

## PDF hosted at the Radboud Repository of the Radboud University Nijmegen

The following full text is a publisher's version.

For additional information about this publication click this link.

<http://hdl.handle.net/2066/101028>

Please be advised that this information was generated on 2021-09-20 and may be subject to change.

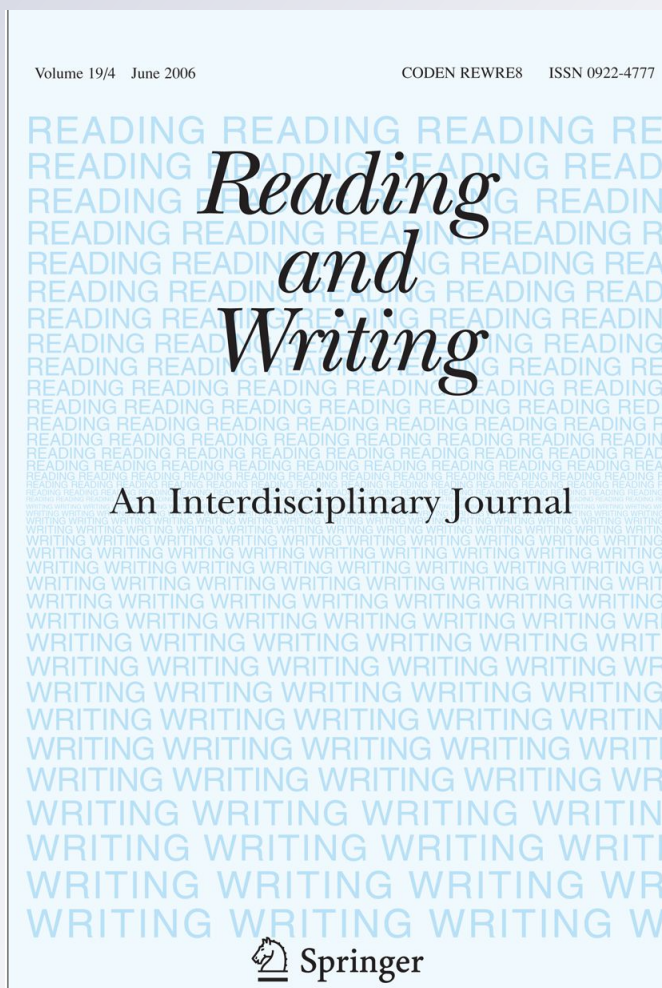
*Improving word reading speed: individual differences interact with a training focus on successes or failures*

**Esther G. Steenbeek-Planting, Wim H. J. van Bon & Robert Schreuder**

**Reading and Writing**  
An Interdisciplinary Journal

ISSN 0922-4777  
Volume 25  
Number 9

Read Writ (2012) 25:2061-2089  
DOI 10.1007/s11145-011-9342-7



**Your article is published under the Creative Commons Attribution license which allows users to read, copy, distribute and make derivative works, as long as the author of the original work is cited. You may self-archive this article on your own website, an institutional repository or funder's repository and make it publicly available immediately.**

## Improving word reading speed: individual differences interact with a training focus on successes or failures

Esther G. Steenbeek-Planting · Wim H. J. van Bon ·  
Robert Schreuder

Published online: 29 September 2011

© The Author(s) 2011. This article is published with open access at Springerlink.com

**Abstract** The effect of two training procedures on the development of reading speed in poor readers is examined. One training concentrates on the words the children read correctly (*successes*), the other on the words they read incorrectly (*failures*). Children were either informed or not informed about the training focus. A randomized controlled trial was conducted with 79 poor readers. They repeatedly read regularly spelled Dutch consonant–vowel–consonant words, some children their successes, others their failures. The training used a computerized flashcards format. The exposure duration of the words was varied to maintain an accuracy rate at a constant level. Reading speed improved and transferred to untrained, orthographically more complex words. These transfer effects were characterized by an Aptitude-Treatment Interaction. Poor readers with a low initial reading level improved most in the training focused on successes. For poor readers with a high initial reading level, however, it appeared to be more profitable to practice with their failures. Informing students about the focus of the training positively affected training: The exposure duration needed for children informed about the focus of the training decreased more than for children who were not informed. This study suggests that neither of the two interventions is superior to the other in general. Rather, the improvement of general reading speed in a transparent orthography is closely related to both the children’s initial reading level and the type of words they

---

E. G. Steenbeek-Planting · W. H. J. van Bon  
Behavioural Science Institute, Radboud University Nijmegen, P.O. Box 9104, 6500 HE Nijmegen,  
The Netherlands

E. G. Steenbeek-Planting · W. H. J. van Bon · R. Schreuder (✉)  
Department of Special Education, Radboud University Nijmegen, P.O. Box 9104,  
6500 HE Nijmegen, The Netherlands  
e-mail: e.steenbeek@pwo.ru.nl

R. Schreuder  
Donders Institute for Brain, Cognition, and Behaviour, Radboud University Nijmegen,  
P.O. Box 9104, 6500 HE Nijmegen, The Netherlands

practice with: common and familiar words when training their successes and uncommon and less familiar words with training their failures.

**Keywords** Training · Intervention · Orthography · Poor readers · Reading speed

Improving the reading fluency of dyslexic and poor readers is a major educational goal (e.g., Report of the National Reading Panel, 2000). Reading fluency is a complex, multifaceted construct defined as accurate, fast and effortless reading with good comprehension (Hudson, Pullen, Lane, & Torgesen, 2009; see also Kuhn, Schwanenflugel, & Meisinger, 2010). Some models that describe the development of word reading fluency proposed a continuous process that is characterized by several phases (Adams, 1990; Ehri, 1998, 1999, 2005). Other models emphasized the increments that occur in the acquisition process (cf. Hinton, McClelland, & Rumelhart, 1986; Perfetti, 1992). The complexity of a writing system affects the development of word reading fluency: children learning to read an alphabetic language with a transparent orthography do so faster and more efficiently than children learning an opaque orthography as English (cf. Aro & Wimmer, 2003; Ellis et al., 2004; Seymour, Aro, & Erskine, 2003; Wimmer & Goswami, 1994). In addition to this, the nature of the orthography influences the types of reading difficulties children may experience (Caravolas, 2007): children reading opaque orthographies make more reading errors than children reading transparent orthographies (Aro & Wimmer, 2003; Frith, Wimmer, & Landerl, 1998; Patel, Snowling, & de Jong, 2004; Seymour et al., 2003). A prime characteristic of reading disorders in languages with a transparent orthography is the impairment in reading speed (de Jong & van der Leij, 2003; Landerl, 2001; Serrano & Defior, 2008).

This study examines the development of word reading fluency in Dutch, an orthographically transparent orthography. We will focus on the reading of words with a phonological consonant–vowel–consonant (CVC) structure. These words—with the exception of a few loan words—are orthographically fully transparent in reading, that is, they can be read by applying only context-free grapheme-phoneme correspondence (GPC) rules. Many polysyllabic words, in contrast, are more complex to decode because they require the use of additional contextual, graphotactical or morphological rules. Verhoeven and van Leeuwe (2009) investigated the development of CVC word reading fluency in Dutch typical and poor readers. Their study showed that gains in accuracy occur very rapidly from the beginning of reading instruction and taper off thereafter. Growth of decoding skills thereafter—in typical and poor readers—was found to be largely a matter of increased reading speed. Remediation has been shown to improve accuracy in Dutch poor readers, but reading speed tends to remain low compared to normally developing children (van der Leij & van Daal, 1989; see also Martens & de Jong, 2008; Scheltinga, van der Leij, & Struiksma, 2010; Yap & van der Leij, 1993). Torgesen, Rashotte and Alexander (2001) showed that a low reading speed in identifying single words is the most important factor accounting for individual differences in reading fluency (see also Jenkins, Fuchs, van den Broek, Espin, &

Deno, 2003; Schwanenflugel, Meisinger, Wisenbaker, Kuhn, Strauss, & Morris, 2006; Vadasy & Sanders, 2009). Therefore, our research on remedial intervention focuses on the identification speed of single words. The present study investigates the differential effects of training methods targeted on the word reading speed of poor readers. The study contrasts a training method focused on words read correctly (*successes*) with a training method focused on words read incorrectly (*failures*) and investigates the interaction of these with the effect of being informed or not about the focus of training material.

To improve the identification of single words, repetition is proposed to be a key element, next to other beneficial measures such as providing immediate and corrective feedback, direct instruction, and scaffolding (see e.g., Chard, Vaughn, & Tyler, 2002; Kuhn & Stahl, 2003; Meyer & Felton, 1999; Swanson, 1999; Swanson, Hoskyn, & Lee, 1999). Poor readers require more practice than normally developing children to speedily identify words (Reitsma, 1983). Berends and Reitsma (2006) argued that it is more beneficial to repeatedly train on a few words over once reading many words. If word repetition is important, the question needs to be answered *which words* should be repeated: words that are read correctly, or words that appear to offer some difficulty, as they are read erroneously. That is the central question of this study. Focusing on successes is operationalized in this study by removing from the training set words that were read incorrectly during flashcard training sessions. In the training focusing on failures, words read correctly during the training are removed from the training set. Arguments favoring either one of these procedures will be discussed below.

In following Ehri's (1998, 1999) framework, Wolf and Katzir-Cohen (2001) emphasized that instruction for fluency development should begin focusing on the accuracy of the word level and its underlying representations. In line with this view, the recommended practice in the remedial teaching of poor readers is to focus on words that are read incorrectly (e.g., Bender, 2004; Martens, Witt, Daly, & Volmer, 1999). Insofar as reading errors offer a window on the reading process that children use to read words (Allen, 1976; Au, 1977; Goodman, 1969; Savage, Stuart, & Hill, 2001; Singleton, 2005; Weber, 1970), the analysis of errors and miscues may be important in determining directions in reading instruction (McKenna & Picard, 2006), and therefore is advocated in popular textbooks for reading instruction in the primary school (e.g., Beard, 1990; Graham & Kelly, 1997; Roberts, 1989). Hall (2003) showed that a study of reading errors can be very instructive and helps teachers to gain an understanding of children and their difficulties. Such considerations and findings emphasized concentration on errors in reading practice and form the foundations of an intervention focused on failures.

There are, however, arguments in favor of the opposite approach, focusing on successes, which entails that children reread words they have successfully decoded. One argument is that, to stimulate fast and effortless reading, children must be able to automatize the application of GPC rules or automatically identify words. According to the self-teaching hypothesis (cf. Share, 1995, 1999), each successful identification of a given word increases the likelihood to successfully read that word again, as the reader obtains word-specific orthographic information. The self-teaching hypothesis argues that fast and accurate word identification depends on the

frequency to which a child has been exposed to a particular word. Words that are read repeatedly are likely to be recognized visually with minimal phonological processing from the very earliest stages of reading acquisition. Thus, repeatedly reading successes should enable readers to catalyze the process of reading words from phonological recoding into reading words that have become increasingly lexicalized.

A second argument stems from our earlier study (Steenbeek-Planting, van Bon, & Schreuder, 2011). We investigated the instability of errors in Dutch CVC words, that is, how often words were read correctly at one time and incorrectly at another. This study showed that typically developing readers in first and second grade, and reading-level-matched poor readers did not repeatedly misread the same items. As for the words that were read inaccurately by the children, only a fourth of those words were read incorrectly twice, and three-fourths of those words were read incorrectly at one time, but correctly at the other. Errors thus were unstable to an important degree, and they were, most probably, not caused by a lack of GPC rule knowledge, but rather by inattentiveness or other, stochastic processes. If errors are determined largely by random factors, a focus on such errors will not be effective for enhancing reading speed because they are too unreliable to determine word selection. Moreover, the words that were read consistently wrong by beginning readers were not representative for the CVC orthographic type as such. They were characterized by a low word frequency, a low bigram frequency and a rather small and low frequency neighborhood. Thus, a training focused on failures inevitably uses words that are orthographically unlike most other CVC words, with sublexical components that are atypical for CVC words. If, children, however, practice on successes, they practice on typical CVC words, and therefore on the common GPC rules, which, consequently, can be expected to bring about greater transfer to CVC word reading than a training focusing on failures.

The effects of a training focusing either on failures or on successes may be influenced by motivational factors. Children experiencing learning problems often feel discouraged by their failure to learn (Riddick, 2010). A reading intervention focusing on failures may therefore discourage the learners even more. Repeatedly being confronted with failure may affect their self-esteem and as a consequence would be counterproductive (Shute, 2008). Some training theorists therefore suggested that learning should focus on successful events paired with positive reinforcement (Latham, 1989). On the other hand, confronting the readers with their performance by means of a training procedure that focuses on failures, might urge them through negative reinforcement to avoid making errors and to overcome their failures by working harder and more efficiently (Kluger & DeNisi, 1996; Locke & Latham, 2002). To investigate whether improvement is additionally affected by positive or negative reinforcement, the children in our study are either informed or not informed about the focus of the training. Children who are informed are expected to show greater learning effects either through positive reinforcement (in the training focused on successes) or through negative reinforcement (in the training focused on failures). It is expected that children who are *not* informed about the focus of the training, will not or to a lesser degree show the aforementioned



effects and will show smaller learning effects, because they do not know that their practicing is associated with either their failures or successes.

This research seeks to answer the following questions: Are reading speed and accuracy differentially affected by a training focused on failures and a training focused on successes? Are such training effects strengthened by informing children about the training focus? The questions are answered by using a randomized controlled trial, in which children are randomly assigned to one of four flashcard training conditions:

1. SI: Focus on successes (S) and informed (I) about this focus.
2. SN: Focus on successes (S), but not informed (N) about this focus.
3. FI: Focus on failures (F) and informed (I) about this focus.
4. FN: Focus on failures (F), but not informed (N) about this focus.

In every condition, all children receive immediate feedback that qualifies each oral reading response as correct or incorrect.

Studies accounting for individual differences in response to intervention (see Bracht, 1970; Shavelson, Berliner, Ravitch, & Loeding, 1974) have resulted in systematic analyses of what has been termed Aptitude-Treatment Interactions (ATIs) (Cronbach & Snow, 1977; Snow, 1991; see also Kanfer & Ackerman, 1989). The data therefore will be used to investigate also whether some poor readers benefit more from a training focused on failures and other poor readers benefit more from a focus on successes, and whether training outcomes are related to the children's initial reading level.

## Method

### Participants

Poor readers ( $n = 83$ , of whom 50 male) were selected from four primary schools for special education. Children are in this type of education (see Eurybase, 2008) because of learning disabilities, mild mental retardation or mild behavioral problems. The majority of these children (73%) are poor readers (van Bon, Bouwmans, & Broeders, 2006). To ensure a sample of typical IQ, children with mental retardation or behavioral problems were excluded from participation. Formal reading instruction in these and other Dutch primary schools is based on phonics instruction and starts in Grade 1.

Four children left the study due to illness or moving to another school. Therefore, results on 79 children (of whom 47 male) will be reported. Children were defined as poor readers because they scored below the 10th percentile for their grade level on at least two of the following standardized screening tests that also served as pretests: Lexical Decision Tests (LDT1, LDT2), Word Decoding Tests (WDT1, WDT2, WDT3) or a Nonword Reading Test (NRT). All children were able to sound out the graphemes according to the Dutch GPC rules, had Dutch as their first language, and did not have any diagnosed neurological problems, nor a speech, hearing or visual impairment. The teachers also classified these children as poor readers and



considered the reading problems of these children as not caused by behavioral problems. Parents were informed about the participation of the children and had given written consent for participation.

Children were randomly assigned to one of the four experimental training conditions SI, SN, FI or FN (see end of introduction). Boys and girls were evenly distributed over experimental groups. The training groups were evenly distributed across schools and classes to reduce school-specific or teacher-specific effects.

In order to assess whether ATI effects of initial reading level apply to our data, children were qualified as having either a *high initial reading level* (HI) or a *low initial reading level* (LI), based on their reading composite score on the WDT1 prior to intervention. Thirty-seven children with a reading composite score at or below the median (35 words read correctly per minute,) were considered LI children; 42 children with a reading composite score above the median formed the group of HI children.<sup>1</sup>

The training groups did not differ in age ( $F < 1$ ), number of months of formal reading instruction ( $F < 1$ ), on any pretest, both as to composite scores as well as to the disaggregated component scores for accuracy and speed ( $F < 1$  for all tests). The training groups did not differ on these variables either when analyses were conducted separately for LI children and HI children ( $F$ s  $< 1$ ). The HI children performed better than the LI children on all pretests, both as to the composite scores ( $F(6,72) = 22.3$ ,  $p < .001$ ,  $\eta_p^2 = .65$ ), as well as to the component scores for accuracy ( $F(7,71) = 6.63$ ,  $p < .001$ ,  $\eta_p^2 = .45$ ), and speed ( $F(7,71) = 14.57$ ,  $p < .001$ ,  $\eta_p^2 = .59$ ). The LI children received fewer months of reading instruction than the HI children ( $F(1,77) = 5.36$ ,  $p < .05$ ,  $\eta_p^2 = .07$ ), however, the groups did not differ in chronological age ( $F < 1$ ).

Table 1 provides an overview of characteristics of the experimental subgroups. Scores on pretests can be found in Tables 5 and 6.

## Procedure and instruments

### *Computerized flashcard training*

*Stimuli in the training set* Words with a phonological CVC<sup>2</sup> structure were selected from the Celex Database (Baayen, Piepenbrock, & van Rijn, 1993). The selected words were the lemmas for words that can occur independently in a language (i.e., nouns, verbs, adjectives, adverbs, and numbers). Words that were considered rather idiosyncratic or shocking, proper names and words with a foreign orthography or phonology were eliminated (Nunn, 1998; Booij, 1995). Additionally, some words were considered unsuitable for training purposes. These were low-frequency words (e.g., *zijkl* [drainage watercourse]) having a homophonic high-frequency counterpart (e.g., *zeil* [sail]). Only the high-frequency homophones (zeil)

<sup>1</sup> It should be noted, that reading performance is studied as a dichotomous instead of a continuous variable. Ideally, the ATI effects should be studied with a multilevel approach. However, data acquisition for such approach was beyond the scope of this exploratory study.

<sup>2</sup> C and V represent single graphemes and thus include digraphs. Dutch CVC words contain maximally five letters (e.g., *leeuw* [lion]).

**Table 1** Composition of the training groups as to gender, mean age and reading instruction (RI) in months

	Low initial level						High initial level									
	SI ( <i>n</i> = 10)		SN ( <i>n</i> = 12)		FI ( <i>n</i> = 9)		FN ( <i>n</i> = 11)		SI ( <i>n</i> = 10)		SN ( <i>n</i> = 6)		FI ( <i>n</i> = 11)		FN ( <i>n</i> = 10)	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Age	100.40	19.28	99.50	20.95	109.67	17.31	104.36	13.90	104.80	15.48	110.00	19.28	101.36	10.14	109.00	12.53
RI	24.50	6.69	27.25	9.51	21.22	7.10	22.09	6.64	27.90	8.86	23.83	7.49	29.36	9.97	30.00	8.01
Gender	8 M/2 F		6 M/6 F		5 M/4 F		6 M/5 F		4 M/6 F		4 M/2 F		7 M/4 F		7 M/3 F	

SI = Training focused on successes and children informed about this focus, SN = Training focused on successes and children not informed about this focus, FI = Training focused on failures and children informed about this focus, FN = Training focused on failures and children not informed about this focus

were used, as otherwise children might memorize the low-frequency written word form (*zjil*) with the high-frequency word meaning. Of the initial set of 1,078 words, 845 thus constituted the training set.

*General training setup* The training consisted of 10 sessions of approximately 20 min each. Children practiced individually once or twice a week. The training used the flashcards format of van den Bosch, van Bon and Schreuder (1995). In each training session 100 CVC words were presented, randomly selected from the training set, one at a time on a computer screen. Children were instructed to name the presented word as accurately and as fast as possible. At the end of the 10 sessions, each child had thus read 1,000 times a word from the training set. CVC words were presented in black, lower case letters (Arial, size 48). on a white background in the centre of a computer screen. The letters had a height of approximately 1.5 cm and words ranged from 2 to 6.5 cm in length. The child was seated in front of the computer screen, at a distance of approximately 60–80 cm.

The exposure duration of the words was varied to maintain the accuracy rate at an approximately constant level. After each trial, the reading accuracy of the last word and the previous five words were evaluated. The exposure duration of the next word was increased by 17 ms if four or more words out of these six had been read incorrectly, and was decreased by 17 ms if five or six out of the six words were read correctly. In the other cases (three or four correct), exposure duration remained unchanged. In this way, the accuracy rate was maintained at an approximately constant level of 67%. Due to software failure, each session started with an exposure duration of 350 ms, instead of the exposure duration the previous session ended with.<sup>3</sup>

Each word was preceded by a fixation cross (a+) in the center of the screen for 800 ms. After a blank screen for 200 ms, the word was presented with varying exposure duration. The word was followed by hash marks (####) in order to prevent further visual processing of the letter string. As soon as a voice key registered a verbal response, the hash marks disappeared and were followed by visual feedback (1,000 ms). The visual feedback on the screen indicated whether the verbal response was correct (smiley) or incorrect (sad face). At the end of each session, the child was shown a computer graph that depicted the presentation times of the words read in the current and in the previous sessions and the meaning of the graph was explained if necessary. This graph visualized the child's progress and should motivate the children to perform well (Kluger & DeNisi, 1996). Each session started with a short practice block of six randomly chosen CV and VC words in order to get accustomed to the training.

<sup>3</sup> Each session started with an exposure duration of 350 ms instead of the exposure duration the previous session ended with. This may have caused the mean accuracy score to turn out somewhat lower (60%). This difference, however, is negligible. All analyses on the data of flashcard training (see “[Results—Flashcard training](#)”) were double checked by reanalysing the data for only the second half of each session. The reanalysis rendered exactly the same main and interaction effects. This shows that the starting exposure duration of each session was not of crucial importance and the adjustment of exposure duration worked well within each session.

*Training conditions* Children received training according to one of the four different conditions (SI, SN, FI and FN). In each session, 100 words, taken randomly without replacement from the training set (initial  $n = 845$ ), were presented. For training groups focusing on successes (SI and SN), the words read incorrectly during the training session were removed from further training, thus reducing the number of words in the training set. In the next session, again 100 words were randomly taken without replacement from the remaining training set. This procedure was applied to all 10 training sessions. For the groups focusing on failures (FI and FN), words read correctly during the training session were removed. In both training conditions, words used in invalid trials were not removed.

Apart from focus, training conditions differed in the information that was given about the focus of the training. In the non-informed condition, children were not told that their training was either focused on failures or on successes. In contrast, children in the informed condition were explicitly shown, reminded by graphics and a short movie,<sup>4</sup> and told at the beginning of each session that they were practicing on their past successes and new items or on their past failures and new items.

Laptops with 14" screens were used. The correctness of verbal responses was recorded by the experimenter, and stored in the computer by means of a buttonbox.

### *Pre- and post-tests*

Children were screened for selection with LDT1 and LDT2 and thereafter, WDT1, 2 and 3, NRT were additionally used for selection. These tests also served as pretest, together with a Sentence Verification Test (SVT). Parallel versions of the tests were used at the posttest.

*Lexical Decision Test (LDT)* The students were asked to complete two versions of a standardized paper-and-pen Lexical Decision Test (van Bon, 2007). Each version involves a card with words distributed across it in columns. LDT1 is composed of CVCC and CCVC words following the dominant orthographic rules. Sixty high-frequency nouns that are likely to be known in their spoken forms by 6-year-olds (Kohnstamm, Schaerlaekens, de Vries, Akkerhuis, & Frooninckxs, 1999) were interspersed with 20 pseudowords. LDT2 is composed of bisyllabic words following the dominant orthographic rules, and has 90 high-frequency nouns likely to be known in their spoken forms by 6-year-olds, interspersed with 30 pseudowords. This test is administered in class and students are asked to silently read the items and cross out every pseudoword. The raw score for each test is the number of words judged within a minute minus the number of errors. Test–retest reliability for children in Grades 1 to 3 is considered sufficient, .81 for LDT1 and .82 for LDT2, (van Bon, 2007). The LDT taps word reading skills and semantic knowledge. We

<sup>4</sup> After a reading error, children in the FI-training saw a short movie in which a word that had been read incorrectly was 'dumped' into a bucket and stored in the computer. A picture at the end of the session illustrated that the failures collected in the bucket were returned to a box with words they would practice in the subsequent sessions, i.e. the training set. Conversely, after a correct response, children in the SI-condition saw the words read correctly being collected and at the end of the session returned to the box with words that they would practice in the subsequent sessions.

incorporated the LDT, as it is an adequate and reliable alternative for using oral reading tests (van Bon, Hoevenaars, & Jongeneelen, 2004).

*Word Decoding Test (WDT)* A standardized word reading test (Verhoeven, 1995) was administered individually to assess the oral reading abilities of the students for words in isolation. This test consists of three cards with words listed in columns (WDT1: simple monosyllabic words; WDT2: monosyllabic words with one or two consonant clusters; WDT3: two-, three-, and four-syllable words). Students are instructed to read the words aloud as quickly and accurately as possible. The composite score for each card is the number of words read correctly in 1 min. The reported reliability of the three cards (Cronbach's  $\alpha$ ) ranges from .86 to .94 (Verhoeven & van Leeuwe, 2003) and is judged sufficient.

*Nonword Reading Test (NRT)* In order to assess the decoding ability of the students for pseudowords, a standardized Nonword Reading Test was administered (van den Bos, Lutje Spelberg, Scheepstra, & de Vries, 1994). The test consists of pseudowords of increasing length. The students are instructed to read the pseudowords aloud as quickly and accurately as possible. The test score is the number of nonwords read correctly in 2 min. The parallel reliability is good, .93 and above (van den Bos et al., 1994).

*Sentence Verification Test (SVT)* In order to determine word reading skill in sentence context and comprehension skill, a computerized sentence verification task derived from van den Bosch et al. (1995) and Wentink (1997) was used. Thirty semantically correct sentences (e.g., *Kaas is geel*. [Cheese is yellow.]) and fifteen semantically incorrect sentences (e.g., *De trein is zuur*. [The train is sour.]) were presented one-by-one on a computer screen in random order. Sentences consisted of high-frequency monosyllabic words that follow the dominant orthographic rules, and that are likely to be known in their spoken forms by 6-year-olds (Kohnstamm et al., 1999). Children were asked to silently read the sentences as quickly and accurately as possible, and then to judge the sentences as meaningful or not meaningful by pressing a button. Thereafter, the sentence disappeared and a new sentence appeared on the screen. A blank screen of 2 s appeared in between the presentation of the sentences. Children got acquainted with the test by judging four training sentences. There were no time limitations to this task and no feedback was given. The score is the percentage of correctly judged sentences. SVT responses and latencies were recorded by the laptops.

## Results

First we present data on word reading collected during the flashcard training. The stimuli changed from session to session for each individual, and therefore we answer the questions whether the content of the training differed for the training groups with respect to the number of words and their word characteristics in the last

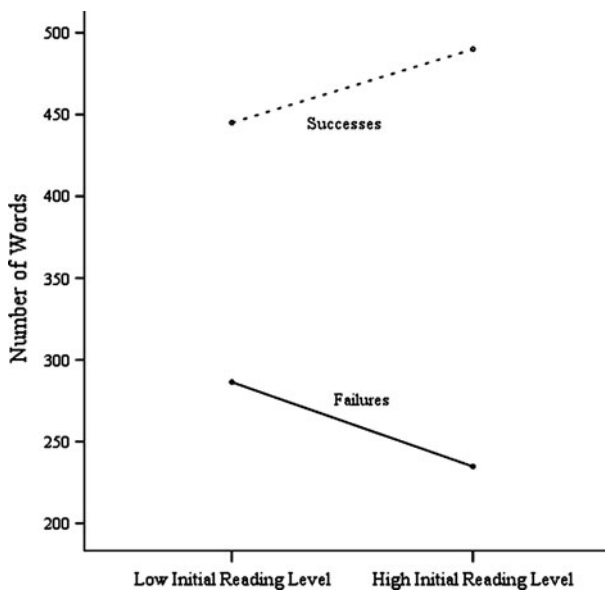
training set. Changes during the training are explored by investigating the accuracy scores and exposure durations from the first to the last training session. Planned comparisons are made with respect to Focus (successes vs. failures), Information (informed vs. non-informed) and Initial Level (high vs. low).

Second, we present the results on transfer measures of word reading skills. We conduct repeated measures MANOVAs with the pre- and post-test component scores for reading speed and accuracy on LDTs, WDTs, NRT, and SVT as dependent variables. Between-subjects factors are Focus, Information, and Initial Level. Time (pre- vs. posttests) are the within-subjects factor. To assess whether our interventions had additional practice value, in the last section, we compare the pre- and posttest scores of the participants on the standardized tests to the respective age norms.

### The flashcard training

#### *Stimuli of the last training set*

An analysis of variance was conducted with Number of words in the last training set as a dependent variable and Focus, Information, and Initial Level as between-subjects factors. A main effect of Focus ( $F(1,71) = 283.67, p < .001, \eta_p^2 = .80$ ) was further qualified by an interaction of Focus by Initial Level ( $F(1,71) = 15.44, p < .001, \eta_p^2 = .18$ ). Post hoc tests with Bonferroni correction showed that children focusing on successes had more words in their last training sets than children focusing on failures (see Fig. 1), both as to the LI children ( $p < .001$ ) and HI



**Fig. 1** Mean number of words in the training set of the last session for children with a low and a high initial reading level in the training focused on failures and on successes

children ( $p < .001$ ). As for the children focusing on failures, the LI children had more words in their last training sets than the HI children ( $p < .05$ ), whereas this effect was (marginally) reversed for the children focusing on successes ( $p = .07$ ). Thus, the difference in number of words between groups focusing on failures versus successes is largest for the HI children and smallest for the LI children.

Next, we analyzed the lexical and sublexical characteristics: (1) length, (2) mean log bigram frequency, (3) frequency (i.e., the natural logarithm of a word's frequency per million, Baayen et al., 1993), (4) number of orthographic neighbors (i.e., words differing in one letter), (5) number of phonological neighbors (i.e., words differing in one phoneme), (6) frequency of the most frequent orthographic neighbor, and (7) frequency of the most frequent phonological neighbor. These characteristics (see Table 2) were reduced to three uncorrelated factors with principal components analysis: *pattern frequency* (marked by high loadings of bigram frequency and frequency for the most frequent orthographic and phonological neighbors), *neighborhood size* (marked by high loadings for the number of phonological and orthographical neighbors and by a negative loading for word length), and *word frequency* (marked by word frequency).

These three factor scores were entered as the dependent variables into a multivariate analysis of variance. Focus, Information, and Initial Level were the between-subjects variables. Words trained on by children focusing on successes have a higher pattern frequency ( $F(1,28,271) = 17.95, p < .001, \eta_p^2 < .01$ ), larger neighborhood size ( $F(1,28,271) = 150.96, p < .001, \eta_p^2 = .01$ ), and higher word frequency ( $F(1,28,271) = 165.73, p < .001, \eta_p^2 = .01$ ) than words trained on by children focusing on failures. As for the groups focusing on failures, words trained on by LI children have a higher word frequency ( $F(1,10,782) = 6.03, p < .05, \eta_p^2 < .01$ ) and a lower neighborhood size ( $F(1,10,782) = 6.64, p < .01, \eta_p^2 < .01$ ) than words trained on by HI children. It seems that LI children do not only have difficulty reading low-frequency words, but also words with a low neighborhood size.

### Accuracy during training

A repeated measures ANOVA was conducted with log odds of the percentage correct of each session (see Allerup & Elbro, 1998) as a dependent variable (see Table 3). Time (Session 1 to Session 10) was entered as a within-subjects factor and Focus, Information, and Initial Level as between-subjects factors.

Results show main effects of Time ( $F(9, 63) = 4.43, p < .001, \eta_p^2 = .39$ ), and Initial Level ( $F(1,71) = 16.92, p < .001, \eta_p^2 = .19$ ), indicating that HI children have higher accuracy scores than the LI children (see Fig. 2). The LI children had lower accuracy scores in the first session than the HI children ( $F(1,77) = 14.08, p < .001, \eta_p^2 = .16$ ), and this difference apparently did not disappear. An interaction of Time by Initial Level was also found (Huyn-Feldt correction applied, see Keselman et al., 1998) ( $F(7.35, 521.47) = 2.33, p < .05, \eta_p^2 = .03$ ). Tests of within-subjects contrasts (polynomial) show an interaction of Time by Initial Level at the cubic level only ( $F(1,77) = 4.87, p < .05, \eta_p^2 = .06$ ). The HI and LI groups increase their accuracy scores from session 1 to session 10, but they do so in a different time course. LI children show an initial stagnation, after which they increase in accuracy and this



**Table 2** Means and standard deviations of word characteristics, grouped within three factors, in the training set of session 10 for the training focused on failures and on successes, split to children with a low and high initial level

Word characteristics	Range	Training groups								
		Low initial level				High initial level				
		Failures		Successes		Failures		Successes		
		M	SD	M	SD	M	SD	M	SD	
Pattern frequency										
WF of orthographic neighbor	-10-4.32	2.33	0.77	2.41	0.73	2.35	0.76	2.39	0.74	
WF of phonological neighbor	.52-4.32	2.57	0.66	2.59	0.65	2.56	0.66	2.58	0.65	
Mean log bigram frequency	11.53-15.79	13.68	0.68	13.76	0.67	13.69	0.69	13.75	0.67	
Neighborhood size										
Number of orthographic neighbors	0-30	12.00	4.98	12.97	4.91	12.23	5.06	12.82	5.00	
Number of phonological neighbors	3-35	17.75	6.04	18.57	5.84	17.90	6.02	18.44	5.89	
Word length	3-5	3.65	0.49	3.60	0.50	3.63	0.49	3.60	0.50	
Word frequency										
Word frequency	-1.60-4.32	0.72	1.02	0.89	1.07	0.68	1.03	0.88	1.05	

WF of orthographic neighbor = Word frequency of the most frequent orthographic neighbor; WF of phonological neighbor = Word frequency of the most frequent phonological neighbor. Word frequency is presented log transformed

**Table 3** Means and standard deviations of reading accuracy (in percentages) in each training session of the training groups

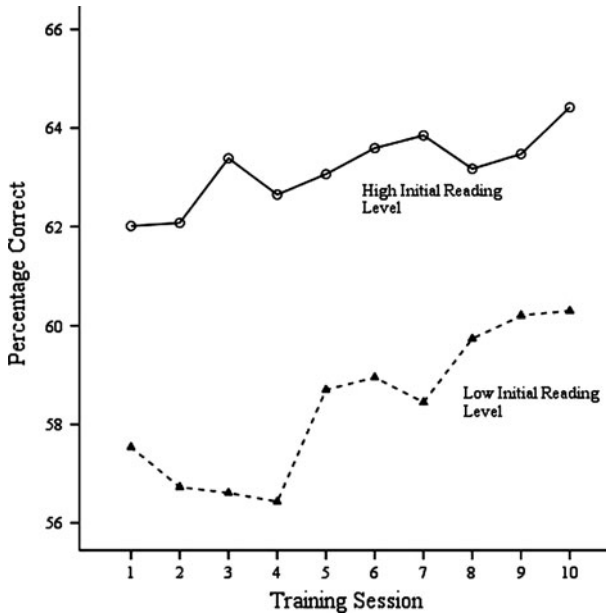
Session	Low initial level							
	SI ( <i>n</i> = 10)		SN ( <i>n</i> = 12)		FI ( <i>n</i> = 9)		FN ( <i>n</i> = 11)	
	<i>M</i>	SD	<i>M</i>	SD	<i>M</i>	SD	<i>M</i>	SD
1	59.24	2.00	55.87	10.50	57.04	6.55	58.21	5.96
2	59.11	3.80	54.08	12.09	56.12	8.61	57.94	10.30
3	61.10	2.02	53.26	13.03	55.96	8.77	56.71	10.09
4	61.18	2.52	54.74	12.83	53.01	10.60	56.75	9.63
5	62.72	2.23	56.94	9.92	58.83	5.20	56.86	8.24
6	62.76	1.61	58.22	9.02	57.98	8.25	57.08	9.60
7	61.91	2.15	56.44	10.33	58.73	5.60	57.26	9.33
8	62.75	3.42	58.59	8.95	58.31	7.00	59.42	5.46
9	64.01	4.36	58.91	8.27	58.73	5.90	59.34	4.29
10	63.25	4.27	58.72	11.05	59.61	3.88	59.87	6.06
Session	High initial level							
	SI ( <i>n</i> = 10)		SN ( <i>n</i> = 6)		FI ( <i>n</i> = 11)		FN ( <i>n</i> = 10)	
	<i>M</i>	SD	<i>M</i>	SD	<i>M</i>	SD	<i>M</i>	SD
1	62.23	1.56	61.72	2.46	62.29	2.04	61.66	2.13
2	62.35	1.63	62.08	1.94	62.06	1.40	61.81	2.00
3	63.04	1.58	68.63	9.33	61.97	2.09	62.14	1.98
4	62.91	2.32	62.79	1.69	62.68	2.49	62.28	2.30
5	63.76	1.52	63.91	1.27	62.99	1.97	61.95	1.75
6	63.85	3.45	64.00	2.00	63.55	1.73	63.13	2.93
7	64.24	1.99	64.87	4.55	63.08	3.10	63.69	2.58
8	64.67	3.89	63.43	2.82	62.62	2.51	62.12	2.32
9	63.56	2.66	64.95	4.54	63.59	2.67	62.36	1.94
10	64.88	2.30	66.82	6.53	63.51	3.64	63.51	4.71

SI = Training focused on successes and children informed about this focus, SN = Training focused on successes and children not informed about this focus, FI = Training focused on failures and children informed about this focus, FN = Training focused on failures and children not informed about this focus

is best described as a linear increase ( $F(1,41) = 22.19, p < .001, \eta_p^2 = .35$ ) and marginally as cubic ( $F(1,41) = 4.01, p = .05, \eta_p^2 = .09$ ). HI children show a slight increase over all ten sessions and their improvement over time can only be described as linear ( $F(1,36) = 9.98, p < .01, \eta_p^2 = .22$ ). No effects of Focus and Information ( $F < 1$ ) on reading accuracy were found.

#### *Exposure duration during training*

A repeated measures ANOVA was conducted with the mean log exposure duration of each session (see Table 4) as a dependent variable, Time (Session 1 to Session



**Fig. 2** Mean accuracy score (in percentages) of each training session for children with a low and high initial reading level

10) as a within-subjects factor, and Focus, Information, and Initial Level as between-subjects factors.

A main effect for Time ( $F(9, 63) = 13.24, p < .001, \eta_p^2 = .65$ ) indicates that the mean exposure duration decreased from the first to the last session. A main effect of Initial Level ( $F(1, 71) = 40.92, p < .001, \eta_p^2 = .37$ ) indicates that the exposure durations for LI children were longer than for HI children. LI children had longer exposure durations in the first session ( $F(1, 77) = 32.62, p < .001, \eta_p^2 = .30$ ), and this difference was apparent throughout all training sessions. Interestingly, two interactions were found. The interaction of Time by Focus ( $F(8.54, 606.61) = 1.96, p < .05, \eta_p^2 = .03$ , with Huyn-Feldt correction) is illustrated by Fig. 3. Tests of within-subjects contrasts (polynomial) show an interaction of Time by Focus at the linear level only ( $F(1, 77) = 6.37, p < .05, \eta_p^2 = .08$ ). For children focusing on successes exposure duration decreased in a linear trend ( $F(1, 40) = 22.61, p < .001, \eta_p^2 = .36$ ), and more rapidly than for children focusing on failures, whose decrease in exposure duration can be described as linear ( $F(1, 37) = 76.47, p < .001, \eta_p^2 = .67$ ), but also as quadratic ( $F(1, 37) = 6.16, p < .05, \eta_p^2 = .14$ ).

The interaction of Time by Information ( $F(8.54, 606.61) = 2.78, p < .01, \eta_p^2 = .04$ ) is further qualified by tests of within-subjects contrasts (polynomial) that show an interaction of Time by Information at the linear level only ( $F(1, 77) = 12.63, p < .001, \eta_p^2 = .14$ ). For children uninformed, the decrease in exposure duration can only be described as being linear ( $F(1, 38) = 15.95, p < .001, \eta_p^2 = .30$ ) and for children who are informed this can be described as being linear ( $F(1, 39) = 102.29, p < .001, \eta_p^2 = .72$ ), but also as quadratic ( $F(1, 39) = 4.75, p < .05, \eta_p^2 = .11$ ). As is

**Table 4** Means and standard deviations of exposure duration (in milliseconds) in each training session of the training groups

Low initial level								
	SI ( <i>n</i> = 10)		SN ( <i>n</i> = 12)		FI ( <i>n</i> = 9)		FN ( <i>n</i> = 11)	
	<i>M</i>	SD	<i>M</i>	SD	<i>M</i>	SD	<i>M</i>	SD
1	210.74	112.97	282.68	256.86	246.02	158.61	221.68	168.81
2	149.46	53.01	291.54	264.60	285.94	186.17	243.91	250.13
3	169.99	77.23	291.09	258.53	277.03	136.19	261.29	274.81
4	152.48	84.19	305.11	301.58	280.45	180.37	256.74	270.92
5	114.21	48.11	241.64	245.63	231.71	165.41	260.26	239.79
6	117.67	69.05	208.23	186.94	211.02	113.06	256.54	250.16
7	102.42	42.95	195.70	170.08	204.40	158.18	203.37	203.06
8	107.81	52.91	212.01	211.28	192.04	156.56	221.94	177.95
9	100.42	73.75	229.80	239.71	204.46	151.16	202.09	103.57
10	99.25	48.90	239.01	231.33	176.84	91.48	232.43	172.46
High initial level								
	SI ( <i>n</i> = 10)		SN ( <i>n</i> = 6)		FI ( <i>n</i> = 11)		FN ( <i>n</i> = 10)	
	<i>M</i>	SD	<i>M</i>	SD	<i>M</i>	SD	<i>M</i>	SD
1	104.93	33.96	83.19	12.81	125.89	31.80	98.18	28.04
2	94.60	21.51	74.25	12.42	128.69	64.27	94.20	29.57
3	83.81	31.72	76.06	13.96	90.22	20.64	88.81	26.14
4	83.26	24.46	68.26	9.04	87.60	31.15	89.72	43.00
5	75.33	32.09	68.73	20.72	85.27	16.90	80.59	36.12
6	65.45	17.76	76.81	16.02	78.89	19.58	84.95	55.16
7	74.26	22.70	74.79	38.86	71.94	18.56	76.66	26.52
8	66.56	27.61	58.80	9.88	71.49	20.00	92.89	41.76
9	63.52	35.13	57.00	14.42	80.05	48.04	80.78	47.31
10	59.20	17.00	61.38	18.42	78.05	36.99	80.95	39.38

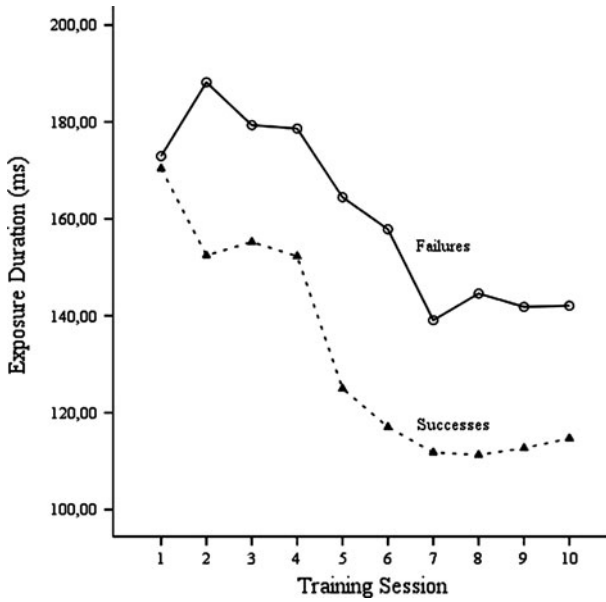
SI = Training focused on successes and children informed about this focus, SN = Training focused on successes and children not informed about this focus, FI = Training focused on failures and children informed about this focus, FN = Training focused on failures and children not informed about this focus

illustrated in Fig. 4, for children who are informed the exposure duration decreased more over sessions than for children who are uninformed.

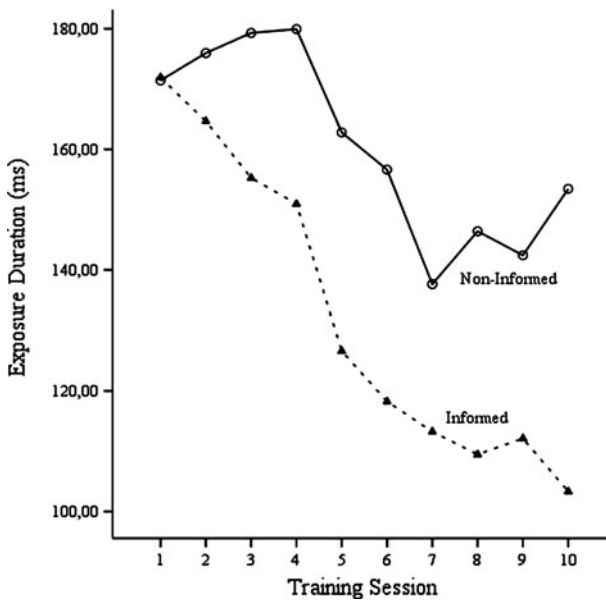
### Transfer measures of word reading skills

#### *Reading accuracy*

A repeated measures MANOVA was conducted with the accuracy scores on the reading tests as the dependent variables. Accuracy for each test was determined as the percentage of correct responses and is displayed in Table 5.



**Fig. 3** Mean exposure duration (in milliseconds) in each training session for children in the training focused on failures and successes



**Fig. 4** Mean exposure duration (in milliseconds) of each training session for children being informed about the focus of the training, and children being uninformed

**Table 5** Reading accuracy (as percentage correct) at pretest and posttest of the Lexical Decision Test (LDT1 and LDT2), Word Decoding Test (WDT1, WDT2 and WDT3), Nonword Reading Test (NRT) and Sentence Verification Test (SVT)

Low initial level																
SI ( <i>n</i> = 10)			SN ( <i>n</i> = 12)			FI ( <i>n</i> = 9)			FN ( <i>n</i> = 11)							
Pretest		Posttest	Pretest		Posttest	Pretest		Posttest	Pretest		Posttest					
<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>					
LDT1	84.04	14.41	85.21	9.61	83.05	6.01	75.62	19.43	81.05	13.40	78.74	9.04	88.64	10.96	79.54	10.44
LDT2	82.73	17.91	80.53	13.70	84.53	13.63	73.38	14.95	82.71	18.24	77.85	14.02	87.92	15.06	70.35	25.53
WDT1	78.74	13.19	81.11	11.11	82.85	8.88	80.22	15.84	79.93	9.45	81.17	15.96	70.65	20.13	76.21	15.71
WDT2	68.19	22.90	67.86	14.93	66.54	25.60	57.10	27.81	58.46	22.23	70.41	18.88	59.73	25.89	65.34	20.31
WDT3	41.59	25.48	55.47	17.42	44.72	20.75	45.41	26.83	45.19	31.24	38.27	28.31	47.09	32.39	48.42	22.54
NRT	36.11	14.72	41.04	21.98	40.62	22.74	30.94	23.04	31.07	18.76	37.13	20.38	35.21	23.18	37.89	20.65
SVT	83.56	9.26	85.56	9.84	77.96	13.94	79.63	10.84	83.65	5.60	84.44	5.44	80.60	10.23	85.65	8.39
High initial level																
SI ( <i>n</i> = 10)			SN ( <i>n</i> = 6)			FI ( <i>n</i> = 11)			FN ( <i>n</i> = 10)							
Pretest		Posttest	Pretest		Posttest	Pretest		Posttest	Pretest		Posttest					
<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>					
LDT1	88.49	9.22	88.96	9.41	96.18	3.56	88.59	7.74	91.60	5.83	90.62	8.14	90.19	9.26	90.80	7.06
LDT2	94.41	5.86	85.63	14.67	98.38	2.55	93.06	5.09	93.96	5.53	93.50	3.57	94.94	5.24	93.71	6.96
WDT1	90.64	7.16	94.50	3.59	93.14	6.16	92.86	6.55	91.02	6.57	92.40	5.98	94.29	3.32	97.03	2.62
WDT2	83.86	16.63	87.35	10.44	93.49	7.31	87.03	5.98	85.20	9.58	88.38	9.02	91.88	5.02	91.79	4.94
WDT3	73.84	14.54	73.59	17.46	75.15	22.69	72.44	12.39	71.34	19.84	75.53	13.61	74.91	9.95	80.98	6.42
NRT	51.12	15.96	58.62	23.48	62.51	6.07	57.90	14.05	50.76	19.65	52.78	21.00	54.58	5.64	57.99	12.46

**Table 5** continued

High initial level															
SI ( <i>n</i> = 10)		SN ( <i>n</i> = 6)		FI ( <i>n</i> = 11)		FN ( <i>n</i> = 10)									
Pretest		Pretest		Pretest		Pretest									
<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>								
87.78	8.13	92.44	7.50	91.11	3.44	94.44	2.72	90.30	4.01	91.71	5.36	87.78	10.90	95.56	2.96
Posttest		Posttest		Posttest		Posttest									
<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>

SI = Training focused on successes and children informed about this focus, SN = Training focused on successes and children not informed about this focus, FI = Training focused on failures and children informed about this focus, FN = Training focused on failures and children not informed about this focus



An effect of Time was observed ( $F(7, 65) = 5.84, p < .001, \eta_p^2 = .39$ ) indicating that accuracy improved from pre- to posttest. An effect of Initial Level was observed ( $F(7, 65) = 7.38, p < .001, \eta_p^2 = .44$ ), indicating that HI children have higher accuracy scores than LI children. This holds for all tests ( $p < .001$ ). No other main effects were observed ( $F < 1$ ). No significant interactions were observed (Time by Information ( $F(7, 65) = 1.34, p > .05, \eta_p^2 = .13$ ); Time by Focus by Information ( $F(7, 65) = 1.34, p > .05, \eta_p^2 = .13$ ); Time by Focus by Initial Level ( $F(7, 65) = 1.74, p > .05, \eta_p^2 = .16$ ); all other interactions  $F < 1$ ).

The effect of Time was significant for LDT1 ( $F(1,71) = 4.16, p < .05, \eta_p^2 = .055$ ), LDT2 ( $F(1,71) = 12.98, p < .001, \eta_p^2 = .16$ ), and SVT ( $F(1,71) = 13.35, p < .001, \eta_p^2 = .16$ ), indicating that accuracy improved for these reading tests only. No improvement in accuracy was observed for the other tests (WDT1 ( $F(1,71) = 2.68, p > .05, \eta_p^2 = .04$ ), WDT2, WDT3 and NRT  $F < 1$ ). Interestingly, the LDTs and the SVT are tests in which the child needs to judge whether an item (word or sentence) is meaningful or not.

These results on reading accuracy indicate that children improve their reading accuracy on the two Lexical Decision Tests and the Sentence Verification Test. No differential effects of training condition on reading accuracy improvement were found, however.

### Reading speed

A repeated measures MANOVA was conducted with reading speed as the dependent variable. Reading speed was determined as the number of words read within 1 min for the LDT, WDT and NWR. For the SVT, reading speed was determined as the log median latency time over the semantically correct sentences. Descriptives are displayed in Table 6.

A main effect of Time was observed ( $F(7, 65) = 22.88, p < .001, \eta_p^2 = .71$ ), indicating that children improved their reading speed from pre- to posttest. A main effect of Initial Level ( $F(7, 65) = 13.66, p < .001, \eta_p^2 = .60$ ) indicates that HI children read faster than LI children, as expected. No other main effects were observed (Focus ( $F(7, 65) = 1.32, p > .05, \eta_p^2 = .12$ ), Information ( $F(7, 65) = 1.82, p > .05, \eta_p^2 = .16$ ). Interestingly, the interaction of Time by Focus by Initial Level ( $F(7, 65) = 2.73, p < .05, \eta_p^2 = .23$ ) was significant. No other interactions were present (Time by Initial Level ( $F(7, 65) = 1.36, p > .05, \eta_p^2 = .13$ ); Time by Information by Initial Level ( $F(7, 65) = 1.75, p > .05, \eta_p^2 = .16$ ); Time by Focus by Information by Initial Level ( $F(7, 65) = 1.54, p > .05, \eta_p^2 = .14$ ); all other interactions  $F < 1$ ).

The effect of Time was significant for each reading test, indicating that children improved their reading speed on all reading tests: LDT1 ( $F(1,71) = 22.23, p < .001, \eta_p^2 = .24$ ), LDT2 ( $F(1,71) = 21.99, p < .001, \eta_p^2 = .24$ ), WDT1 ( $F(1,71) = 73.10, p < .001, \eta_p^2 = .51$ ), WDT2 ( $F(1,71) = 43.61, p < .001, \eta_p^2 = .38$ ), WDT3 ( $F(1,71) = 35.50, p < .001, \eta_p^2 = .33$ ), NRT ( $F(1,71) = 14.54, p < .001, \eta_p^2 = .17$ ), and SVT ( $F(1,71) = 87.00, p < .001, \eta_p^2 = .55$ ).

The interaction of Time by Focus by Initial Level was significant for LDT2 ( $F(1,71) = 4.05, p < .05, \eta_p^2 = .05$ ), WDT3 ( $F(1,71) = 10.13, p < .01, \eta_p^2 = .13$ ),

**Table 6** Reading speed at pretest and posttest measured as number of words read per minute of the Lexical Decision Test (LDT1 and LDT2) and Word Decoding Test (WDT1, WDT2 and WDT3); as number of words read per 2 min of the Nonword Reading Test (NRT); and as latency time (in MS) for the Sentence Verification Test (SVT)

Low initial level																
	SI (n = 10)			SN (n = 12)			FI (n = 9)			FN (n = 11)						
	Pretest		Posttest	Pretest		Posttest	Pretest		Posttest	Pretest		Posttest				
	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD				
LDT1	15.10	4.28	17.80	4.94	15.67	7.76	17.83	7.20	12.44	6.06	21.00	17.26	16.18	4.56	18.09	7.57
LDT2	12.80	2.53	14.10	5.09	13.25	6.86	16.67	8.34	9.78	4.74	22.44	25.58	13.55	6.25	19.73	12.31
WDT1	32.50	5.87	41.80	9.09	32.67	10.88	43.33	13.83	27.00	7.43	35.00	11.52	31.45	7.09	39.09	12.28
WDT2	19.00	5.08	28.00	8.88	20.42	7.83	27.58	11.18	16.78	6.16	22.56	9.44	20.82	6.49	25.73	8.89
WDT3	14.10	3.48	20.80	7.42	15.33	6.83	22.75	8.53	11.44	4.22	15.44	8.05	14.18	6.00	16.91	8.12
NRT	28.10	8.63	37.20	14.63	24.42	7.00	34.42	13.53	29.67	15.72	29.22	8.83	27.27	4.15	30.55	8.30
SVT	7,496	1,519	7,025	2,346	8,398	4,794	7,047	5,004	8,795	4,170	9,128	4,464	8,609	3,166	7,478	3,132

High initial level																
	SI (n = 10)			SN (n = 6)			FI (n = 11)			FN (n = 10)						
	Pretest		Posttest	Pretest		Posttest	Pretest		Posttest	Pretest		Posttest				
	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD				
LDT1	27.00	6.25	32.90	13.63	27.17	5.85	31.67	7.34	24.36	7.20	32.18	14.34	22.90	7.39	32.40	7.26
LDT2	21.40	8.00	33.30	13.38	23.83	7.14	30.33	9.27	21.18	7.40	28.64	13.40	22.40	7.79	25.10	5.76
WDT1	54.10	9.59	57.60	12.95	65.00	19.87	74.50	25.31	54.36	12.30	63.45	12.39	51.70	10.73	61.70	13.00
WDT2	37.00	9.79	41.80	11.32	49.17	19.41	54.50	30.40	36.73	14.01	44.00	11.20	39.30	13.02	49.00	12.27
WDT3	28.80	5.18	32.30	8.55	40.17	25.96	38.17	18.30	28.27	12.03	34.45	12.88	27.40	7.20	34.20	8.00
NRT	40.50	9.96	42.20	16.27	50.50	20.71	55.50	33.31	38.27	14.60	50.73	23.05	36.80	7.83	40.30	7.69
SVT	5,264	822	4,695	878	5,023	485	5,585	1,300	4,744	1,313	4,341	922	4,646	848	4,688	925

SI = Training focused on successes and children informed about this focus, SN = Training focused on successes and children not informed about this focus, FI = Training focused on failures and children informed about this focus, FN = Training focused on failures and children not informed about this focus

and NRT ( $F(1,71) = 4.77, p < .05, \eta_p^2 = .06$ ). Other tests did not show this interaction (WDT1 ( $F(1,71) = 1.73, p > .05, \eta_p^2 = .02$ ); WDT2 ( $F(1,71) = 2.27, p > .05, \eta_p^2 = .03$ ); LDT1 and SVT both  $F < 1$ ). The interaction of Time by Focus by Initial Level observed in LDT2, WDT3 and NRT shows an ATI effect. In the groups training on successes, LI children improved their reading speed more than HI children (post hoc tests with Bonferroni correction, LDT2:  $p < .001$ , WDT3:  $p < .001$ , NRT:  $p < .01$ ). And, conversely, in the groups training on failures, HI children improved more than LI children (LDT2:  $p < .05$ , WDT3:  $p < .001$ , NRT:  $p < .05$ ). Thus, a training on failures is most beneficial for HI children, and conversely, a training focused on successes is most beneficial for LI children. Crucially, as the previous part of the results section pointed out, the training groups did *not* show any differential effect on reading accuracy. Improved reading speed evidently did not go together with a detrimental effect on reading accuracy. In other words, no speed-accuracy trade-off is observed as a result of the flashcard training focusing either on successes or on failures.

### *Comparison with normal reading improvement*

The pre- and posttest scores of the LDTs, WDTs and NRT were classified into five levels: below the 10th percentile, the next 15%, and each of the next quartiles. The Wilcoxon Signed Ranks Test indicated that children improved from pre- to posttest by scoring in a higher norm class on the LDT1 ( $Z = -2.14, p < .05$ ) and WDT1 ( $Z = -2.11, p < .05$ ) (On LDT2, WDT2, WDT3 and NRT progress was not significant.). This suggests that for poor readers, intervention focused on CVC words was beneficial as children improved their reading of untrained monosyllabic words more than can be expected from regular education.

## **Discussion**

The present study compares the effects of two training procedures in poor readers. A randomized controlled trial design was used to answer the questions (a) whether a training focused on failures versus a training focused on successes differentially improve reading speed and accuracy, (b) whether the effect of training focus interacts with the effect of being informed or not about the training focus, and (c) whether such training effects interact with the child's initial reading level. The interventions were focused on reading speed, while maintaining an approximately constant level of reading accuracy. The training material consisted of regularly spelled Dutch CVC words and outcome transfer measures consisted of a Lexical Decision Test (LDT), Word Decoding Test (WDT), Nonword Reading Test (NRT), and a Sentence Verification Test (SVT).

The main findings of the pretest–posttest comparison are as follows. An improvement of *reading speed* was observed for all reading tests. Improvement was characterized by an ATI effect (Cronbach & Snow, 1977; Snow, 1991). The HI children among the poor readers improved their reading speed most by the training focused on failures, and conversely, the LI children in the training focused on

successes. It should be noted that the ATI effect was restricted to the reading speed of untrained words with an orthographical structure that can be characterized as more complex than the orthographical structure of the trained words (observed in LDT2, WDT3, and NRT that is composed primarily of polysyllabic words). A possible cause for this immediate transfer to polysyllabic words may be a more efficient syllable processing. Poor readers tend to read longer words letter-by-letter. Possibly, as a result of the CVC training, the poor readers progressively shifted towards a more syllable-bound decoding strategy. Such a shift after training is evidenced for Dutch poor readers by Wentink, van Bon, and Schreuder (1997) (see also Huemer, Aro, Landerl, & Lyytinen, 2010). An ATI effect was not found for words of the CVC structure used in the training. Improved reading competence in a rather transparent orthography as Dutch is mostly a matter of increased speed (Verhoeven & van Leeuwe, 2009) and low reading speed is an important characteristic of Dutch poor readers (van der Leij & van Daal, 1989). The transfer of increased reading speed to untrained words in our study, therefore, is promising.

*Reading accuracy*, on the other hand, improved equally for all children on CVC word reading in LDT1, on the LDT2, which uses bisyllabic words, and intermediately transferred to the SVT, which uses high-frequency monosyllabic words. Interestingly, increases in accuracy included exactly those tests (lexical decision and sentence verification) that assess semantic processing. This might reflect that the training enhanced the connection of semantic properties to orthographic word features (see Ehri, 1998, 1999, 2005). During training, the LI and HI children showed different growth trajectories in accuracy. The HI children improved slightly over all sessions, whereas the LI children showed initial stagnation after which they increased in accuracy. During training immediate feedback was provided on whether an item had been correct. It is an issue of future research to verify whether children show steeper improvement if scaffolding or corrective feedback is provided. Importantly, no significant speed-accuracy trade-off occurred as an effect of the training. Evidently, gains in reading speed did not lead to a loss of accuracy.

The ATI effect found for reading speed implies that the two intervention approaches *interact* with the children's reading level at the start of the training. It can be concluded, therefore, that neither of the two intervention approaches is superior to the other in general (cf. related studies of Eckert, Dunn, & Ardoin, 2006; and Worsdell, Iwata, Dozier, Johnson, Neidert, & Thomason, 2005), but that the best approach depends on the reading level of the child.

As an alternative approach to assess whether our intervention had additional practice value, we compared the improvement in the training groups to normed scores. Results showed that the training groups improved more on the monosyllabic words that were practiced (LDT1 and WDT1) than can be expected from normal reading education. This suggests that the intervention was beneficial. Comparisons of the training groups to customary improvement could not be made on the non-normed SVT. However, the effect sizes for this test were large, which might point towards a similar improvement on an intermediate transfer to sentence comprehension as well.

To enhance fast and effortless reading for the LI children, training should focus on words of a type—CVC, in this case—they are able to read with sufficient accuracy. As the analysis of the stimuli in the last training set shows, these words

can be considered representative for many other words of the same orthographic structure. They have sublexical units with a high chance of occurrence in the language, are characterized by high bigram frequency, and they typically have many and high-frequency orthographic and phonological neighbors. Also, children are likely to re-encounter these words in later reading, as they are characterized by a high frequency of use. Therefore, mastering these words reflects mastering a common core of (CVC) reading material. Improved competence in reading this common core apparently transfers to improved reading speed of words that are orthographically more complex. In contrast, for the HI children, the training approach focused on failures is the most effective one. Children in this training group have been practicing on words that can be viewed as being in the periphery of the CVC common core. These words have a low probability of occurrence in the language and their constituent sublexical units occur less in other words. If HI children focus on these rather uncommon and less familiar words, reading speed improves and transfers to untrained words that are orthographically more complex than the words they have practiced. For the HI children, these CVC words may have been more of a challenge to read and probably are in their zone of proximal development (Vygotsky, 1978). They may involve practicing at a level of difficulty that is neither too high nor too low, the level at which optimal learning takes place. Perhaps, practicing on successes was too easy for the HI children, whereas practicing on failures was too hard for the LI children. Surprisingly, the training on failures did not prove to be most beneficial for the LI children, as would be predicted by Ehri (1998, 1999). Rather, it seems that repeatedly practicing successes increases item-based knowledge (Share, 1995, 1999). Improving reading speed in the HI children by practicing on the more demanding words is consistent with expert learning theories of Gobet (2005) and Ericsson (2004), who show that performance can be optimized by continued practice on failures accompanied by detailed feedback. Our results also partly underscore Podsakoff and Farh's (1989) conclusion that goals that are hard to reach lead to better results than goals that are easy to meet.

The content of the different trainings was further investigated by exploring the characteristics of the words in the last training set. Analyses on the number of words show that the ATI effect is not explained by the *number* of different words children have practiced with, because children with the fewest words in their last training set, be it successes or failures, do not consistently show the greatest improvement in reading speed of untrained words. Thus, improvement of general reading speed is *not* primarily related to word specific training effects. This differs from a conclusion by Berends and Reitsma (2006) that the practical value of repeated reading with Dutch poor readers lies in its word specific training effects. In contrast to Berends and Reitsma, who only found effects on trained monosyllabic words with consonant clusters and no transfer effects to untrained words, our study leads to the contrary conclusion, that decoding many different words transfers to untrained words, probably by improving skill in applying GPC rules. The explanation for the contrasting outcomes perhaps is, that it is not so much the *number* of words in training that is crucially for improving general reading skills, but rather *which* words children practice with: either their failures or successes.

The training data suggest that informing students about the focus of the training positively affects training: Exposure durations for children informed about the focus of the training decreased more than for children who were not informed. This is in line with studies of Swanson et al. (1999), Conte and Hintze (2000), and Kluger and DeNisi (1998) who stress the importance of goal-setting for students in education. Possibly, the informed children increased their learning not only through positive or negative reinforcement, but also through a clearer metacognitive focus on the goal of the training. However, the effect of informing children was not found in the pretest–posttest comparison. The additional effects of informing children apparently were not robust or powerful enough for a transfer to general reading performance.

It should be noted that our training improved reading speed at the word level and intermediately transferred to the sentence level. Improved reading speed at the sentence level as a result of a training targeted at the sentence level has been reported for poor readers (Breznitz, 2006; Karni et al., 2005; Snellings, van der Leij, de Jong, & Blok, 2009). Our study goes beyond these sentence-focused trainings, as our focus on the word level affects reading speed at the sentence level as well. Future research should verify whether our training brings about not only intermediate transfer to sentence comprehension, but a far transfer to passage reading and text comprehension as well.

Our study focused on increasing reading speed, because a low reading speed is a prime characteristic of reading disorders in languages with a transparent orthography. Children with reading disorders in opaque orthographies generally suffer more from a low accuracy. Moreover, Ehri (2005) indicated that children reading transparent orthographies seem to progress faster through developmental phases of reading fluency than children reading opaque orthographies. Whether children reading in an opaque language benefit from a differential training like we observed, is an issue for future research.

Our study suggests that the improvement of general reading speed in a transparent orthography is closely related to both the type of words children practice with (common and familiar words vs. uncommon and less familiar words) and to their initial reading level. The training approach that focuses on CVC words that are representative for many other CVC words, is the most effective for the poor readers with a low initial reading level. These readers constitute only a small number of the population at large: We selected the participating students with a reading score below the 10th percentile and thereafter split the group in half. Thus, our poor readers with a low initial reading level would constitute only 5% of the population. It is these children, however, who are the readers that deserve the most effort to improve their reading skill.

**Acknowledgments** We are grateful to Drs. Hubert Voogd for his assistance in creating computer software for the training. We also thank Drs. Ans Bouten, Drs. Marjan Feenstra and Drs. Jorita Bruining for their assistance in data collection.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution Non-commercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

## References

- Adams, M. J. (1990). *Beginning to read: Thinking and learning about print*. Cambridge, MA: MIT Press.
- Allen, P. D. (1976). Some general implications concerning specific taxonomy categories. In P. D. Allen & D. J. Watson (Eds.), *Findings of research in miscue analyses: Classroom implications* (pp. 70–93). Urbana: ERIC Clearinghouse on Reading and Communication Skills.
- Allerup, P., & Elbro, C. (1998). Comparing differences in accuracy across conditions or individuals: An argument for the use of log odds. *The Quarterly Journal of Experimental Psychology*, *51A*, 409–424.
- Aro, M., & Wimmer, H. (2003). Learning to read: English in comparison to six more regular orthographies. *Applied Psycholinguistics*, *24*, 621–635.
- Au, K. H. (1977). Analyzing oral reading errors to improve instruction. *The Reading Teacher*, *31*, 46–49.
- Baayen, R. H., Piepenbrock, R., & van Rijn, H. (1993). *The CELEX lexical database (CD-ROM) [Computer Software]*. Philadelphia, PA: Linguistic Data Consortium, University of Pennsylvania.
- Beard, R. (1990). *Developing reading 3–13* (2nd ed.). London: Hodder & Stoughton.
- Bender, W. N. (2004). *Learning disabilities. Characteristics, identification, and teaching strategies*. Boston, MA: Pearson.
- Berends, I. E., & Reitsma, P. (2006). Remediation of fluency: Word specific or generalised training effects? *Reading and Writing: An Interdisciplinary Journal*, *19*, 221–234.
- Booij, G. (1995). *The phonology of Dutch*. Oxford: Clarendon Press.
- Bracht, G. H. (1970). Experimental factors related to aptitude-treatment interactions. *Review of Educational Research*, *40*, 627–645.
- Breznitz, Z. (2006). *Fluency in reading: Synchronization of processes*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Caravolas, M. (2007). The nature and causes of dyslexia in different languages. In M. J. Snowling & C. Hulme (Eds.), *The science of reading: A Handbook* (pp. 336–355). Oxford: Blackwell Publishing.
- Chard, D. J., Vaughn, S., & Tyler, B. (2002). A synthesis of research on effective interventions for building reading fluency with elementary students with learning disabilities. *Journal of Learning Disabilities*, *35*, 386–406.
- Conte, K. L., & Hintze, J. M. (2000). The effects of performance feedback and goal setting on oral reading fluency within curriculum-based measurement. *Diagnostique*, *25*, 85–98.
- Cronbach, L., & Snow, R. (1977). *Aptitudes and instructional methods: A handbook for research on interactions*. New York, NY: Irvington.
- de Jong, P. F., & van der Leij, A. (2003). Developmental changes in the manifestation of a phonological deficit in dyslexic children learning to read a regular orthography. *Journal of Educational Psychology*, *95*, 22–40.
- Eckert, T. L., Dunn, E. K., & Ardoin, S. P. (2006). The effects of alternate forms of performance feedback on elementary-aged students' oral reading fluency. *Journal of Behavioral Education*, *15*, 149–162.
- Ehri, L. C. (1998). Grapheme-phoneme knowledge is essential for learning to read words in English. In J. L. Metsala & L. C. Ehri (Eds.), *Word recognition in beginning literacy* (pp. 3–40). Mahwah, NJ: Erlbaum.
- Ehri, L. C. (1999). Phases of development in learning to read words. In J. Oakhill & R. Beard (Eds.), *Reading development and the teaching of reading* (pp. 79–108). Oxford: Blackwell.
- Ehri, L. C. (2005). Learning to read words: Theory, findings, and issues. *Scientific Studies of Reading*, *9*, 167–189.
- Ellis, N. C., Natsume, M., Stavropoulou, K., Hoxhallari, L., van Daal, V. H. P., Polyzoe, N., et al. (2004). The effects of orthographic depth on learning to read alphabetic, syllabic, and logographic scripts. *Reading Research Quarterly*, *39*, 438–468.
- Ericsson, K. A. (2004). Deliberate practice and the acquisition and maintenance of expert performance in medicine and related domains. *Academic Medicine*, *79*(10 suppl.), S70–S81.
- Eurybase. (2008). *Organization of the education system in the Netherlands*. Retrieved June 4, 2010, from Eurydice Web site: [http://eacea.ec.europa.eu/education/eurydice/eurybase\\_en.php](http://eacea.ec.europa.eu/education/eurydice/eurybase_en.php).
- Frith, U., Wimmer, H., & Landerl, K. (1998). Differences in phonological recoding in German- and English-speaking children. *Scientific Studies of Reading*, *2*, 31–54.
- Gobet, F. (2005). Chunking models of expertise: Implications for education. *Applied Cognitive Psychology*, *19*, 183–204.
- Goodman, K. S. (1969). Analysis of oral reading miscues: Applied psycholinguistics. *Reading Research Quarterly*, *5*, 9–30.



- Graham, J., & Kelly, A. (Eds.). (1997). *Reading under control: Teaching reading in the primary school*. London: David Fulton.
- Hall, K. (2003). *Listening to Stephen read: Multiple perspectives on literacy*. Buckingham: Open University Press.
- Hinton, G. E., McClelland, J. L., & Rumelhart, D. E. (1986). Distributed representations. In D. E. Rumelhart, J. L. McClelland, & The PDP Research Group (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition. Vol. 1: Foundations* (pp. 77–109). Cambridge, MA: MIT Press.
- Hudson, R. F., Pullen, P. C., Lane, H. B., & Torgesen, J. K. (2009). The complex nature of reading fluency: A multidimensional view. *Reading and Writing Quarterly*, 25, 4–32.
- Huemer, S., Aro, M., Landerl, K., & Lyytinen, H. (2010). Repeated reading of syllables among Finnish-speaking children with poor reading skills. *Scientific Studies of Reading*, 14, 317–340.
- Jenkins, J. R., Fuchs, L. S., van den Broek, P., Espin, C. L., & Deno, S. L. (2003). Sources of individual differences in reading comprehension and reading fluency. *Journal of Educational Psychology*, 95, 719–729.
- Kanfer, R., & Ackerman, P. L. (1989). Motivation and cognitive abilities: An integrative/aptitude-treatment interaction approach to skill acquisition. *Journal of Applied Psychology*, 74, 657–690.
- Karni, A., Morocz, I. A., Bitan, T., Shaul, S., Kushnir, T., & Breznitz, Z. (2005). An fMRI study of the differential effects of word presentation rates (reading acceleration) on dyslexic readers' brain activity patterns. *Journal of Neurolinguistics*, 18, 197–219.
- Keselman, H. J., Huberty, C. J., Lix, L. M., Olejnik, S., Cribbie, R. A., Donahue, B., et al. (1998). Statistical practices of educational researchers: An analysis of their ANOVA, MANOVA, and ANCOVA analyses. *Review of Educational Research*, 68, 350–386.
- Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin*, 119, 254–284.
- Kluger, A. N., & DeNisi, A. (1998). Feedback interventions: Toward the understanding of a double-edged sword. *Current Directions in Psychological Science*, 7(3), 67–72.
- Kohnstamm, G. A., Schaeerlaekens, A. M., de Vries, A. K., Akkerhuis, G. W., & Frooninckxs, M. (1999). *Nieuwe streeflijst woordenschat voor 6-jarigen gebaseerd op onderzoek in Nederland en België (Herziene versie) [New vocabulary target list for 6-year-olds based on research in the Netherlands and Belgium (revised version)]*. Lisse: Swets & Zeitlinger.
- Kuhn, M. R., Schwanenflugel, P. J., & Meisinger, E. B. (2010). Aligning theory and assessment of reading fluency: Automaticity, prosody, and definitions of fluency. *Reading Research Quarterly*, 45, 230–251.
- Kuhn, M. R., & Stahl, S. A. (2003). Fluency: A review of developmental and remedial practices. *Journal of Educational Psychology*, 95, 3–21.
- Landerl, K. (2001). Word recognition deficits in German: More evidence from a representative sample. *Dyslexia*, 7, 183–196.
- Latham, G. P. (1989). Behavioral approaches to the training and learning process. In I. L. Goldstein (Ed.), *Training and development in organizations* (pp. 256–295). San Francisco, CA: Jossey-Bass.
- Locke, E. A., & Latham, G. P. (2002). Building a practically useful theory of goal setting and task motivation. *American Psychologist*, 57, 705–717.
- Martens, V. E. G., & de Jong, P. F. (2008). Effects of repeated reading on the length effect in word and pseudoword reading. *Journal of Research in Reading*, 31, 40–54.
- Martens, B. K., Witt, J. C., Daly, E. J., & Vollmer, T. (1999). Behavior analysis: Theory and practice in educational settings. In C. R. Reynolds & T. B. Gutkin (Eds.), *Handbook of school psychology* (3rd ed., pp. 638–663). New York, NY: Wiley.
- McKenna, M. C., & Picard, M. (2006). Revisiting the role of miscue analysis in effective teaching. *The Reading Teacher*, 60, 378–380.
- Meyer, M. S., & Felton, R. H. (1999). Repeated reading to enhance fluency: Old approaches and new directions. *Annals of Dyslexia*, 49, 283–306.
- National Reading Panel. (2000). *Report of the national reading panel. Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications of reading instruction: Reports of the subgroups* (NIH Publication No. 00-4754). Washington, DC: US Government Printing Office.
- Nunn, A. N. (1998). *Dutch orthography: A systematic investigation of the spelling of Dutch words*. The Hague, The Netherlands: Holland Academic Graphics.

- Patel, T. K., Snowling, M. J., & de Jong, P. F. (2004). A cross-linguistic comparison of children learning to read in English and Dutch. *Journal of Educational Psychology, 96*, 785–797.
- Perfetti, C. A. (1992). The representation problem in reading acquisition. In P. B. Gough, L. C. Ehri, & R. Treiman (Eds.), *Reading acquisition* (pp. 145–174). Hillsdale, NJ: Erlbaum.
- Podsakoff, P. M., & Farh, J. L. (1989). Effects of feedback sign and credibility on goal setting and task performance. *Organizational Behavior and Human Decision Processes, 44*, 45–67.
- Reitsma, P. (1983). Printed word learning in beginning readers. *Journal of Experimental Child Psychology, 75*, 321–339.
- Riddick, B. (2010). *Living with dyslexia: The social and emotional consequences of specific learning difficulties/disabilities* (2nd ed.). London: Routledge.
- Roberts, G. (1989). *Teaching children to read and write*. Oxford: Blackwell.
- Savage, R. S., Stuart, M., & Hill, V. (2001). The role of scaffolding errors in reading development: Evidence from a longitudinal and a correlational study. *British Journal of Educational Psychology, 71*, 1–13.
- Scheltinga, F., van der Leij, A., & Struiksma, C. (2010). Predictors of response to intervention of word reading fluency in Dutch. *Journal of Learning Disabilities, 43*, 212–228.
- Schwanenflugel, P. J., Meisinger, E. B., Wisenbaker, J. M., Kuhn, M. R., Strauss, G. P., & Morris, R. D. (2006). Becoming a fluent and automatic reader in the early elementary school years. *Reading Research Quarterly, 41*, 496–522.
- Serrano, F., & Defior, S. (2008). Dyslexia speed problems in a transparent orthography. *Annals of Dyslexia, 58*, 81–95.
- Seymour, P. H. K., Aro, M., & Erskine, J. M. (2003). Foundation literacy acquisition in European orthographies. *British Journal of Psychology, 94*, 143–174.
- Share, D. L. (1995). Phonological recoding and self-teaching: *Sine qua non* of reading acquisition. *Cognition, 55*, 151–218.
- Share, D. L. (1999). Phonological recoding and orthographic learning: A direct test of the self-teaching hypothesis. *Journal of Child Psychology, 72*, 95–129.
- Shavelson, R. J., Berliner, D. C., Ravitch, M. M., & Loeding, D. (1974). Effects of position and type of question on learning from prose material: Interaction of treatments with individual differences. *Journal of Educational Psychology, 66*, 40–48.
- Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research, 78*, 153–189.
- Singleton, C. (2005). Dyslexia and oral reading errors. *Journal of Research in Reading, 28*, 4–14.
- Snellings, P., van der Leij, A., de Jong, P. F., & Blok, H. (2009). Enhancing the reading fluency and comprehension of children with reading disabilities in an orthographically transparent language. *Journal of Learning Disabilities, 42*, 291–305.
- Snow, R. E. (1991). Aptitude-treatment interaction as a framework for research on individual differences in psychotherapy. *Journal of Consulting and Clinical Psychology, 59*, 205–216.
- Steenbeek-Planting, E. G., van Bon, W. J. H., & Schreuder, R. (2011). Instability of word reading errors of typical and poor readers.
- Swanson, H. L. (1999). Reading research for students with LD: A meta-analysis of intervention outcomes. *Journal of Learning Disabilities, 32*, 504–532.
- Swanson, H. L., Hoskyn, M., & Lee, C. (1999). *Interventions for students with learning disabilities: A meta-analysis of treatment outcomes*. New York, NY: Guilford.
- Torgesen, J. K., Rashotte, C. A., & Alexander, A. (2001). Principles of fluency instruction in reading: Relationships with established empirical outcomes. In M. Wolf (Ed.), *Dyslexia, fluency and the brain* (pp. 333–355). Parkton, MD: York.
- Vadasy, P. F., & Sanders, E. A. (2009). Supplemental fluency intervention and determinants of reading outcomes. *Scientific Studies of Reading, 13*, 383–425.
- van Bon, W. H. J. (2007). *De Doorstreepleestoets [Paper-and-pen lexical decision task]*. Leiden, The Netherlands: PITS.
- van Bon, W. H. J., Bouwmans, M., & Broeders, I. N. L. D. C. (2006). The prevalence of poor reading in Dutch special elementary education. *Journal of Learning Disabilities, 39*, 482–495.
- van Bon, W. H. J., Hoevenaars, L. T. M., & Jongeneelen, J. J. (2004). Using paper-and-pencil lexical decision tests to assess word decoding skill: Aspects of validity and reliability. *Journal of Research in Reading, 27*, 58–68.
- van den Bos, K. P., Lutje Spelberg, H. C., Scheepstra, A. J. M., & De Vries, J. R. (1994). *De Klepel-Verantwoording, Diagnostiek en Behandeling [Nonword Reading Test]*. Nijmegen, The Netherlands: Berkhout.

- van den Bosch, K., van Bon, W. H. J., & Schreuder, R. (1995). Poor readers' decoding skills: Effects of training with limited exposure duration. *Reading Research Quarterly*, *30*, 110–125.
- van der Leij, A., & van Daal, V. H. P. (1989). Repeated reading and severe reading disability. In E. Mandl, E. de Corte, N. Bennett, & H. F. Friedrich (Eds.), *Learning and instruction* (Vol. 2.2, pp. 235–251). Oxford: Pergamon Press.
- Verhoeven, L. T. W. (1995). *Drie-Minuten-Toets [Three-Minutes-Test]*. Arnhem, The Netherlands: Cito.
- Verhoeven, L. T. W., & van Leeuwe, J. (2003). Ontwikkeling van decodeervaardigheid in het basisonderwijs [Development of decoding skills in elementary school]. *Pedagogische Studiën*, *80*, 257–271.
- Verhoeven, L. T. W., & van Leeuwe, J. (2009). Modeling the growth of word-decoding skills: Evidence from Dutch. *Scientific Studies of Reading*, *13*, 205–223.
- Vygotsky, L. S. (1978). *Mind and society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press.
- Weber, R. M. (1970). A linguistic analysis of first-grade reading errors. *Reading Research Quarterly*, *5*, 427–451.
- Wentink, H. W. M. J. (1997). *From graphemes to syllables: The development of phonological decoding skills in poor and normal readers*. Nijmegen, The Netherlands: Dissertation KUN.
- Wentink, H. W. M. J., van Bon, W. H. J., & Schreuder, R. (1997). Training of poor readers' phonological decoding skills: Evidence for syllable-bound processing. *Reading and Writing: An Interdisciplinary Journal*, *9*, 163–192.
- Wimmer, H., & Goswami, U. (1994). The influence of orthographic consistency on reading development: Word recognition in English and German children. *Cognition*, *51*, 91–103.
- Wolf, M., & Katzir-Cohen, T. (2001). Reading fluency and its intervention. *Scientific Studies of Reading*, *5*, 211–239.
- Worsdell, A. S., Iwata, B. A., Dozier, C. L., Johnson, A. D., Neidert, P. L., & Thomason, J. L. (2005). Analysis of response repetition as an error-correction strategy during sight-word reading. *Journal of Applied Behavior Analysis*, *38*, 511–527.
- Yap, R., & van der Leij, A. (1993). Word processing in dyslexics: An automatic decoding deficit? *Reading and Writing: An Interdisciplinary Journal*, *5*, 261–279.