josse, G., & Price, C. J. (2011). Automatic top-down processing explains common left occipito-temporal responses to visual words and objects. *Cerebral Cortex, 21*, 103–114. http://dx.doi.org/10.1093/cercor/bhq063.

Kliegl, R., Mayr, U., & Krampe, R. T. (1994). Time-accuracy functions for determining process and person differences: An application to cognitive aging. *Cognitive Psychology, 26*, 134–164. http://dx.doi.org/10.1006/cogp. 1994.1005.

Kronbichler, M., Bergmann, J., Hutzler, F., Staffen, W., Mair, A., Ladurner, G., et al (2007). Taxi vs. Taksi: On orthographic word recognition in the left ventral occipitotemporal cortex. *Journal of Cognitive Neuroscience, 19*(10), 1584–1594. http://dx.doi.org/10.1162/jocn.2007.19.10.1584.

Kronbichler, M., Hutzler, F., Wimmer, H., Mair, A., Staffen, W., & Ladurner, G. (2004). The visual word form area and the frequency with which words are encountered: Evidence from a parametric fMRI study. *Neuroimage, 21*, 946–953. http://dx.doi.org/10.1016/j.neuroimage.2003.10.021.

Lachmann, T., & van Leeuwen, C. (2008). Different letter-processing strategies in diagnostic subgroups of developmental dyslexia. *Cognitive Neuropsychology, 25*, 730–744. http://dx.doi.org/10.1080/02643290802309514.

Liberman, A. M. (1992). The relation of speech to reading and writing. In R. Frost & L. Katz (Eds.), *Orthography, phonology, morphology, and meaning* (pp. 167–178). Amsterdam: Elsevier.

Liberman, I. Y., Shankweiler, D., Liberman, A. M., Fowler, C., & Fischer, F. (1977). Phonetic segmentation and recoding in the beginning reader. In A. Reber & D. Scarborough (Eds.), *Toward a psychology of reading: The proceedings of the CUNY conference* (pp. 207–227). Hillsdale, NJ: Lawrence Erlbaum.

Loftus, G. R. (1978). On interpretation of interactions. *Memory & Cognition, 6*, 312–319. http://dx.doi.org/10.3758/BF03197461.

Loftus, G. R. (1985). Evaluating forgetting curves. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 11*, 396–405. http://dx.doi.org/10.1037//0278-7393.11.2.397.

Loftus, G. R., & Irwin, D. E. (1998). On the relations among different measures of visible and informational persistence. *Cognitive Psychology, 35*, 135–199. http://dx.doi.org/10.1006/cogp. 1998.0678.

Loftus, G. R., Oberg, M. A., & Dillon, A. M. (2004). Linear theory, dimensional theory, and the face-inversion effect. *Psychological Review, 111*(4), 835–863. http://dx.doi.org/10.1037/0033-295X.111.4.835.

Loftus, G. R., Truax, P. E., & Nelson, W. W. (1987). Age-related differences in visual information processing: Qualitative or quantitative? In C. Schooler & K. W. Schaie (Eds.), *Cognitive functioning and social structure over the life course* (pp. 59–77). Norwood, NJ: Ablex.

Long, J. S. (1997). *Regression models for categorical and limited dependent variables*. Thousand Oaks, California: Sage Publications, Inc..

Lyon, G. R., Shaywitz, S. E., & Shaywitz, B. A. (2003). A definition of dyslexia. *Annals of Dyslexia, 53*, 1–14. http://dx.doi.org/10.1007/s11881-003-0001-9.

Manis, F. R., Seidenberg, M. S., Doi, M. L., McBride-Chang, C., & Petersen, A. (1996). On the bases of two subtypes of developmental dyslexia. *Cognition, 58*, 157–195. http://dx.doi.org/10.1016/0010-0277(95)00679-6.

Marinus, E., & de Jong, P. F. (2010). Size does not matter, frequency does: Sensitivity to orthographic neighbors in normal and dyslexic readers. *Journal of Experimental Child Psychology, 106*, 129–144. http://dx.doi.org/10.1016/j.jecp. 2010.01.004.

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. http://dx.doi.org/10.1016/j.bandl.2006.04.003.

Martensen, H. E., Maris, E., & Dijkstra, A. F. J. (2000). When does inconsistency hurt? On the relation between phonological-consistency effects and the reliability of sublexical units. *Memory and Cognition, 28*(4), 648–656. http://dx.doi.org/10.1016/j.bandl.2006.04.003.

Maurer, U., Brem, S., Bucher, K., & Brandeis, D. (2005). Emerging neurophysiological specialization for letter strings. *Journal of Cognitive Neuroscience, 17*(10), 1532–1552. http://dx.doi.org/10.1162/089892905774597218.

Maurer, U., Brem, S., Kranz, F., Bucher, K., Benz, R., Halder, P., et al (2006). Coarse neural tuning for print peaks when children learn to read. *Neuroimage, 33*(2), 749–758. http://dx.doi.org/10.1016/j.neuroimage.2006.06.025.

McGuinness, D. (1997). *Why our children can't read. And what we can do about it*. New York: The Free Press.

McKague, M. (2005). *The development of orthographic representations in skilled readers*. Unpublished Doctoral Dissertation. The University of Melbourne, Victoria, Australia.

McKague, M., Davis, C., Pratt, C., & Johnston, M. B. (2008). The role of feedback from phonology to orthography in orthographic learning: An extension of item-based accounts. *Journal of Research in Reading, 31*(1), 55–76. http://dx.doi.org/10.1111/j.1467- 9817.2007.00361.x.

Metsala, J. L., & Walley, A. C. (1998). Spoken vocabulary growth and the segmental restructuring of lexical representations: Precursors to phonemic awareness and early reading ability. In J. L. Metsala & L. C. Ehri (Eds.), *Word recognition in beginning literacy* (pp. 89–120). NJ: Lawrence Erlbaum.

Morais, J., Cary, L., Algeria, J., & Bertelson, P. (1979). Does awareness of speech as a sequence of phones arise spontaneously? *Cognition, 7*, 323–331. http://dx.doi.org/10.1016/0010-0277(79)90020-9.

Morton, J. (1969). Interaction of information in word recognition. *Psychological Review, 76*, 165–178. http://dx.doi.org/10.1037/h0027366.

Murphy, L., & Pollatsek, A. (1994). Developmental dyslexia: Heterogeneity without discrete subgroups. *Annals of Dyslexia, 44*, 120–146. http://dx.doi.org/10.1007/BF02648158.

Muthén, L. K., & Muthén, B. O. (1998/2010). *MPlus user's guide* (sixth ed.). Los Angeles, CA: Muthén & Muthén.

Nation, K., & Cocksey, J. (2009). Beginning readers activate semantics from sub-word orthography. *Cognition, 110*, 273–278. http://dx.doi.org/10.1016/j.cognition.2008.11.004.

Newell, B. R., & Dunn, J. C. (2008). Dimensions in data: testing psychological models using state-trace analysis. *Trends in Cognitive Sciences, 12*(8), 285–290. http://dx.doi.org/10.1016/j.tics.2008.04.009.

Norris, D. G., McQueen, J. M., & Cutler, A. (2000). Merging information in speech recognition: Feedback is never necessary. *Behavioural and Brain Sciences, 23*, 299–370. http://dx.doi.org/10.1017/S0140525X00003241.

Paizi, D., De Luca, M., Zoccolotti, P., & Burani, C. (in press). A comprehensive evaluation of lexical reading in Italian developmental dyslexics. *Journal of Research in Reading.* http://dx.doi.org/10.1111/j.1467-9817.2011.01504.x.

Palmer, E. D., Brown, T. T., Petersen, S. E., & Schlaggar, B. L. (2004). Investigation of the functional neuroanatomy of single word reading and its development. *Scientific Studies of Reading, 8*(3), 203–223. http://dx.doi.org/10.1207/s1532799xssr0803_2.

Patel, T. K., Snowling, M. J., & De Jong, P. F. (2004). Learning to read in Dutch and English: A cross-linguistic comparison. *Journal of Educational Psychology, 96*, 785–797. http://dx.doi.org/10.1037/0022-0663.96.4.785.

Paulesu, E., Demonet, J. F., Fazio, F., McCrory, E., Chanoine, V., Bruswick, N., et al (2001). Dyslexia: Cultural diversity and biological unity. *Science, 291*, 2165–2167. http://dx.doi.org/10.1126/science.1057179.

Paulesu, E., Frith, U., Snowling, M., Gallagher, A., Morton, J., Frackowiac, R. S. J., et al (1996). Is developmental dyslexia a disconnection syndrome? Evidence from PET scanning. *Brain, 119*, 143–157. http://dx.doi.org/10.1093/brain/119.1.143.

Pecher, D. (2001). Perception is a two-way junction: Feedback semantics in word recognition. *Psychonomic Bulletin & Review, 8*(3), 545–551. http://dx.doi.org/10.3758/BF03196190.

Peereman, R., Content, A., & Bonin, P. (1998). Is perception a two-way street? The case of feedback consistency in visual word recognition. *Journal of Memory and Language, 39*, 151–174. http://dx.doi.org/10.1006/jmla.1998.2573.

Perfetti, C. A. (1985). *Reading ability*. New York: Oxford University press.

Perfetti, C. A. (1992). The representation problem in reading acquisition. In P. Gough (Ed.), *Reading Acquisition* (pp. 145–174). Hillsdale, NJ: Erlbaum.

Perfetti, C. A., Liu, Y., & Tan, L. H. (2005). The lexical constituency model: Some implications of research on Chinese for general theories of reading. *Psychological Review, 112*(1), 43–59. http://dx.doi.org/10.1037/0033-295X.112.1.43.

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*(2), 273–315. http://dx.doi.org/10.1037/0033-295X.114.2.273.

Perry, C., Ziegler, J. C., & Zorzi, M. (2010). Beyond single syllables: Large-scale modeling of reading aloud with the connectionist dual process (CDP++) model. *Cognitive Psychology, 61*, 106–151. http://dx.doi.org/10.1016/j.cogpsych.2010.04.001.

Pexman, P. M., Lupker, S. J., & Reggin, L. D. (2002). Phonological effects in visual word recognition: Investigating the impact of feedback activation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*, 572–584. http://dx.doi.org/10.1037//0278-7393.28.3.572.

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. http://dx.doi.org/10.1037//0033-295X.103.1.56.

Poldrack, R. A. (2006). Can cognitive processes be inferred from neuroimaging data? *Trends in Cognitive Sciences, 10*(2), 59–63. http://dx.doi.org/10.1016/j.tics.2005.12.004.

Poldrack, R. A. (2008). The role of fMRI in cognitive neuroscience: Where do we stand? *Current Opinion in Neurobiology, 18*, 223–227. http://dx.doi.org/10.1016/j.conb.2008.07.006.

Poldrack, R. A., Wagner, A. D., Prull, M. W., Desmond, J. E., Golver, G. H., & Gabrieli, J. D. (1999). Functional specialization for semantic and phonological processing in the left inferior prefrontal cortex. *Neuroimage, 10*, 15–35. http://dx.doi.org/10.1006/nimg.1999.0441.

Price, C. J., & Devlin, J. T. (2011). The interactive account of ventral occipitotemporal contributions to reading. *Trends in Cognitive Sciences, 15*(6), 246–253. http://dx.doi.org/10.1016/j.tics.2011.04.001.

Price, C. J., Winterburn, D., Giraud, A. L., Moore, C. J., & Noppeney, U. (2003). Cortical localization of the visual and auditory word form areas: A reconsideration of the evidence. *Brain and Language, 86*, 272–286. http://dx.doi.org/10.1016/S0093-934X(02)00544-8.

Prince, M., Brown, S., & Heathcote, A. (2012). The design and analysis of state-trace experiments. *Psychological Methods, 17*(1), 78–99. http://dx.doi.org/10.1037/a0025809.

Pugh, K. R., Frost, S. J., Sandak, R., Landi, N., Rueckl, J. G., Constable, R. T., et al (2008). Effects of stimulus difficulty and repetition on printed word identification: An fMRI comparison of nonimpaired and reading-disabled adolescent cohorts. *Journal of Cognitive Neuroscience, 20*, 1146–1160. http://dx.doi.org/10.1162/jocn.2008.20079.

Pugh, K. R., Mencl, W. E., Jenner, A. R., Katz, L., Frost, S. J., Lee, J. R., et al (2000). Functional neuroimaging studies of reading and reading disability (developmental dyslexia). *Mental Retardation & Developmental Disabilities Research Reviews, 6*, 207–213. http://dx.doi.org/10.1002/1098-2779(2000) 6:3<207::AID-MRDD8>3.3.CO;2-G.

Pugh, K. R., Mencl, W. E., Jenner, A. R., Katz, L., Frost, S. J., Lee, J. R., et al (2001). Neurobiological studies of reading and reading disability. *Journal of Communication Disorders, 34*, 479–492. http://dx.doi.org/10.1016/S0021-9924(01)00060-0.

Rack, J. P., Snowling, M. J., & Olson, R. K. (1992). The nonword-reading deficit in developmental dyslexia: A review. *Reading Research Quarterly, 27*, 29–53. http://dx.doi.org/10.2307/747832.

Raftery, A. E. (1996). Bayesian model selection in social research. In P. V. Marsden (Ed.). *Sociological methodology* (Vol. 25, pp. 111–163). Oxford: Basil Blackwell.

Ramus, F. (2003). Developmental dyslexia: Specific phonological deficit or general sensorimotor dysfunction? *Current Opinion in Neurobiology, 13*, 212–218. http://dx.doi.org/10.1016/S0959-4388(03)00035-7.

Ramus, F., & Szenkovits, G. (2008). What phonological deficit? *The Quarterly Journal of Experimental Psychology, 61*(1), 129–141. http://dx.doi.org/10.1080/17470210701508822.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models. Applications and data analysis methods*. Thousand Oaks, California: Sage Publications, Inc..

Reitsma, P. (1983). Word-specific knowledge in beginning reading. *Journal of Research in Reading, 6*, 41–56. http://dx.doi.org/10.1111/j.1467-9817.1983.tb00237.x.

Reynolds, C. R. (1981). The fallacy of "two years below grade level for age" as a diagnostic for reading disorders. *Journal of School Psychology, 19*(4), 350–358. http://dx.doi.org/10.1016/0022-4405(81)90029-7.

Richlan, F., Kronbichler, M., & Wimmer, H. (2009). Functional abnormalities in the dyslexic brain: A quantitative meta-analysis of neuroimaging studies. *Human Brain Mapping, 30*(10), 3299–3308. http://dx.doi.org/10.1002/hbm.20752.

Rogosa, D. R., & Willett, J. B. (1985). Understanding correlates of change by modeling individual differences in growth. *Psychometrika, 50*, 203–228. http://dx.doi.org/10.1007/BF02294247.

Salthouse, T. A., & Hedden, T. (2002). Interpreting reaction time measures in between-group comparisons. *Journal of Clinical and Experimental Neuropsychology, 24*(7), 858–872. http://dx.doi.org/10.1076/jcen.24.7.858.8392.

Sandak, R., Mencl, W. E., Frost, S. J., & Pugh, K. (2004). The neurobiological basis of skilled and impaired reading: Recent findings and new directions. *Scientific Studies of Reading, 8*, 273–292. http://dx.doi.org/10.1207/s1532799xssr0803_6.

Schlaggar, B. L., Brown, T. T., Lugar, H. M., Visscher, K. M., Miezin, F. M., & Petersen, S. E. (2002). Functional neuroanatomical differences between adults and school-age children in the processing of single words. *Science, 296*, 1476–1479. http://dx.doi.org/10.1126/science.1069464.

Schlaggar, B. L., & McCandliss, B. D. (2007). Development of neural systems for reading. *Annual Review of Neuroscience, 30*, 475–503. http://dx.doi.org/10.1146/annurev.neuro.28.061604.135645.

Schneider, W., Eschman, A., & Zuccolotto, A. (2002). *E-Prime Version 1.0*. Pittsburgh: Psychology Software Tools, Inc..

Schurz, M., Sturm, D., Richlan, F., Kronbichler, M., Ladurner, G., & Wimmer, H. (2010). A dual-route perspective on brain activation in response to visual words: Evidence for a length by lexicality interaction in the visual word form area (VWFA). *Neuroimage, 49*, 2649–2661. http://dx.doi.org/10.1016/j.neuroimage.2009.10.082.

Seidenberg, M. S., & McClelland, J. L. (1989). A distributed developmental model of word recognition and naming. *Psychological Review, 96*, 523–568. http://dx.doi.org/10.1037//0033-295X.96.4.523.

Seymour, P. H. K., Aro, M., & Erskine, J. M. (2003). Foundation literacy acquisition in European orthographies. *British Journal of Psychology, 94*, 143–174. http://dx.doi.org/10.1348/000712603321661859.

Share, D. L. (1995). Phonological recoding and self-teaching: *sine qua non* of reading acquisition. *Cognition, 55*, 151–218. http://dx.doi.org/10.1016/0010-0277(94)00645-2.

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. http://dx.doi.org/10.1006/jecp. 1998.2481.

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. http://dx.doi.org/10.1016/j.jecp. 2004.01.001.

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. http://dx.doi.org/10.1037/0033-2909.134.4.584.

Share, D. L., & Shalev, C. (2004). Self-teaching in normal and disabled readers. *Reading and Writing, 17*, 1–31. http://dx.doi.org/10.1007/s11145-004-2658-9.

Share, D. L., & Stanovich, K. E. (1995). Cognitive processes in early reading development: Accommodating individual differences into a model of acquisition. *Issues in Education, 1*, 1–57.

Shaywitz, B. A., Shaywitz, S. E., Blachman, B. A., Pugh, K. R., Fulbright, R. K., Skudlarski, P., et al (2004). Development of left occipitotemporal systems for skilled reading in children after a phonologically-based intervention. *Biological Psychiatry, 55*, 926–933. http://dx.doi.org/10.1016/j.biopsych.2003.12.019.

Shaywitz, S. E., Shaywitz, B. A., Fulbright, R. K., Skudlarski, P., Mencl, W. E., Constable, R. T., et al (2002). Disruption of posterior brain systems for reading in children with developmental dyslexia. *Biological Psychiatry, 52*, 101–110. http://dx.doi.org/10.1016/S0006-3223(02)01365-3.

Shaywitz, S. E., Shaywitz, B. A., Pugh, K. R., Fulbright, R. K., Constable, R. T., Mencl, W. E., et al (1998). Functional disruption in the organization of the brain for reading in dyslexia. *Proceedings of the National Academy of Sciences, 95*, 2336–2641.

Shaywitz, B. A., Skudlarski, P., Holahan, J. M., Marchione, K. E., Constable, R. T., Fulbright, R. K., et al (2007). Age-related changes in reading systems of dyslexic children. *Annals of Neurology, 61*(4), 363–370. http://dx.doi.org/10.1002/ana.21093.

Simos, P. G., Breier, J. I., Fletcher, J. M., Bergman, E., & Papanicolaou, A. C. (2000). Cerebral mechanisms involved in word reading in dyslexic children: A magnetic source imaging approach. *Cerebral Cortex, 10*(8), 809–816. http://dx.doi.org/10.1093/cercor/10.8.809.

Simos, P. G., Fletcher, J. M., Bergman, E., Breier, J. I., Foorman, B. R., Castillo, E. M., et al (2002). Dyslexia-specific brain activation profile becomes normal following successful remedial training. *Neurology, 58*, 1203–1213.

Snowling, M. J., Bryant, P. E., & Hulme, C. (1996). Theoretical and methodological pitfalls in making comparisons between developmental and acquired dyslexia: Some comments on A. Castles and M. Coltheart (1993). *Reading and Writing, 8*, 443–451.

Song, Y., Bu, Y., Hu, S., Luo, Y., & Liu, J. (2010). Short-term language experience shapes the plasticity of the visual word form area. *Brain Research, 1316*, 83–91. http://dx.doi.org/10.1016/j.brainres.2009.11.086.

Song, Y., Hu, S., Li, X., Li, W., & Liu, J. (2010). The role of top-down task context in learning to perceive objects. *Journal of Neuroscience, 30*(29), 9869–9876. http://dx.doi.org/10.1523/JNEUROSCI.0140-10.2010.

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*(2), 217–235. http://dx.doi.org/10.1207/s15326942dn2702_2.

Staels, E., & Van den Broeck, W. (submitted for publication). *Orthographic learning and the role of text-to-speech software in Dutch disabled readers*.

Stanley, J. C. (1964). *Measurement in today's school*. Englewood Cliffs.

Stanovich, K. E., & Cunningham, A. E. (1993). Where does knowledge come from? Specific associations between print exposure and information acquisition. *Journal of Educational Psychology, 85*, 211–229. http://dx.doi.org/10.1037/0022-0663.85.2.211.

Stanovich, K. E., & Siegel, L. S. (1994). Phenotypic performance profile of children with reading disabilities: A regression-based test of the phonological-core variable-difference model. *Journal of Educational Psychology, 86*, 24–53. http://dx.doi.org/10.1037/0022-0663.86.1.24.

Stanovich, K. E., Siegel, L. S., & Gottardo, A. (1997). Converging evidence for phonological and surface subtypes of reading disability. *Journal of Educational Psychology, 89*, 114–127. http://dx.doi.org/10.1037//0022-0663.89.1.114.

Stanovich, K. E., & West, R. F. (1989). Exposure to print and orthographic processing. *Reading Research Quarterly, 24*, 402–433. http://dx.doi.org/10.2307/747605.

Stone, G. O., & Van Orden, G. C. (1994). Building a resonance framework for word recognition using design and system principles. *Journal of Experimental Psychology: Human Perception and Performance, 20*, 1248–1268. http://dx.doi.org/10.1037//0096-1523.20.6.1248.

Stone, G. O., Vanhoy, M., & Van Orden, G. C. (1997). Perception is a two-way street: Feedforward and feedback phonology in visual word recognition. *Journal of Memory and Language, 36*, 337–359. http://dx.doi.org/10.1006/jmla.1996.2487.

Stuart, A., & Ord, K. (1994). *Kendall's advanced theory of statistics. Distribution theory* (Vol. I). London: Arnold.

Temple, E., Deutsch, G. K., Poldrack, R. A., Miller, S. L., Tallal, P., Merzenich, M. M., et al (2003). Neural deficits in children with dyslexia ameliorated by behavioral remediation: Evidence from functional MRI. *Proceedings of the National Academy of Sciences, 100*(5), 2860–2865. http://dx.doi.org/10.1073/pnas.0030098100.

Thurstone, L. L. (1926). The mental age concept. *Psychological Review, 33*, 268–278. http://dx.doi.org/10.1037/h0072344.

Treiman, R., Mullennix, J., Bijeljac-Babic, R., & Richmond-Welty, E. D. (1995). The special role of rimes in the description, use, and acquisition of English orthography. *Journal of Experimental Psychology: General, 124*, 107–136. http://dx.doi.org/10.1037//0096-3445.124.2.107.

Turkeltaub, P. E., Gareau, L., Flowers, D. L., Zeffiro, T. A., & Eden, G. F. (2003). Development of neural mechanisms for reading. *Nature Neuroscience, 6*(6), 767–773. http://dx.doi.org/10.1038/nn1065.

Uttal, W. R. (2001). *The new phrenology. The limits of localizing cognitive processes in the brain*. MIT Press.

van den Bos, K. P., & lutje Spelberg, H. C. (1994). Word identification routes and reading disorders. In K. P. van den Bos, L. S. Siegel, D. J. Bakker, & D. L. Share (Eds.), *Current directions in dyslexia research* (pp. 201–219). Lisse: Swets & Zeitlinger.

Van den Broeck (2002). The misconception of the regression-based discrepancy operationalization in the definition and research of learning disabilities. *Journal of Learning Disabilities, 35*(3), 194–204. http://dx.doi.org/10.1177/002221940203500301.

Van den Broeck, W., Geudens, A., & Staels, E. (2011). *A new reliable method to study subtypes in developmental dyslexia*. Paper presented at the annual conference of the Society of the Scientific Studies of Reading. St. Pete Beach, Florida.

Van den Broeck, W., Geudens, A., & van den Bos, K. P. (2010). The nonword reading deficit of disabled readers: A developmental interpretation. *Developmental Psychology, 46*, 717–734. http://dx.doi.org/10.1037/a0019038.

van der Leij, A., & Morfidi, E. (2006). Core deficits and variable differences in Dutch poor readers learning English. *Journal of Learning Disabilities, 39*, 74–90. http://dx.doi.org/10.1177/00222194060390010701.

van der Leij, A., & van Daal, V. (1999). Automatization aspects of dyslexia: Speed limitations in word identification, sensitivity to increasing task demands, and orthographic compensation. *Journal of Learning Disabilities, 32*(5), 417–428. http://dx.doi.org/10.1177/002221949903200507.

van der Linden, W. (2009). Conceptual issues in response-time modeling. *Journal of Educational Measurement, 46*(3), 247–272. http://dx.doi.org/10.1111/j.1745-3984.2009.00080.x.

Van der Marck, S., Bucher, K., Maurer, U., Schulz, E., Brem, S., Buckelmüller, J., et al (2009). Children with dyslexia lack multiple specializations along the visual word form (VWF) system. *Neuroimage, 47*, 1940–1949. http://dx.doi.org/10.1016/j.neuroimage.2009.05.021.

van der Mark, S., Klaver, P., Bucher, K., Maurer, U., Schulz, E., Brem, S., et al (2011). The left occipitotemporal system in reading: Disruption of focal fMRI connectivity to left inferior frontal and inferior parietal language areas in children with dyslexia. *NeuroImage, 54*, 2426–2436. http://dx.doi.org/10.1016/j.neuroimage.2010.10.002.

van IJzendoorn, M. H., & Bus, A. G. (1994). Meta-analytic confirmation of the nonword reading deficit in developmental dyslexia. *Reading Research Quarterly, 30*, 266–275. http://dx.doi.org/10.2307/747877.

Van Orden, G. C., & Goldinger, S. D. (1994). Interdependence of form and function in cognitive systems explains perception of printed words. *Journal of Experimental Psychology: Human Perception and Performance, 20*, 1269–1291. http://dx.doi.org/10.1037//0096-1523.20.6.1269.

Van Orden, G. C., & Paap, K. R. (1997). Functional neuroimages fail to discover pieces of mind in the parts of the brain. *Philosophy of Science (Proceedings), 64*, S85–S94.

Van Orden, G. C., Pennington, B. F., & Stone, G. O. (1990). Word identification in reading and the promise of subsymbolic psycholinguistics. *Psychological Review, 97*, 488–522. http://dx.doi.org/10.1037//0033-295X.97.4.488.

Vandermosten, M., Boets, B., Wouters, J., Ghesquière, P. (in press). A qualitative and quantitative review of diffusion tensor imaging studies in reading and dyslexia. *Neuroscience and Biobehavioral Reviews*. http://dx.doi.org/10.1016/j.neubiorev.2012.04.002.

Vellutino, F. R., Fletcher, J. M., Snowling, M. J., & Scanlon, D. M. (2004). Specific reading disability (dyslexia): What have we learned from the past four decades? *Journal of Child Psychology and Psychiatry, 45*, 2–40. http://dx.doi.org/10.1046/j.0021-9630.2003.00305.x.

Vellutino, F. R., Scanlon, D. M., & Tanzman, M. S. (1994). Components of reading ability: Issues and problems in operationalizing word identification, phonological coding, and orthographic coding. In G. R. Lyon (Ed.), *Frames of reference for the assessment of learning disabilities: New views on measurement issues* (pp. 279–332). Baltimore, MD: Brookes.

Verhaeghen, P., & Cerella, J. (2002). Aging, executive control, and attention: A review of meta-analyses. *Neuroscience and Biobehavioral Reviews, 26*, 849–857. http://dx.doi.org/10.1016/S0149-7634(02)00071-4.

Verhaeghen, P., Steitz, D. W., Sliwinski, M. J., & Cerella, J. (2003). Aging and dual-task performance: A meta-analysis. *Psychology and Aging, 18*(3), 443–460. http://dx.doi.org/10.1037/0882-7974.18.3.443.

Vinckier, F., Dehaene, S., Jobert, A., Dubus, J. P., Sigman, M., & Cohen, L. (2007). Hierarchical coding of letter strings in the ventral stream: Dissecting the inner organization of the visual word-form system. *Neuron, 55*, 143–156. http://dx.doi.org/10.1016/j.neuron.2007.05.031.

Yapp, R., & van der Leij, A. (1993). Word processing in dyslexics: An automatic decoding deficit? *Reading and Writing: An Interdisciplinary Journal, 5*, 261–279.

Yoncheva, Y. N., Blau, V. C., Maurer, U., & McCandliss, B. D. (2010). Attentional focus during learning impacts N170 ERP responses to an artificial script. *Developmental Neuropsychology, 35*(4), 423–445. http://dx.doi.org/10.1080/87565641.2010.480918.

Yoncheva, Y. N., Zevin, J. D., Maurer, U., & McCandliss, B. D. (2010). Auditory selective attention to speech modulates activity in the visual word form area. *Cerebral Cortex, 20*(3), 622–632. http://dx.doi.org/10.1093/cercor/bhp129.

Ziegler, J. C., & Goswami, U. (2005). Reading acquisition, developmental dyslexia, and skilled reading across languages: A psycholinguistic grain size theory. *Psychological Bulletin, 131*(1), 3–29. http://dx.doi.org/10.1037/0033-2909.131.1.3.

Ziegler, J. C., & Goswami, U. (2006). Becoming literate in different languages: Similar problems, different solutions. *Developmental Science, 9*(5), 429–453. http://dx.doi.org/10.1111/j.1467-7687.2006.00509.x.

Ziegler, J. C., Montant, M., & Jacobs, A. M. (1997). The feedback consistency effect in lexical decision and naming. *Journal of Memory and Language, 37*, 533–554. http://dx.doi.org/10.1006/jmla.1997.2525.

Ziegler, J. C., & Perry, C. (1998). No more problems in Coltheart's neighborhood: Resolving neighborhood conflicts in the lexical decision task. *Cognition, 68*, 53–62. http://dx.doi.org/10.1016/S0010-0277(98)00047-X.

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

Zoccolotti, P., De Luca, M., Di Filippo, G., Judica, A., & Martelli, M. (2009). Reading development in an orthographically regular language: Effects of length, frequency, lexicality and global processing ability. *Reading and Writing, 22*, 1053–1079. http://dx.doi.org/10.1007/s11145-008-9144-8.

Zoccolotti, P., De Luca, M., Judica, A., & Spinelli, D. (2008). Isolating global and specific factors in developmental dyslexia: A study based on the rate and amount model (RAM). *Experimental Brain Research, 186*, 551–560. http://dx.doi.org/10.1007/s00221-007-1257-9.