Available online at www.sciencedirect.com

ScienceDirect

Cognitive Psychology 54 (2007) 251–282

Cognitive Psychology

www.elsevier.com/locate/cogpsych

Reading ability is negatively related to Stroop interference ☆

Athanassios Protopapas [a,*], Anastasia Archonti [b,1], Christos Skaloumbakas [c,2]

[a] *Institute for Language and Speech Processing/Athena, Artemidos 6 & Epidavrou, GR-15125 Maroussi, Greece*
[b] *Panteion University, Greece*
[c] *Children's Psychiatric Hospital of Attiki, Greece*

Accepted 21 July 2006
Available online 7 September 2006

Abstract

Stroop interference is often taken as evidence for reading automaticity even though young and poor readers, who presumably lack reading automaticity, present strong interference. Here the relationship between reading skills and Stroop interference was studied in a 7th-grade sample. Greater interference was observed in children diagnosed with reading disability (dyslexia) than in unimpaired children. Moreover, poorer reading skills were found to correlate with greater Stroop interference in the general school population. In correlation and regression analyses, interference was primarily associated with reading speed, with an additional unique contribution of reading accuracy. Color naming errors were few and not comparably related to reading skills. The relation of reading skill to Stroop interference was examined in computational modeling simulations. The production model of Roelofs [Roelofs, A. (2003). Goal-referenced selection of verbal action: modeling attentional control in the Stroop task. *Psychological Review*, *110*, 88–125], in which interference is primarily due to word stimuli having direct access to word form encoding whereas color naming must pass through concept activation and lemma selection, was found to account well for the human data after imposing covariation constraints on parameters controlling word processing and blocking latency, in modifications

---

☆ We are grateful to A. Roelofs for providing code implementing his model and for his helpful comments. We also thank T. Carr for substantial editorial guidance, E. Altmann for help with the modeling literature, and R. Poldrack for comments on a previous version of this manuscript.

* Corresponding author. Fax: +30 210 6854270.
E-mail address: protopap@ilsp.gr (A. Protopapas).

[1] Present address: Department of Human Communication Science, University College London, UK.

[2] Present address: Department of Special Education, University of Birmingham, UK.

doi:10.1016/j.cogpsych.2006.07.003

not affecting the model's previous fit to other data. The connectionist model of Cohen, Dunbar, and McClelland [Cohen, J. D., Dunbar, K., & McClelland, J. L. (1990). On the control of automatic processes: a parallel distributed processing account of the Stroop effect. *Psychological Review*, *97*, 332–361], in which interference is caused by differential route strength, implementing an automaticity account, approximated the observed patterns with network-wide parameter manipulations not specific to reading, such as processing speed and response threshold, likely to affect previously optimized performance. On the basis of the empirical and modeling data we argue for a direct link between reading skill and interference, beyond the effects of executive functioning.

Keywords: Stroop effect; Reading; Automaticity; Dyslexia; Computational modeling

1. Introduction

Skillful reading is considered largely automatic at the word level; the development of automaticity is generally taken to be a major goal of reading instruction and practice (Samuels & Flor, 1997; Wolf, Miller, & Donnelly, 2000). The color-word Stroop interference task is a classic naming task in which color naming is slowed down by an interfering (incongruent) printed word (Stroop, 1935). This interference can be taken as a measure of reading automaticity (Logan, 1997). Therefore, intuitively, reading skill should be expected to be positively correlated with Stroop interference, with better readers exhibiting stronger interference. Even though expression of such a general expectation is found in the literature (Samuels, 1999), certain research findings also exist that are difficult to reconcile with it (e.g., Everatt, Warner, & Miles, 1997). In the present study we test this hypothesis that Stroop interference is positively related to reading skill and discuss the implications both for the notion of automaticity in reading and for the theoretical interpretation of the Stroop task. We present empirical evidence replicating and extending previous reports of a *negative* relation between Stroop interference and reading skill. We then conduct simulations of the observed findings in two prominent but quite different computational models of the Stroop task, one due to Roelofs (2003; 2005; Roelofs and Hagoort, 2002) and the other due to Cohen and colleagues (Cohen, Dunbar, & McClelland, 1990; Cohen, Usher, & McClelland, 1998). We use these simulations to address the puzzle of why the relation between Stroop interference and reading ability is negative rather than positive.

1.1. Automaticity in reading

Automaticity is a complex notion, generally considered to be a graded feature of task performance related to (a) speed, (b) voluntariness (in initiation, control, and termination), (c) cognitive resource requirements (attention as effort) and (d) conscious awareness (attention as focused concentration) (Logan, 1997). It can be modeled as a process feature (e.g., Cohen et al., 1990; LaBerge & Samuels, 1974) or as a transition from algorithmic processing to memory retrieval (Logan, 1988; Logan, 2002; Strayer & Kramer, 1990). Automatic processes are performed rapidly, without conscious intent or guidance, and with little effort, thus allowing the simultaneous performance of other tasks at little or no cost. Skilled reading is considered to be a largely automatic process, at least at the word level, including word decoding, up to lexical access, although text-level automaticity has also been proposed (Logan, 1997; Samuels & Flor, 1997). This means that little effort or attention is required to derive word meanings from the written letter strings. Thus, in skilled readers, sufficient cognitive resources are thought to be available for computation of the meaning of the text (Adams, 1990; LaBerge & Samuels, 1974; Samuels, 1999).

Beginning reading, on the other hand, is an extremely demanding process, inefficiently carried out, with little speed or accuracy (Kuhn & Stahl, 2003). Children attending first grade struggle with letter sounds, then with conscious decoding of letter strings into words, gradually acquiring larger orthographic units that can be processed efficiently to access the lexicon (Ehri, 1995; Samuels & Flor, 1997). The gradual process of reading automatization takes a number of school years to complete. Sometimes it fails, such as in cases of reading disability, resulting in poor reading accuracy and speed (Savage, 2004; van der Leij & van Daal, 2000; Yap & van der Leij, 1993). Nonfluent reading hampers comprehension and academic performance, therefore reading automatization is considered one of the critical objectives of remedial instruction (Kuhn & Stahl, 2003; Levy, 2001; Wolf & Katzir–Cohen, 2001; Wolf et al., 2000).

However, because of its complexity, the notion of automaticity is not entirely clear as it relates to the development of reading skill. Automaticity does entail flawless and rapid processing of text, that is, increased speed and accuracy of reading, but, consistent with Logan's (1997) multi-component analysis, it is probably not a unitary construct in its application to reading development. Stanovich (1990) has separated the components of obligatoriness, speed, and resource limitations, arguing that even though skilled reading, being an automatic process, is seen as obligatory, fast, and resource-free at the same time, reading practice past the point of accurate decoding does not necessarily achieve these goals simultaneously. In particular, obligatoriness may develop quickly while speed may still lag behind.

Complicating matters further, reading researchers often refer to reading "fluency," a term theoretically distinct from "automaticity" (Kuhn & Stahl, 2003; Wolf & Katzir–Cohen, 2001) but in practice greatly overlapping, because fluency is difficult to assess outside the domains of reading speed and accuracy. Fluency as commonly understood is certainly not simply a measure of reading speed; fast but errorful reading is not considered fluent, and neither is reading with incorrect prosody (Kuhn & Stahl, 2003). Conceding that there are no consensual definitions of what is meant by the term, Wolf and Katzir–Cohen have suggested that "reading fluency involves every subprocess and subskill involved in reading," while acknowledging that a "subset of time-related terms [are] most frequently related to [fluency] (e.g., automaticity, speed of processing, reading rate/speed, and word recognition rate/proficiency)" (p. 213). Nevertheless, quantifiable operational definitions are necessary for research designs. Thus, it is not uncommon to see fluency treated as if it were simply measured by reading speed, with the understanding that at least the automaticity aspect of fluency is properly assessed. Components of automaticity other than speed, although often acknowledged, are rarely addressed in reading research. As will be seen, this is problematic for explaining the relationships among reading ability, automaticity, and performance in the Stroop task.

1.2. Automaticity and the Stroop task

The experimental task of Stroop (1935) produces one of the best known and most robust findings in cognitive psychology. Naming the color of the ink in which a stimulus is printed takes longer when the stimulus is a word with a different color meaning (e.g., the word "green" printed in red ink) than when the stimulus is a plain rectangular patch or a string of letters (e.g., "XXXXXX" printed in red ink). MacLeod (1991) has reviewed the massive evidence from variants of this task and put a number of alternative hypothesis to the test of accounting for the data. Although none of the candidates was entirely successful, it seems that early notions of relative processing speed were inadequate, whereas the notion of automaticity was almost sufficient, especially in its modern, graded form (as opposed to the early binary distinction of controlled versus automatic processing). More recent proposals along the lines of parallel distributed processing seemed to fare best, as they apparently combine desirable aspects of both processing speed and an automaticity continuum.

Despite certain inadequacies in accounting for particular experimental findings, the notion of automaticity remains central to our understanding of the Stroop task, in particular the autonomy (obligatoriness) aspect of it (Logan, 1997). As MacLeod (1991) stated, "the basic idea is that processing of one dimension requires much more attention than does processing of the other dimension. Thus, naming the ink color draws more heavily on attentional resources than does reading the irrelevant word. Moreover, reading the word is seen as obligatory, whereas naming the ink color is not. Presumably, this imbalance derives from our extensive history of reading words as opposed to naming ink colors" (p. 188). There can hardly be a simpler account of the observed interference: reading is much more practiced than color naming and therefore reading dominates naming without regard for attention, causing interference. In support of this view, MacLeod and Dunbar (1988) tested the time course of interference as a function of practice for a new task (shape naming) and found a gradual reversal of the interference pattern, consistent both with a graded notion of automaticity and with the central role of relative automaticity in determining interference.

Restating from attentional control (in automatic processing) to strength of processing, Cohen et al. (1990) proposed that "the relative strength of two competing processes determines the pattern of interference effects observed" (p. 334). In their model, "the speed and accuracy with which a task is performed depends on the speed and accuracy with which information flows along the appropriate processing pathway" (p. 335), which depends on the connections between units along the pathway, referred to as "the *strength* of the pathway." The difference from the automaticity view is not so much in what seems to cause the effect but in how automaticity is operationally defined. Therefore, in the context of this model, and as verified in specific simulations, practice on a task increases its speed of processing and allows it to interfere with tasks of lower processing speed. The more practiced tasks are then still the ones causing the most interference. Reading and naming are seen as differing only along a practice continuum, with no fundamental asymmetries related, e.g., to word-form encoding and production.

1.3. Failures of reading automatization

Children who have not yet mastered fluent reading are, by definition, not reading automatically. They may or may not pronounce written words accurately (their difficulties depending to some extent on the characteristics of the orthographic system of their language) but they certainly do not read fast enough to be considered good readers (Samuels, 1999; van der Leij & van Daal, 2000). This is true of young children learning to read, who have mastered the alphabetic mapping but not fully the reading process itself, and also of older children who have failed to learn to read well despite years of school instruction, due to some learning disability. According to Samuels (1999), "poor readers who do not recognize words automatically find [the Stroop color-word test] to be an easy task" (p. 183), meaning that poor readers do not have difficulty pronouncing the correct ink color, despite the conflicting printed word, in contrast to good readers. Samuels thus suggested to use the Stroop task "as an indicator to determine if a student recognizes words automatically," a suggestion clearly in line with an automaticity-like concept underlying performance on the Stroop task, as discussed in the preceding section.

With respect to developing reading skills, the pattern reported in the literature is that Stroop interference in children learning to read English increases over the course of grades 1–2 (Schiller, 1966) and then diminishes slowly into adulthood, to increase again after age 60 (Comalli, Wapner, & Werner, 1962). Given that mastering reading English at the foundational level can take 2 school years or more (Seymour, Aro, & Erskine, 2003), it seems that attaining full automaticity (after grade 2) is associated with decreasing interference whereas declining reading skills in old age are associated with increasing interference, a pattern opposite from the one expected (as also noted by Roelofs, 2003; Roelofs & Hagoort, 2002).

Turning to the performance of poor readers, who are not reading automatically, robust Stroop interference has been repeatedly reported (Alwitt, 1966; Everatt et al., 1997; Helland & Asbjørnsen, 2000; Kelly, Best, & Kirk, 1989; van der Schoot, Licht, Jorsley, & Sergeant, 2000). Moreover, Everatt et al. (1997) found *more* interference in children with dyslexia than in age-matched controls. Interestingly, in the study of Everatt et al. there was no difference in interference between children with dyslexia and reading-level matched, younger controls, suggesting that Stroop interference may be directly related to reading skill regardless of age or status of reading disability.

1.4. Study plan

The present study was designed to determine whether Stroop interference is positively or negatively related to reading skill and to identify components of reading skill that may best predict performance on the Stroop task. We first compare children with dyslexia to good readers of the same age in an attempt to replicate the findings of Everatt et al. (1997). Specifically, we test whether children diagnosed with reading disability exhibit more, less, or equal interference as children who read well. Then we examine the relationship of interference with reading skills in the general school population, studying the correlations among component skills and the position of Stroop interference among them.

Our sample consists of children in grades 6–8, the great majority of them attending 7th grade. This age is sufficiently advanced for automaticity to have developed in reading Greek, at least to some intermediate stage (cf. Samuels, LaBerge, & Bremer, 1978, for English). It has been estimated that foundation reading skills in Greek are fully achieved before the end of 1st grade, possibly because of the relative transparency of the Greek orthographic system for reading (Ellis et al., 2004; Seymour et al., 2003). On the other hand, this is an age in which requests for learning disability assessment make it possible to identify well defined reading-disabled populations for controlled testing. At the same time, adult compensation strategies to counter poor reading are not expected to be effective yet, therefore this age is very good for a relatively clear distinction between good and poor readers.

2. General methods

2.1. Participants

The participant population and testing battery (except for the Stroop task) were part of a larger study on learning disability assessment (reported in detail in Protopapas & Skaloumbakas, in press). For the present studies, children were selected from the original sample based on availability of full data sets for the measures of interest (listed under "materials" below) and on additional specific criteria, different for each study, as detailed in the corresponding sections. All children were native speakers of Greek.

The full sample included children from the general school population ("school sample"), recruited in eight public schools from various regions of the Greek province of Attiki, which includes the greater metropolitan area of Athens. Children in this sample were self-selected in that their parents consented to participate in the study, responding to a written request distributed by a teacher to every child in class. The overall performance range of these students was very large, thus this sample is to some extent representative of the general school population.

A second group of children was recruited at the Children's Psychiatric Hospital of Attiki ("clinical sample"), where they requested assessment services related to learning problems at school. These children were of at least average intelligence, free from neurological disorders and primary behavioral or emotional disorders, and were diagnosed with a specific learning disability (reading disability, or "dyslexia") at the hospital by an interdisciplinary team on the basis of their consistently and substantially impaired performance in reading and spelling.

2.2. Task and materials

Reading was assessed using pseudowords (a list of 20 items), single words (a list of 84 items of varying length, phonological complexity, and written frequency), and passages (three passages 72–90 words long, varying in genre and complexity), in each case measuring the number of reading errors and the total reading time. For the passages, text comprehension was also assessed with specific questions asked after reading aloud each passage. Spelling was assessed with dictation of a 49-word passage, easy in meaning and containing well-known words, and with dictation of a list of 21 isolated words chosen to be frequent and to provide opportunities for a variety of spelling errors. More information on testing materials and procedures can be found in Protopapas and Skaloumbakas (in press).

Phoneme awareness was assessed in a phoneme deletion task with a set of 22 two-syllable and three-syllable pseudowords constructed following Greek phonotactic structure, including a high proportion of consonant clusters. For each pseudoword, one phoneme was the designated deletion target, varying greatly in phonetic features, word position and syllabic position.

In addition to the reading measures, children were tested with the digit span and arithmetic subscales of the Greek version of the WISC-III (Georgas, Paraskevopoulos, Bezevegis, & Giannitsas, 1997). The full, 60-item version of the Standard Progressive Matrices (SPM) test of nonverbal intelligence (Raven, 1976) was also administered. Since all children were of about the same age, no conversion to standard scores was undertaken for these three measures (raw scores are reported).

Finally, for the Stroop task, two sheets were prepared, one with color stimuli made up of 6 repetitions of the letter X with no spaces (control condition), and the second with color stimuli being the Greek words for red, green, blue, yellow, and brown (incongruent/color word condition). There were 60 stimuli on each sheet, arranged on 3 columns of 20. The stimuli were printed in colored ink, with 12 items on each sheet printed in each of the five named colors, randomly arranged throughout the sheet. In the word condition, each color ink was used three times for every color word except the word for its own color; thus all items were word-color incongruent.

2.3. Procedure and scoring

Each child was tested individually, by a special education professional (clinical sample) or specially trained graduate student (school sample), in a quiet room at the hospital (clinical sample) or school (school sample).

For the pseudoword reading task, the child was given a sheet of paper, on which the 20 items to be read were printed in one column, and was asked to read aloud the pseudowords "quickly, but not rushing, to avoid mistakes." Individual responses were noted when incorrect, and the duration of reading was timed using a stopwatch. Test results include the number of incorrectly read items (regardless of number of errors per item; unread or incomplete items count as one error each) and the total reading time, in seconds.

For the word reading task, the child was given a sheet of paper, on which the 84 items to be read were printed in 3 columns of 28 words each. Instructions and scoring were as for pseudoword reading.

For the text reading task, each passage was presented individually on a sheet of paper. The child was asked to read the passage aloud, during which the number of reading errors and total reading time were noted. Except for the easiest first passage, the child was then allowed an additional 1 min of silent study. The printed sheet was then taken from the child's view and 3–4 comprehension questions were asked. Points were given for correct responses (with partial points for predefined approximations). Test results include (a) the total number of reading errors from all three passages; (b) the total reading time, in seconds (added from all three passages); and (c) the total comprehension score (points added from all three passages).

For the spelling tasks, the material (passage or words) was dictated at a child-determined rate. The total number of spelling errors was noted in each task (more than one per word were possible).

For the phoneme deletion task, each pseudoword was presented orally and, once repeated correctly, was presented again along with the phoneme to be deleted. The total number of incorrect responses was noted.

For color naming, the child was first given a sheet of paper, on which five large color patches were printed, one in each of the five colors used, and was asked to name the colors, to ensure adequate color vision and verify that the intended color names were used. Then the sheet of colored XXXXXX was given, and the child was asked to name aloud the colors progressing down each column. Errors and total time were noted on the scoring sheet. Finally, the sheet of colored words was given, and the child was asked to do the same thing and not to read the words. Errors and total naming time were again noted. For both Xs and words, 30 s into each task, the experimenter noted the item reached by that time. The number of correctly named items up to that point provided the additional measure of "30 s Items."

Administration of the full testing battery, including several tasks not reported here, took 60–70 min. All tasks except for the Raven's SPM were recorded on tape and recordings were subsequently used to verify and correct the scoring sheet entries prior to analysis.

3. Study 1: good vs. poor readers

3.1. Participants

For this study, children from the school sample were included only if their performance had been evaluated by two independent experts and judged to be free from learning disabilities. No other special criteria applied. Thus the "NI" (nonimpaired) group from the school sample comprised 35 boys and 37 girls, of age 140–157 months. All children from the clinical sample were included, forming the "RD" (reading disabled) group, which consisted of 12 boys and 4 girls diagnosed with reading disability, of age 131–164 months.

3.2. Results and discussion

Table 1 summarizes the results from all measures for each group. Stroop interference is calculated here as the difference in time to name the color of all color-word items minus time to name the color of all XX items. The effect is seen most clearly in Fig. 1, where mean naming times for each condition by each group are plotted.

Table 1
Per-group mean and standard deviation for each measure and a one-way ANOVA test of the corresponding group differences

Effect measure	NI group ($N = 72$)		RD group ($N = 16$)		$F(1, 86)$	p
	Mean	*SD*	Mean	*SD*		
Age	149.3	4.1	150.5	7.6	.86	.357
Pseudoword read errors	3.4	2.8	9.0	3.6	47.44	<.0005
Pseudoword read time	43.2	12.4	64.8	15.9	35.49	<.0005
Word read errors	1.3	1.6	9.3	4.6	141.70	<.0005
Word read time	88.0	17.6	149.5	61.2	54.47	<.0005
Text read time	100.1	15.2	171.9	49.1	110.08	<.0005
Text read errors	3.7	2.6	8.9	5.2	34.09	<.0005
Text comprehension	11.3	3.1	8.7	2.4	9.91	.002
Text spell errors	2.6	2.2	15.4	7.9	143.78	<.0005
Word spell errors	.9	1.3	7.2	4.5	105.45	<.0005
Phoneme deletion errors	4.5	3.3	8.9	3.8	22.15	<.0005
Raven's SPM raw score	40.0	9.3	33.6	8.3	6.43	.013
Digit span raw score	13.7	2.7	11.5	2.2	9.38	.003
Arithmetic raw score	18.1	2.6	16.8	2.4	3.27	.074
Naming time XX	50.5	9.7	61.2	9.0	16.47	<.0005
Items 30 s XX	37.7	7.3	32.1	5.0	8.76	.004
Errors XX	.1	.5	.4	.6	3.08	.083
Naming time color words	83.7	17.6	112.1	23.7	29.89	<.0005
Items 30 s color words	23.5	5.9	17.7	4.2	14.00	<.0005
Errors color words	.6	1.2	1.3	2.1	2.72	.103
Stroop interference	33.3	14.2	50.9	17.9	18.49	<.0005

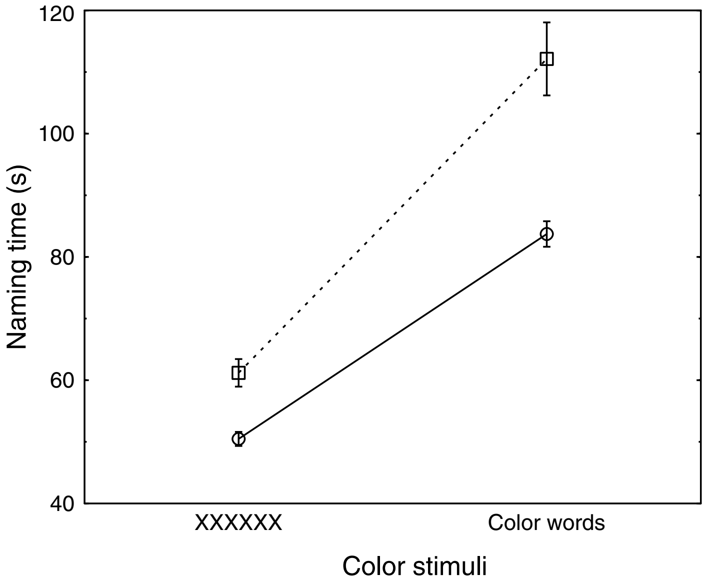

Fig. 1. Color naming time for the two types of stimuli (baseline condition XXXXXX, and interference condition, color words) by each group. Circles and solid line: NI group (school sample); Squares and dotted line: RD group (clinical sample). Error bars show standard error.

These raw times were entered into a 2 (stimulus conditions) × 2 (groups) repeated-measures MANCOVA, with Raven's SPM raw score as a covariate (because it was significantly different in one-way ANOVA between the two groups, see Table 1). The effect of the SPM covariate was not significant ($F(1,85) = 2.40$, $p = .125$), indicating that any nonverbal intelligence differences between the two groups were not related to their differences in Stroop interference. The 18 s main effect of group was significant ($F(1,85) = 24.18$, $p < .0005$), as was the 42 s main effect of stimulus condition ($F(1,85) = 30.19$, $p < .0005$). The 18 s interaction was also significant ($F(1,85) = 18.69$, $p < .0005$), reflecting the larger difference among stimulus conditions for the RD group than for the NI group. Therefore, Stroop interference is significantly *larger* in the RD group than in the NI group, in agreement with the findings of Everatt et al. (1997).

Color naming errors were then subjected to the same 2 × 2 MANCOVA, in which the effect of the SPM covariate was again nonsignificant ($F(1,85) < 1$). The main effect of group was significant ($F(1,85) = 4.38$, $p = .039$) but the main effect of stimulus condition was not ($F(1,85) < 1$), and there was no significant interaction between the two ($F(1,85) = 1.41$, $p = .238$). Therefore the RD group appears to make more errors overall, but not consistently more so in the color-word stimulus condition. The apparent difference between the stimulus conditions seen in Table 1 is due to some of the children making several errors and does not reflect a consistent tendency; consider that the median is zero errors in all four combinations of group and stimulus condition.

In χ^2 analysis of the proportions of children in each group making each number of errors (Table 2), a marginally significant difference emerged for the baseline (XXX) condition ($\chi^2 = 6.272$, $df = 2$, exact $p = .051$, two-sided), reflecting the fact that 67% of the RD children made zero errors, contrasting with 90% of the NI children. However, there was no corresponding significant difference for the color-word condition ($\chi^2 = 9.517$, $df = 6$, exact $p = .181$, two-sided), consistent with the interpretation that stimulus incongruence does not affect the error rates of RD children more than those of NI children.

Table 2
Number of children in each participant group (RD and NI) making the indicated number of color naming errors in the baseline (XXX) and incongruent (color word) condition in Study 1

Number of errors	XXX		Color word	
	NI	RD	NI	RD
0	65	11	48	9
1	4	4	16	2
2	3	1	4	2
3	0	0	2	2
⩾4	0	0	2	1
Total	72	16	72	16

Table 3
Per-group mean and standard deviation for each alternative Stroop interference measure and an ANOVA test of the corresponding group difference

Effect measure	NI group ($N = 72$)		RD group ($N = 16$)		$F(1, 86)$	p
	Mean	*SD*	Mean	*SD*		
Time difference	33.3	14.2	50.9	17.9	18.49	<.0005
Time ratio	1.7	.3	1.8	.3	4.02	.048
Items 30 s difference	14.3	6.8	14.4	5.8	.00	.946
Items 30 s ratio	1.7	.5	1.9	.6	3.03	.085

The interaction test for naming times is similar to testing the time difference directly (in one-way ANOVA, as shown in Table 1) because in each case what is tested for significance is the between-group difference among the between-condition time differences. This corresponds to the standard calculation of the Stroop effect. However, there are alternative ways to express the underlying effect, that is, the increased difficulty in color naming when the items are color words versus when the items are meaningless letter strings. An obvious alternative would be the time ratio. If one group is slower in general (and slower to name colors in particular), then a given percentage increase in time would correspond to an equal *relative* increase in difficulty even though the absolute time difference would be larger. Other alternatives can be formed by taking into account the number of items named correctly within a given time period, such as in 30 s, instead of the time taken to name all the items. The effect could then be calculated as either the difference or the ratio of the number of items in each condition (XX versus color words).

Table 3 presents these four ways to express the interference effect. All four indices of interference are significantly different from zero, in separate analyses, for both the NI group ($t(71) \geqslant 17.7, p < .0005$) and the RD group ($t(15) \geqslant 9.9, p < .0005$). Therefore there is no question that the RD group, like the NI group, presents robust Stroop interference, regardless of how it is measured.

In each of the four ways to compute interference, the magnitude of the interference effect is larger for the RD group than for the NI group, but only in the usual calculation of the effect is this difference clearly statistically significant.[3] On the one hand, there is clearly no statistically significant effect of group if the interference is expressed in terms of number

---

[3] Because the distributions of these interference measures are not normal, the statistical comparisons were also performed on transformed variables, with identical results in terms of statistical significance.

of items correctly named for color. On the other hand, both time-derived measures of interference are statistically significantly different between the two groups. However, the difference of the time ratio measures appears substantially less strong than the one based on the time differences. This might reflect lack of power, due to small sample size, or it might indicate a spurious, unreliable effect. Therefore it is necessary to examine further the relation between Stroop interference and reading ability before any firm conclusions can be drawn.

4. Study 2: general population

Using one end of the reading-skill continuum (the RD group) as a group contrasting with the center of the same distribution (the NI group) has the disadvantage that within-group differences work against the effect in question. This would be fine if the two groups were qualitatively different on color naming such that within-group interference differences were randomly distributed with respect to reading skill. If, however, reading skill is continuously related to Stroop interference, then within-group differences could be used to enhance our confidence in the significance of the relationship. Therefore, in this section, correlation and regression analyses are carried out on the measures from the school population only.

4.1. Participants

All 156 children (81 girls) from the school sample participated in this study, regardless of whether they had been evaluated for reading difficulties or not. Their ages ranged between 136 and 172 months ($M=151$, $SD=6$). Data from the clinical sample were not used.

4.2. Results and discussion

Multivariate data analysis procedures are not as robust to deviations from normality as ANOVA is. Therefore, after examining the detailed descriptive statistics, histograms, and Q–Q plots of all the variables, transformations were undertaken to approximate normal distributions for them. Generally, time measures required an inverse transformation whereas error measures required a square root transformation. The distributions of text comprehension, Raven's SPM, digit span, and arithmetic raw scores were judged to be sufficiently close to normal already and therefore these variables were not transformed. Three of the four measures of interference were also transformed accordingly. The color naming error measures (in both XX and color word stimulus conditions) were not transformed, despite high skewness, because they were composed of a majority of zero values and so no transformation could possibly bring their distributions to approximate normality.

Because not only the strength but also the precise nature of the relationship between Stroop interference and reading skills is at question here, the concern arises that nonlinear transformations may introduce unwanted distortions. Therefore, subsequent analyses are carried out with the transformed or untransformed variables, as appropriate, or both, when possible, to ensure that interpretations are not confused by the statistical procedures.

Turning first to the correlations among the measures, in this unselected school sample, which is presumably representative of the general population, Stroop interference was found to be moderately correlated to several reading measures. Table 4 shows the correlation coefficients among the measures (standard Pearson's product-moment coefficients among the transformed variables and non-parametric Spearman's ρ among the untransformed variables). The relationship between interference and reading skills is a *negative* one: higher reading skills are associated with less Stroop interference.[4]

Neither age nor any of the intelligence-related measures are related to interference, thus a potential source of uncontrolled complexity is removed. Although reading speed and accuracy are positively related to interference in all four ways it can be computed, clearly the standard measure of interference (time difference) is most strongly related to reading whereas the item difference is least strongly related.

Moderate correlations are obtained between color naming time (in the control XX condition) and all reading speed and accuracy measures. This is expected because naming speed is strongly related to reading ability, particularly to reading speed (Leinonen et al., 2001; van den Bos, Zijlstra, & van den Broeck, 2003; Wolf & Bowers, 1999). In fact this relation is sufficiently important that recent developments in reading disability indicate that a speed measure, as assessed by naming tasks, should be considered alongside phonological skills as a major contributing factor to reading skill development (or failure; hence the "double-deficit" theory of reading disability. See Wolf & Bowers, 1999; Wolf et al., 2002). Because of this relationship, it is possible that baseline naming speed is the major determinant of interference, by affecting color naming speed for color words. If this were the case, then reading variables should not contribute to the prediction of interference once the effects of baseline color naming are removed by statistical regression.

To test this hypothesis, linear multiple regression analyses were carried out, with each of the four interference indices in turn being the dependent variable (DV). Baseline (XX) color naming time was always forced first into the equation.[5] The six reading accuracy and time measures and the two spelling accuracy measures were entered as additional independent variables (IV), without specifying manually which should enter the regression equation (if any) or in what order. In every case,[6] pseudoword reading time entered the equation first, passing an F-probability criterion of .05, and contributing 8–15% of additional interference variance.

Table 5 shows the corresponding statistics for the 2nd step of each regression (with baseline color naming and pseudoword reading time), including additional unique variance contributed by each variable (R^2 change) and corresponding standardized coefficient. Not shown in the table, the same results are obtained if, for the interference measures derived from the number of color-named items, the baseline number of items is entered first into the equation instead of the baseline color naming time.

---

[4] Note that, because of the inverse transformation, correlations with transformed time measures should be interpreted in the opposite direction than indicated by the sign of the coefficient.

[5] Color naming errors were also used originally as a control variable in the first step (analyses not reported) but were not found to contribute any significant variance, as might be expected by their low correlations with all relevant measures, and were subsequently dropped from further analyses.

[6] Except for untransformed Items 30 s ratio, in which case pseudoword reading time entered third, after text spell errors. The values in Table 5 correspond to a run for this DV in which PRT was forced second. The discrepancy may be caused by the extreme deviation from normality exhibited only for this measure of interference, which required an inverse transformation instead of a square root.

Table 4

Correlations between Stroop interference indices (lines and rows 1–4), color naming measures (lines and rows 5–10), and all the
sures (lines and rows 11–23), for the unselected school sample

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1	Time difference		**.89**	**.49**	**−.71**	−.27	**−.86**	−.17	**−.79**	.06	.13	**.32**	**−.48**	**.41**	**−.39**	**−.47**	**.37**	−.15	**.37**
2	Time ratio	**.89**		**.76**	**−.82**	.18	**−.56**	.26	**−.52**	.02	.08	.27	−.27	**.34**	−.18	**−.31**	.22	−.10	.25
3	Items 30 s difference	**.52**	**.78**		**−.90**	**.54**	−.12	**.69**	−.23	−.06	.09	.18	−.02	.18	.04	−.08	.07	−.03	.13
4	Items 30 s ratio	**.76**	**.85**	**.89**		−.19	**.44**	**−.35**	**.56**	.01	−.13	−.25	.24	**−.29**	.16	.28	−.16	.10	−.24
5	Naming time XX	.25	−.17	**−.53**	−.14		**.70**	**.93**	**.64**	−.07	−.09	−.14	**.47**	−.18	**.47**	**.39**	**−.29**	.13	−.26
6	Naming time color words	**.85**	**.56**	.13	**.49**	**.69**		**.58**	**.92**	−.08	−.15	**−.32**	**.60**	**−.39**	**.52**	**.55**	**−.41**	.18	**−.39**
7	Items 30 s XX	−.13	.26	**.67**	**.30**	**−.93**	**−.56**		**.55**	−.15	−.10	−.08	**.40**	−.16	**.41**	**.32**	−.22	.10	−.18
8	Items 30 s color words	**−.80**	**−.53**	−.23	**−.61**	**−.62**	**−.92**	**.53**		−.14	−.24	**−.31**	**.56**	**−.42**	**.51**	**.52**	**−.38**	.17	**−.39**
9	Naming errors XX	.09	.03	−.04	−.01	.06	.09	−.11	−.13		**.40**	.20	−.03	.24	−.05	−.11	.17	−.13	.17
10	Naming errors color words	.18	.08	.03	.11	.13	.20	−.15	−.25	**.42**		**.28**	−.16	.22	−.17	−.20	**.30**	−.09	.22
11	Pseudoword read errors	**.33**	**.28**	.22	**.28**	.11	**.32**	−.04	**−.32**	.17	**.28**		**−.40**	**.61**	**−.44**	**−.39**	**.53**	−.15	**.55**
12	Pseudoword read time	**.54**	**.31**	.04	**.30**	**.50**	**.66**	**−.40**	**−.62**	.10	.16	**.38**		**−.43**	**.72**	**.65**	**−.55**	.13	**−.47**
13	Word read errors	**.42**	**.36**	.22	**.34**	.12	**.40**	−.09	**−.40**	.22	**.29**	**.58**	**.44**		**−.59**	**−.46**	**.58**	−.24	**.62**
14	Word read time	**.41**	.20	−.02	.21	**.46**	**.55**	**−.40**	**−.52**	.11	.19	**.38**	**.71**	**.54**		**.73**	**−.61**	**.28**	**−.62**
15	Text read time	**.48**	**.30**	.09	**.30**	**.39**	**.57**	**−.31**	**−.53**	.15	.19	**.39**	**.66**	**.51**	**.77**		**−.58**	**.40**	**−.70**
16	Text read errors	**.35**	.25	.14	.26	.22	**.39**	−.14	**−.37**	.22	**.34**	**.51**	**.50**	**.56**	**.60**	**.54**		−.22	**.63**
17	Text comprehension	−.12	−.07	−.04	−.09	−.11	−.16	.09	.13	−.12	−.06	−.14	−.14	−.27	**−.29**	**−.41**	−.21		**−.32**
18	Text spell errors	**.33**	.22	.13	.24	.20	**.36**	−.13	**−.35**	.15	.26	**.51**	**.43**	**.61**	**.58**	**.63**	**.56**	**−.31**	
19	Word spell errors	**.30**	.21	.10	.19	.15	**.32**	−.08	−.27	.11	.15	**.44**	**.31**	**.57**	**.48**	**.55**	**.55**	**−.28**	**.72**
20	Phoneme deletion errors	.25	.20	.16	.22	.12	.27	−.07	**−.29**	.15	.17	**.50**	**.30**	**.46**	**.42**	**.43**	**.45**	**−.35**	**.48**
21	Raven’s SPM raw score	−.03	.04	.02	−.03	−.13	−.09	.12	.13	−.20	−.13	−.24	−.06	−.23	−.11	−.14	−.19	**.40**	**−.29**
22	Digit span raw score	−.27	−.16	.03	−.10	−.25	**−.34**	.23	**.30**	−.22	−.20	**−.37**	**−.35**	**−.29**	**−.43**	**−.44**	**−.30**	**.31**	**−.31**
23	Arithmetic raw score	−.11	−.02	.08	.01	−.20	−.20	.20	.18	−.12	−.10	**−.36**	−.23	**−.29**	**−.36**	**−.40**	**−.37**	**.39**	**−.43**
24	Age	−.04	−.02	.03	−.01	−.09	−.07	.06	.03	.12	.00	.11	−.05	.13	−.02	−.08	.00	.02	−.04

Lower left of the diagonal: Spearman’s non-parametric ρ coefficients among raw (untransformed) measures; Upper ri
product-moment correlation r coefficients among measures appropriately transformed to approach normal distributions (ex
Correlations significant to $p < .0005$ are shown in boldface type.

Table 5

Results of regression analyses with the listed interference measures as dependent variables, predicted by baseline (XX) color naming time (XCN; forced into the equation first) and by pseudoword reading time (PRT; entering the equation second)

Interference measure	Transformed variables				Untransformed variables			
	R^2 change		Standardized β		R^2 change		Standardized β	
	XCN	PRT	XCN	PRT	XCN	PRT	XCN	PRT
Time difference	.073	.153	.061	.444	.079	.110	.121	.368
Time ratio	.032	.155	−.388	.446	.033	.112	−.343	.371
Items 30 s difference	.290	.097	−.704	.352	.293	.079	−.677	.313
Items 30 s ratio	.035	.137	−.384	.419	.024	.095	−.305	.342

The sign of β in the transformed variables analyses was adjusted manually in this table to maintain the same direction as for the untransformed variables.

For the pseudoword reading time, both the variance proportion and the standardized regression coefficient come out remarkably stable, taking into account that the nature of the four interference measures as well as their distributions differ substantially. This uniformity across measures of interference and across variable transformation suggests that the effect is robust and not due to nonlinearities either introduced or distorted by the transformation procedures. Importantly, reading speed is strongly correlated even with the items measures once the effects of baseline naming speed are partialled out.

To alleviate any concerns that the obtained relation may be caused by outliers or certain segments of the distribution only, Fig. 2 plots interference (standard time difference measure) against pseudoword reading time (normalized values for both variables). Despite the presence of a few multivariate outliers, it is clear in this scatterplot that the relation holds across the entire range of decoding speed.

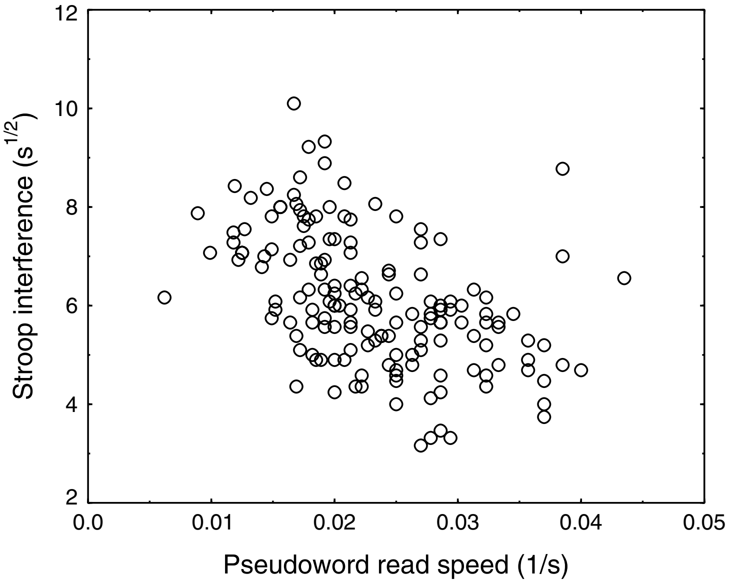

Fig. 2. Transformed Stroop interference (square root of the time difference between color naming conditions) plotted against transformed pseudoword reading speed (inverse time), showing the strong relationship between the two variables across their entire range.

Returning to the regression analyses, an error measure always enters the equation as 3rd predictor IV, after pseudoword read time, accounting for unique additional variance. This is the word read errors variable for all of the analyses with transformed variables, replaced by the pseudoword read errors or the text spell errors in some of the analyses with untransformed variables. Therefore, a consistent pattern of results emerges, which indicates that color-word naming interference is continuously negatively related to reading skills (better reading, less interference). The strongest relation is with a straightforward decoding speed measure, namely pseudoword read time, and this holds after partialling out the effects of baseline color naming speed. Significant additional interference variance is accounted for by one or more reading accuracy measures.

Reading accuracy is primarily associated with the contribution of phonological processing to reading ability (Leinonen et al., 2001). However, reading fluency comprises both a speed and an accuracy aspect, with the speed measure more directly relevant. These relationships have been verified for the Greek language (Protopapas & Skaloumbakas, in press). In particular, a three-factor structure has been found to underlie a comprehensive testing battery designed to assess learning disability, which includes the reading measures reported here. The three factors appear to reflect reading accuracy, reading fluency (primarily reading speed but also including some variance from reading errors), and intelligence.

Because both errors and time are found here to be correlated with Stroop interference, and because time seems more strongly correlated with interference than errors, one may hypothesize that Stroop interference is associated with the fluency dimension, in which both time and error measures contribute. To test this hypothesis, we can examine the factor structure of the set of skills measured, including the baseline naming and interference variables.

The transformed variables were thus entered into a maximum likelihood factor analysis with varimax rotation, in which three factors with eigenvalues greater than 1.0 were extracted, together accounting for 53.3% of the total variance. The structure of the resulting rotated matrix is shown in Table 6.[7] Both baseline color naming time and standard Stroop interference load on Factor 2, which apparently lies along a reading speed dimension. However, interference also has a smaller but non-negligible loading on Factor 1 and, moreover, there is not a very clean separation of accuracy from speed in this factor structure.

Alternatively, the relation of interference to the speed and accuracy dimension can be tested by computing the correlation coefficients of interference with the regressed factor scores of the full assessment scale analysis of Protopapas and Skaloumbakas (in press, derived from 213 children with more measures), where a clearer factor structure was found. The (square-root transformed) standard index of interference (time difference) was the one with the strongest correlations. The coefficients (Pearson's r) were .37 ($p < .0005$) with Factor 1 termed "accuracy," −.44 ($p < .0005$) with Factor 2 termed "fluency," and .12 ($p > .1$) with Factor 3 termed "intelligence." As with the preceding factor analysis, these results are not consistent with the hypothesis that correlations of Stroop interference with error and

---

[7] It was not possible to enter all color naming measures and derived interference indices into the same analysis because they are statistically redundant (all based on the same set of original measures). Plain color naming (of the non-word XX stimuli) and standard "Stroop interference" (computed as time difference) were judged to be the most relevant measures and are shown here.

Table 6
Rotated factor matrix ($N = 156$)

Measure	Factor		
	1	2	3
Word spell errors	**.83**	−.20	−.29
Text spell errors	**.79**	−.33	−.31
Text read errors	**.52**	**−.47**	−.27
Word read errors	**.51**	−.38	−.29
Pseudoword read errors	**.41**	−.31	−.36
Pseudoword read time	−.22	**.85**	
Word read time	−.39	**.75**	.22
Text read time	**−.51**	**.62**	.26
Color naming XX time		**.51**	.12
Stroop interference	.25	**−.47**	
Arithmetic raw score	−.20	.13	**.65**
Text comprehension	−.15	.11	**.54**
Raven's SPM raw score	−.20		**.59**
Digit span raw score		.38	**.53**
Phoneme deletion errors	.32	−.24	**−.57**

Loadings of .4 or higher are shown in bold; loadings less than .1 are not shown at all.

time measures reflect a single, simpler underlying dimension of fluency only (i.e., of fast accurate reading). In contrast, it seems that interference receives independent contributions from both accuracy and speed.

5. General discussion

The purpose of our studies was to examine the relationship between color-word naming interference (the Stroop effect) and reading skills. Our results show that interference is significantly related to reading skills, over and above baseline naming speed, and in particular most strongly to decoding speed. The relationship is negative, that is, better reading skills are associated with less interference. The relationship holds continuously in the general school population and also between good readers and reading disabled children, the latter exhibiting stronger interference than the former. Both reading accuracy and speed apparently contribute independently to the interference variance, therefore interference seems to be determined by overall reading expertise and not solely by some component skill.

5.1. Interference in development and impairment

Our results are consistent with previous reports in the literature that reading-impaired children show robust interference (Alwitt, 1966; Everatt et al., 1997; Helland & Asbjørnsen, 2000; Kelly et al., 1989; van der Schoot et al., 2000). In particular, our results are consistent with reports that dyslexic children show *more* interference than nondyslexic children (Everatt et al., 1997; Kelly et al., 1989). In agreement with the negative relation between skills and interference found here, Everatt et al. found interference for dyslexic children to be the same as that for reading-level matched younger controls. On the other hand, both Everatt et al. and Kelly et al. (1999) failed to find very strong relations between Stroop interference and reading ability. This is also consistent with our findings insofar as both Everatt et al. (1999) and Kelly et al. used primarily reading comprehension measures as indices of reading ability and not measures of reading speed and accuracy. As our data indicate, at this age reading comprehension is only slightly correlated with single word and pseudoword reading speed and accuracy, and not at all with color naming interference.

Our results are also consistent with the developmental findings (Comalli et al., 1962) that interference decreases with age as children grow up from 7 to 20 years (as well as with the findings of Su, 1997, for picture-word naming interference from 2nd to 6th grade). Because reading skills develop throughout childhood, application of a longitudinal interpretation to the cross-sectional data indicates that the decreasing interference is negatively related to the increasing reading ability across time. Therefore, because interference is obviously not possible in the absence of any reading ability, the increasing interference through school grades 1 and 2 (Schiller, 1966) most likely reflects the onset of reading, that is, *emerging* reading skills, and not the gradual attainment of adult fluency. The increased interference reported in old age (over 60 years) may correspond to an overall decrease in reading efficiency, though this hypothesis warrants further testing.

The developmental pattern of increasing interference with initial skill emergence and decreasing interference with subsequent automatization (speedup) has also been derived experimentally, with 1st grade children. Ehri and Wilce (1979) found that training recognition accuracy for unfamiliar words increases interference, whereas training recognition speed for familiar words decreases interference. In an examination of the early stages of word recognition, Stanovich, Cunningham, and West (1981) likewise found increasing interference during the first 6 months of first grade, indicating emerging reading skills, but little further increase in the next 2 months. Presumably, had they continued testing through 3rd or 4th grade they would also have found decreasing interference associated with further automatization of reading skills.

5.2. Interference and reading automaticity

These findings are counterintuitive with respect to our understanding of the Stroop task and its relation to reading automaticity, because it is generally assumed that, as skill efficiency increases with practice, the potential of this skill to interfere with other, less practiced skills, also increases. This assumption is consistent with experimental results from training studies (MacLeod & Dunbar, 1988). Therefore it seems necessary to reconsider the course of automatization of reading skills and the effect of individual aspects of automaticity on naming interference.

On the one hand, word recognition speed continues to increase long after the obligatory aspect of word recognition has developed (Samuels & Flor, 1997; Stanovich, 1990). On the other hand, Stroop interference is taken as an index of the "autonomy associated with automatic processing" (Logan, 1997, p. 126) and not of processing speed (cf. MacLeod, 1991). Therefore, Stroop interference should quickly increase to its maximum level with the attainment of obligatoriness during early reading acquisition, and then remain stable. If further skill improvement were to have an effect, in accordance with the training study of MacLeod and Dunbar (1988), this effect would have to be an *increase* of interference. The finding that interference is *negatively* related to reading skill runs counter to this line of reasoning.

Recent progress in our understanding of the role of attention in "obligatory" processing, including Stroop interference, suggests that processing of the distracting stimuli requires some attentional allocation, which can be manipulated by spatial context and task demands (Lachter, Forster, & Ruthruff, 2004; Risko, Stolz, & Besner, 2005). However, in the particular form of the Stroop task that was used in our studies, with fully co-extensive target and distracting stimuli of unlimited duration, the classical view of obligatoriness in visual and semantic processing still applies. Therefore, if interference reflects obligatoriness, then effects on interference should be mediated by a modulation of obligatoriness. For our particular finding of reading skill effects, it is difficult to imagine that the obligatory aspect of word recognition diminishes with greater reading skill. Certainly such a proposal would require substantial empirical evidence to be supported. It seems, then, preferable to explore the possibility that Stroop interference does not simply reflect the autonomy of automatic processing but is also related to the speed of processing, in agreement with conceptualizations of automaticity that dissociate its constituent components.

In the context of the instance theory of automaticity (Logan, 1988), Stroop interference hinges on obligatory retrieval, that is, the assumption "that attention to an object or event is sufficient to cause things that were associated with it in the past to be retrieved from memory" (Logan, 1997, p. 131). A possible mechanism, in this context, that might account for the observed relation between interference and reading ability, derives from the assumption that retrieval involves a race between traces. Recent developments of the instance theory have used a "counter" or a "random-walk" model to allow conflicting situations to modulate response time while retaining response accuracy (Logan, 2002).

Alternatively, if a mechanism can be postulated to allow suppression of race winners based on task demands, then a possible course of processing might be as follows: as soon as obligatory retrieval is active, word reading dominates the naming task. Correct performance (color naming) is possible only to the extent that reading responses (which typically win the race) can be suppressed at the output stage. Because the response can be produced only after inappropriate candidates have been blocked, the time at which the appropriate (color naming) response can be selected cannot be earlier than the time at which the dominating (read word) response is suppressed. Therefore, slower word reading will lead to later suppression of the read response and later selection of the naming response; and thus to slower color naming in the incongruent condition, that is, greater interference. This account is similar to the one proposed by Miozzo and Caramazza (2003), where active blocking of distractors was hypothesized to explain the finding that distractor word interference in picture naming is *inversely* correlated to the frequency of the distractor.

5.3. The hypothesis of reduced inhibition

One possible source of increased Stroop interference may be poor cognitive control resulting in reduced inhibition. Consistent with findings in aging populations (Spieler, Balota, & Faust, 1996), inhibition-based explanations have been proposed by researchers who found marked interference with poor readers. For example, Everatt et al. (1997) noted that dyslexics, like non-dyslexics, are unable to stop word processing prior to the point of interference. They suggested that some level of control of this interference is possible, but "this control process…is detrimentally affected in young dyslexics" (p. 228). Helland and Asbjørnsen (2000) attributed Stroop interference in their dyslexic group to impaired executive function (because comparable deficits were found on other tasks such as dichotic listening and card sorting), in particular in "the ability to select relevant stimuli and to complete tasks involving those stimuli in an efficient manner" (p. 45). Kelly et al. (1989) found that reading disabled boys (with "severe decoding difficulties") "performed like adult prefrontal patients on the Stroop interference task [in that] they were unable to maintain selective attention to the color of the ink and to inhibit responses to the color words themselves" (p. 288).

In another study, van der Schoot et al. (2000) divided their dyslexic participants into (fast but inaccurate) "guessers" and (accurate but slow) "spellers." Guessers presented with more interference and also with an impulsive pattern of executive deficits, particularly inhibitory deficits, typically associated with ADHD. In contrast, spellers were found to be "capable of inhibiting inappropriate responding" (p. 309), leading to the suggestion that the distinction into guessers and spellers may overlap with "some inhibitory dimension" such that spellers may have "an overactive inhibition system" (p. 310). Although the grouping and findings of van der Schoot et al. do not entirely align with ours, the suggestion of inhibitory deficits remains prominent in the account of non-reading deficits of reading impaired children.

Our studies do not allow a final conclusion to be drawn regarding the potential role of executive control in general, and inhibition in particular, because no independent measures were taken to assess these cognitive domains. Further investigation will be necessary to identify the role of such attentional and executive factors in reading and interference. However, the inhibition hypothesis is not likely to offer a complete explanation of our findings for two reasons: first, in the comparison between nonimpaired and dyslexic children there were no significant group differences in color naming errors accompanying the highly significant differences in color naming time. In fact, in the critical incongruent condition the distribution of participants over number of errors was statistically indistinguishable for the two groups (Table 2). In aging and demented populations, where increased interference is attributed to deficient inhibition, naming errors increase substantially along with increasing interference (Spieler et al., 1996). The same pattern was found in the language-impaired dyslexic participants of Helland and Asbjørnsen (2000) and the dyslexic "guessers" of van der Schoot et al. (2000). If poor inhibition was the main source of the increased interference observed in our Study 1, it remains to be explained why the impaired inhibition failed to result in a higher proportion of naming errors as well.

Second, our strongest findings do not concern reading disability and the small special group of impaired readers in Study 1. Rather, we have found a continuous negative relation of reading ability, particularly speed, with color naming interference, which is not accounted for by other variables and is not at all correlated with color naming errors (note especially the nonparametric coefficients in Table 4). If the inhibition hypothesis is the correct explanation of increased interference, despite the lack of an effect on errors, then on the basis of these findings it would have to be expanded to encompass the general population. That is, poor executive inhibition should be found to be continuously associated with increased interference as well as with development of poor reading skills in the general, otherwise cognitively unimpaired, population, resulting in slow, inefficient reading. Whether a priori considered likely or not, and despite being at present less parsimonious than an account of interference that requires consideration of reading skill only, this is an empirically testable hypothesis, and one worth investigating further, in order to elucidate the role of attentional and executive factors of cognitive control in both reading performance and Stroop interference. Further insight into the need to consider both reading skill and executive control can be obtained by applying current computational models of the Stroop effect to the present results.

6. Computational modeling

6.1. Production systems

How could an account of the effects of slower reading be captured by existing models of the Stroop effect? In the production systems approach, Altmann and Davidson (2001) and Lovett (2001) have put forth models based on the ACT-R theory (Anderson & Lebiere, 1998). In the model of Altmann and Davidson (2001), a word stimulus automatically activates the corresponding lemma, so it is not obvious how the speed of word recognition might be dissociated from its obligatoriness. In the model of Lovett (2001), base-level activation combines with strategic preferences to produce different amounts of interference. Word reading is favored over color naming because of the greater utility value associated with the reading production, thus expressing the obligatory character of reading. The one-production bottleneck dictated by ACT-R theory may offer a mechanism to allow poor reading to produce greater interference than good reading if the speed of the reading production firing can be modulated by a "reading ability" factor.

More recently, Roelofs (2003) has developed a model for the Stroop task based on a more general word production model (WEAVER++; Levelt, Roelofs, & Meyer, 1999). The basic assumption underlying this model is that word reading can directly activate both lemma retrieval and word-form encoding, whereas color naming must pass through conceptual identification before activating lemma retrieval and subsequently word-form encoding. Word input ("reading") and color input (via conceptual activation) both contribute activation at the lemma level at which they may compete for selection. Lemma selection is determined by reference to the current task-dependent goal (e.g., "color naming") at a verification stage, so a lemma cannot be selected for response unless it matches the current goal regardless of how much it has been activated. The main difference between reading and color naming is the following: because word reading is inherently granted direct access to word forms, it can prevent activation of the word-form response that corresponds to the perceived color. In contrast, color naming cannot prevent activation of the lemma that corresponds to the written word but must wait until the incorrect response is suppressed by blocking rules that prevent goal-inappropriate responses and turn off goal-inappropriate stimulus input. Therefore, in this model, color-word Stroop interference is not so much a matter of experience as it is determined by a fundamental asymmetry between reading and concept naming. Crucially, because of this asymmetry, there is no reverse interference in the model (from an incongruent ink color to the word reading task).

This model has been applied to the "key findings" on the Stroop task listed by MacLeod (1991) and to the results from additional experiments with great success. In particular, the model accounts for the inverse relationship between interference and age via manipulation of the duration of word input (parameter *du*). This parameter "determines the gain of the distractor input relative to the target input and is thus a central attentional parameter in the model" (Roelofs, 2003, p. 101). The way the *du* parameter actually functions in the model is by setting the latency of a blocking rule (P3 in Roelofs, 2003). That is, lemmata are activated by their corresponding stimuli for an amount of time equal to *du*, at which point P3 applies and blocks the contribution of the task-inappropriate stimulus to lemma activation, thus allowing the activation of the task-appropriate (color naming) lemma to exceed the activation of the task-inappropriate (reading) one, eventually enabling correct lemma selection. Executive input control by P3 is delayed presumably because the model cannot verify an active lemma against the flagged goal concept and determine that it is inappropriate before it has accumulated sufficient activation. Factors slowing down the accumulation of activation, such as low frequency or inefficient processing skills, can potentially delay blocking. If, therefore, *du* is allowed to be frequency sensitive, then findings such as those of Miozzo and Caramazza (2003) can be accounted for (Roelofs, 2005).

Thus parameter *du* has both the conceptual potential as well as the practical flexibility to allow the model of Roelofs (2003) to account for our counterintuitive empirical relationship between reading skill and interference. In the following section we present a simulation using this model, exploring the effects of variability in reading skill introduced by linking *du* to word-form encoding efficiency.

6.2. Simulation 1

This simulation was carried out using code (kindly provided by A. Roelofs) implementing the lemma retrieval stage of the Roelofs (2003) model, because this is the only stage at which interference takes place. Average latencies for each stage of the model, as given by Roelofs, are shown in Table 7. Because word stimuli activate word-form encoding directly, no time is spent on lemma retrieval for the reading task. In contrast, color naming must go through lemma retrieval on the basis of conceptual activation, which is why color naming is slower than word reading. When an incongruent word stimulus is present during the color naming task, then retrieval of the appropriate lemma takes even longer because the word activates an incorrect lemma for the duration of *du*.

To implement the effect of word reading speed on interference we assume that the amount of time the incongruent word stimulus is allowed to activate its lemma (i.e., *du*) is a function of the amount of time needed for the word stimulus to complete word form encoding if unimpeded. That is, we assume that word-form encoding latency comprises a variable, skill-dependent component and a fixed component; and that the output of the skill-dependent component controls blocking by furnishing the information that a word activation is available, at which point the appropriateness of this available information can be verified. Put simply, blocking of word input depends on the latency of word form encoding (WFE), which is manipulated to simulate the slow reader, so that *du* is a function of WFE.

Table 7
Simulated latencies (ms) per stage for the Stroop model of Roelofs (2003), in total corresponding well to the response times observed by Glaser and Glaser (1982)

Stage	Word reading	Color naming
Visual processing	100	100
Lemma retrieval	—	105
Word-form encoding	147	147
Phonetic processing	200	200
Model total	447	552
Glaser and Glaser (1982)	452	540

Tabulated based on Roelofs (2003, p. 108).

In the context of this particular model, affecting WFE latency only and not other components of the process allows other visual tasks, as well as phonetic planning and articulation, to remain unaffected. Thus the modifications necessary for modeling our data do not have unrealistic implications for unrelated tasks with which reading is not correlated and in which poor readers typically have no deficits. The particular choice of the WFE stage is also in agreement with the current conceptualization of dyslexia as a difficulty primarily rooted in phonological representations supporting word identification and with neuroimaging findings supporting the hypothesis that impaired reading concerns inefficient mapping from orthography to phonology (via specific letter-string processing) and suboptimal phonological processing (Perfetti & Bolger, 2004; Salmelin & Helenius, 2004).

The actual parameter values were determined so that the original behavior of the model would not be affected by our extension. Therefore, based on the default values of 147 ms for word-form encoding (Table 7) and 100 ms for *du* (Roelofs, 2003, p. 124), we set $du = \text{WFE} - 47$ ms. Word reading was modeled simply by adjusting WFE and adding the durations of the three relevant stages shown in Table 7. The control condition of color naming (i.e., without a word stimulus) was modeled by setting *du* at the selected value and word input fixed to zero, letting the lemma retrieval stage run to selection, then adding the resulting retrieval time to the remaining three stages with a corresponding adjustment for WFE. Finally, the incongruent color naming condition, in which a competing word stimulus is present, was modeled by setting *du* at the selected value, and WFE correspondingly, and then running the lemma retrieval stage allowing the word stimulus to activate its lemma for as long as dictated by P3 latency (i.e., *du*). All other parameters were set to the values indicated by Roelofs (2003, p. 124).

Fig. 3 (leftmost column) shows the results of the simulations with WFE ranging between 97 and 297 ms in steps of 5 ms, resulting in *du* ranging between 50 and 250 ms, also in steps of 5 ms. Specifically, the three panels on the left side of the figure plot color naming latency in the control and incongruent condition, as well as Stroop interference (computed as simple time difference), against word reading time (with the same parameter settings). All modeled times are in milliseconds, as determined from the model using the standard step size of 25 ms. On the right side of the figure we have plotted corresponding data from the school sample ($N = 156$). “Reading time” here is a composite formed by the mean of the normalized (*z*-score) values for (untransformed) pseudoword, word, and text reading time. Color naming and interference times are in seconds, as in the preceding analyses. All three panels in each column span the same vertical range, to facilitate slope comparison across measures; however, the relationship between measured seconds (on the right side) and modeled milliseconds (on the left) is arbitrary. Dashed lines through the human data panels correspond to the estimated slope (raw β) from linear regression of each naming measure on the reading time composite.

As expected, and clearly seen in the figure, the effect of our “reading manipulation” on color naming performance in the model is precisely linear. What is the shape of the curve best fitting the corresponding human data? Although the measures are not perfectly comparable, it is possible to test the assumption of linear relationship. A series of monotonic curve-fitting models on the data shown in the panels on the right side of Fig. 3 showed that a gradually decreasing-slope function, such as a power function or S-curve, consistently provided the best and most robust fit. However, after removing a few extreme outliers (two values at $z > 4$ not shown in the figure), the fit of a linear function (adjusted-R^2 of .23 for control, .40 for incongruent, and .28 for interference) was within .02 of the best-fitting

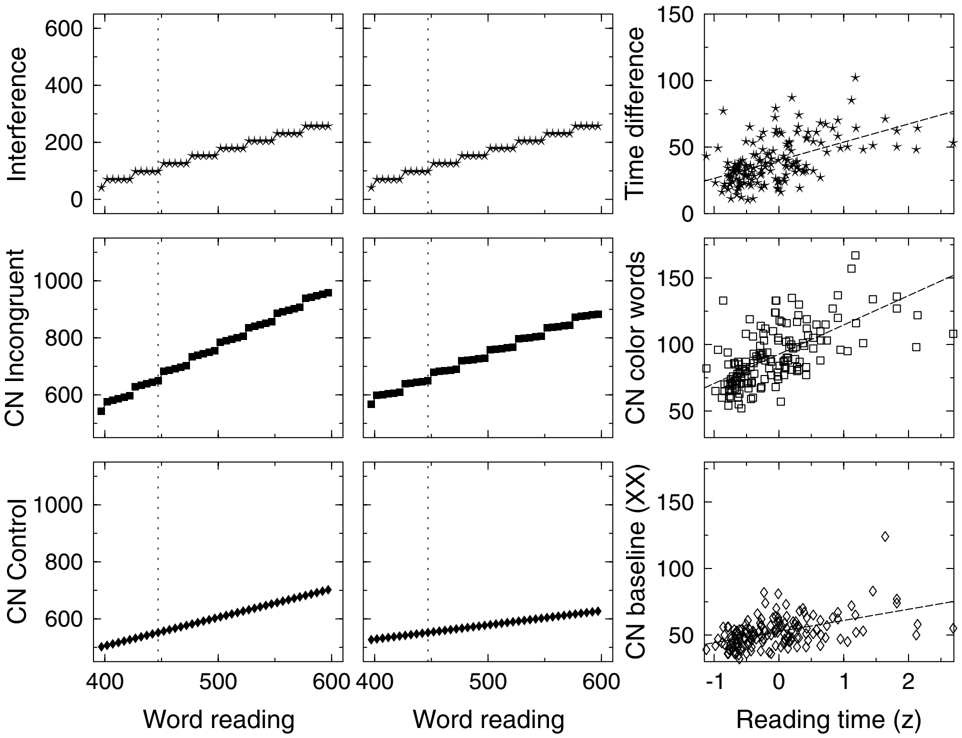

Fig. 3. Results of simulations with the Roelofs (2003) model (left) and corresponding human data from Study 2 (right). Model results in simulated time (ms) per item; human results in seconds per task sheet. The leftmost column plots results for *du* depending on word-form encoding only; the middle column for *du* jointly determined by word-form encoding and visual processing of the word in equal proportions. The dotted vertical line in the model data marks performance at the default parameter settings. In the human data, the dashed line plots the equation resulting from linear regression analysis. CN: color naming.

monotonic function, indicating that linearity constitutes a reasonable approximation for a wide range of reading performance.

The choice of WFE as the locus of poor reading accounts without further assumptions for the positive relationship between reading and color naming, because naming responses also have to pass through the WFE stage before being articulated. This desirable consequence of our modeling choice is in line with our (still incomplete) understanding of naming performance and its relationship to reading. However, even though naming speed for a variety of elements, including letters, digits, colors, and object drawings, is related to reading skill, it is the rapid naming of letters (and secondarily of digits) that is most impaired in poor readers (Wolf et al., 2002). Color naming, in particular, is sometimes found not to be strongly related to reading past the ages of the youngest readers (Semrud-Clikeman, Guy, & Griffin, 2000). The reason for these differences seems to lie in the shared cognitive processes (Wile & Borowsky, 2004) and corresponding brain circuits (Misra, Katzir, Wolf, & Poldrack, 2004) between the reading and rapid naming tasks. Letter naming shares with reading all the regions that color naming shares with reading, such as phonological encoding, plus visual letter-shape processing, a component of letter-string processing, which is known to be impaired in poor readers. Therefore, in our simple modification of the Roelofs (2003) model, the effects of slow reading on color naming may be overestimated. The extent of the overestimate depends on the (unknown) relative contribution of letter-string processing and phonological processing to reading skill variance.

At a second approximation, we can model reading skill variability as arising from two separate factors, including an effect on word reading from WFE, which will affect all color naming tasks because they pass through word-form encoding, and an effect from visual (letter-string) processing, which will not affect baseline color naming at all but will affect color naming in the incongruent condition to the extent that distractor-blocking latency (i.e., *du*) depends on total reading speed. Let LVIS be the total visual processing time for letter-string inputs (relevant for the reading task only[8] and distinct from the modeled visual processing time for color inputs), and let part of this total time contribute to *du* (P3 latency). That is, $du = (\mathrm{WFE} - w) + (\mathrm{LVIS} - l)$ ms. The values of w and l correspond to the fixed components of WFE and LVIS, i.e., the parts not contributing to reading skill variability. (The previous model corresponds to the case in which $w = 47$ and $l = \mathrm{LVIS} = 100$.)

LVIS and WFE can be varied to model reading skill component variability, with the constraint that the value of $du = 100$ will always correspond to $\mathrm{WFE} = 147$ and $\mathrm{LVIS} = 100$, matching the original parameter settings of Roelofs (2003). The relative proportion of influence from each factor is given by the ratio of $(\mathrm{WFE} - w)$ to $(\mathrm{LVIS} - l)$. The results of arbitrarily constraining this proportion to 1:1 ($w = 97$, $l = 50$) for a range of *du* values are shown in the middle column of Fig. 3. Note that the slopes of both color naming conditions with respect to simulated reading speed are decreased, but the magnitude of the interference remains unaffected as compared to the simpler case of allowing *du* to vary as a function of WFE only. Therefore the model is robust in exhibiting the displayed reading–interference relationship, which is of main interest here.

So far we have simply imposed a covariation constraint between pre-existing model parameters, based on a conceptual analysis of the model and on the previous usage of the parameters, and we have allowed the parameters to vary under this constraint. There is no a priori reason that the effect of this constraint on the different tasks should be in proportion to the corresponding relationships in human behavior. Crucially, the extent to which the control and incongruent color naming conditions are affected by covariation of *du* and WFE (and, optionally, LVIS) has not been optimized. As a quantitative index of model fit we can calculate the ratio of the slopes: incongruent vs. reading over control vs. reading. This ratio expresses the relative influence of reading skill on the two color naming conditions. The ratio is 2.061 for the simple version of the model (*du* depending on WFE only), 3.120 for the second version (with *du* depending on WFE and LVIS in equal proportions) and 2.624 for the human data shown in Fig. 3.

A third version of the model, in which the proportions of the two stages affecting reading latency were set to 2:1 (i.e., *du* is determined by two-thirds from WFE and one-third from LVIS) brought the simulated slope in agreement with the human observations at 2.62. The value of this particular optimization is limited, because the human tasks (taking

---

[8] The control color-naming condition is also composed of letters (a string of X's). We assume that this condition is not affected by letter-string processing efficiency, because (a) the letters need not be processed for naming the color, therefore they do not affect color naming directly; (b) the letters cannot be processed as part of word reading, because no word (or pronounceable nonword) is formed, therefore they do not affect color naming indirectly via interference from reading; and (c) even though individual letters can be named we have no reason to assume that there is an obligatory or well-practiced letter-naming process to interfere with color naming. In support of the latter assertion, Brown, Roos-Gilbert, & Carr (1995, Experiment 2) have found that rows of X's do not interfere with color naming, even though they produce Stroop dilution.

several seconds to read lists of words and name lists of colored items) do not precisely match the modeled task (responding to individual stimuli within a few hundred milliseconds). Nevertheless, it indicates the potential of this model for matching the relations between tasks while robustly exhibiting the main aspect of performance under study. Moreover, it generates testable predictions regarding the covariation among tasks and component skills including reading, naming, and interference.

In conclusion, successful manipulation of *du* using the Roelofs (2003) model highlights the role of attentional control in interference, via a mechanism that blocks inappropriate responses, and reinforces the notion that the increase in speed typically associated with reading automatization can be thought of as an index of resource utilization complementary to autonomy (obligatoriness). It remains to determine what type of manipulation might allow modeling of the separate contributions of decoding speed and word reading accuracy to Stroop interference within such a modeling framework.

6.3. Connectionist networks

In a connectionist approach to the Stroop task, the model of Cohen et al. (1990; more recently improved by Cohen et al., 1998) offered a way to dissociate obligatoriness (through the attentional selection “task demand” nodes) from efficiency of processing (through the regular learned pathways from input units). This model combines two layers of units, receiving “stimulus” and “task” activation from a (third) input layer, and two noisy accumulators implementing a separate decision stage based on the activity of the output units. There are two input nodes for each type of stimulus (word and color), one for each color; for example, there is a node signaling the presence of the written word “red” and another indicating red-colored ink. Each “task” (word reading and color naming) is carried out in a dedicated pathway towards the two output nodes (one for each possible color-name response). The strength of each pathway is a result of the amount of training given to the corresponding task, and is therefore a direct index of experience-based efficiency. The task-demand nodes serve to bias the processing of the model in favor of the corresponding pathway, so that the response will be based on the word input when the task is to read, and on the color input when the task is to name the color. The accumulators gather evidence in their favor, in the form of output unit activation, and the first one to exceed a preset threshold is selected for response.

In this model, Stroop interference arises because of the differential efficiency of the two pathways. In a direct implementation of the automaticity-based account of interference, the model needs some time to overcome the activation from the strong reading pathway in the context of the color naming task, whereas color naming is too weak to cause delays in the reading task. In the following section we report simulations using this model, in an effort to identify potential modifications that might bring the automaticity-based account of interference in line with the empirical results observed in our experiments, which seem to run counter to the automaticity predictions. We chose to use the original (1990) version of the model and not a more recent version, because it is the one most clearly and completely described and thoroughly tested in several simulated tasks, and because there is no evidence that the newer models depart significantly from the important properties that make this approach worth examining here.

6.4. Simulation 2

For this simulation we implemented the model as described in Cohen et al. (1990) with the parameter corrections noted by Wiles, Chenery, Hallinan, Blair, and Naumann (2000, i.e., $\sigma = .01$ and $\tau = .1$). All results reported below are based on 1000 repetitions for each input pattern (combination of stimulus and task). Connection weights were copied directly from Cohen et al. (p. 339) without training. We first tried to model the slow reader directly, in the spirit of the original model design, by varying the weights along the reading pathway. Fig. 4 shows the effects of modifying the connection weights of the reading pathway on the mean output latency for each condition of interest (word reading without color input; baseline color naming with no interference; and color naming with an incongruent word), as well as the resulting Stroop interference computed as simple time difference.

As expected from the design of the model, the effects of varying these weights were all in the same direction. Whether modifying only the connections from the input to the hidden units (leftmost column), only from the hidden to the output units (second from left), or

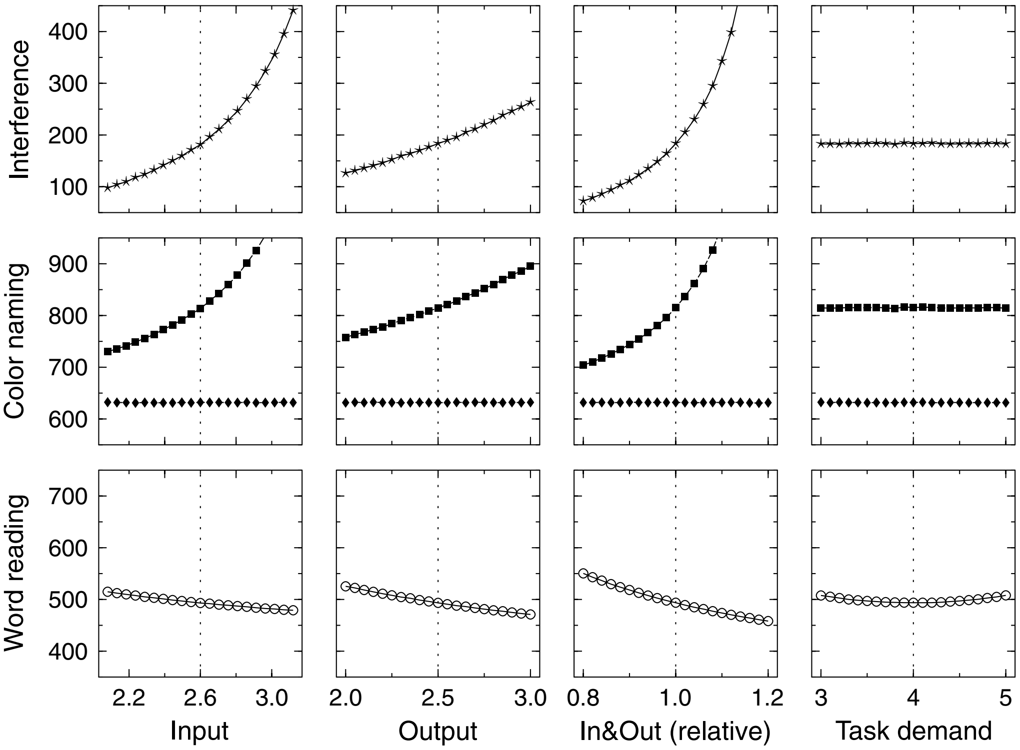

Fig. 4. Results of simulations with a reimplementation of the model of Cohen et al. (1990), in simulated time (ms) as given by the formula $12 * \text{cycles} + 206$. Each point plots the mean of 1000 runs. Bottom row: Model response latencies with the task demand node for reading set to 1.0, and a single stimulus node on the reading pathway input activated (to 1.0). Middle row: Results with the task demand node for color naming set to 1.0, and either only a single node on the color naming pathway input activated (baseline control condition; filled diamonds) or a node on the color naming pathway input and the opposite-label node on the reading pathway input simultaneously activated (color word interference condition; filled squares). Top row: Difference between the corresponding two middle row points, as an index of simulated Stroop interference. Each column plots the effects of varying a set of connection weights on the reading pathway only. From left to right: input-to-hidden; hidden-to-output; both input-to-hidden and hidden-to-output in equal relative proportions; and task demand-to-hidden. Actual connection weight values on the abscissa (common for all rows) except in 2nd from right in which relative proportions of default are shown.

both simultaneously (in the same proportion; third column), a reduction in connection strength resulted in increased reading times and decreased interference, affecting color naming only in the incongruent and not in the control condition. Varying the connection strength from the task demand to the hidden units had negligible effects on either reading or naming response latency (rightmost column).

Thus, affecting the connection weights in this model, unsurprisingly, presents the picture expected from the automaticity account, as far as the relationship between reading efficiency and interference is concerned, in contrast to the empirically observed data. However, this is not the only possible way to affect the performance of the model. In our interpretation of the experimental findings, we have hypothesized that the observed differences in interference are directly attributable to reading skill and not mediated by other factors, for example attentional or other executive function. However, Wiles et al. (2000) reported that the effects of damage to the inhibitory component of the Cohen et al. (1990) model were to increase reaction times marginally in the control condition and markedly in the incongruent condition, a pattern reminiscent of the comparison between good and poor readers. The question arises, therefore, whether such purely "attentional" interventions might provide a reasonable approximation of the empirical findings.

Fig. 5 shows the results of varying the most important model parameters, as they affect the critical measures of word reading and color naming response latency in the conditions of interest. With the exception of bias, modification of which has no appreciable effect on reading performance, an interesting set of behaviors is displayed by the other parameters. Modification of τ, that is, generally slowing down or speeding up the model by affecting the rate of net input integration in each unit, has a consistent result on all measures (leftmost column). Slowing down of the color naming response in the incongruent condition occurs at a higher rate than slowing down in the control condition, thus producing a pattern of increased interference along with slower reading (in contrast to the result reported by Wiles, Chenery, Hallinan, Blair, & Naumann, 2000, perhaps due to a different range of τ values).

Modifications of the response accumulator characteristics, either on the rate or on the threshold (columns on the right side of Fig. 5), also have the effect of simultaneously increasing or decreasing reading and color naming latency. Like the effect of τ modifications, interference also increases with slower reading, because color naming responses in the incongruent condition become slower to a greater extent than responses in the control condition. Thus, on the basis of the direction of these trends alone, there are three potential candidate mechanisms for simulating the empirical relationship between reading and interference. Note that the most fundamentally "attentional" modification, that is, the bias parameter, is not a viable candidate, despite its great effect on interference, because it fails to affect reading, consistent with what one might expect from impoverished attentional inhibition alone.

To further examine the role of the three parameters, Fig. 6 shows the relationship between reading, color naming, and interference, in the same format and scale as Fig. 3 (left), to facilitate comparisons between the two modeling approaches. The simulated times are directly comparable between the two figures as both models claim to simulate veridical response latencies. The ratio of slopes (incongruent vs. reading over control vs. reading) was 2.089 for the τ manipulation, 1.729 for threshold, and 1.633 for bias. Thus for the "general slowing" of the model, the relative effect on interference is as close to the human observations as that of the Roelofs (2003) model with *du* depending only on WFE. However, the slope of the relationship is gradually increasing (panel on left column, middle row of

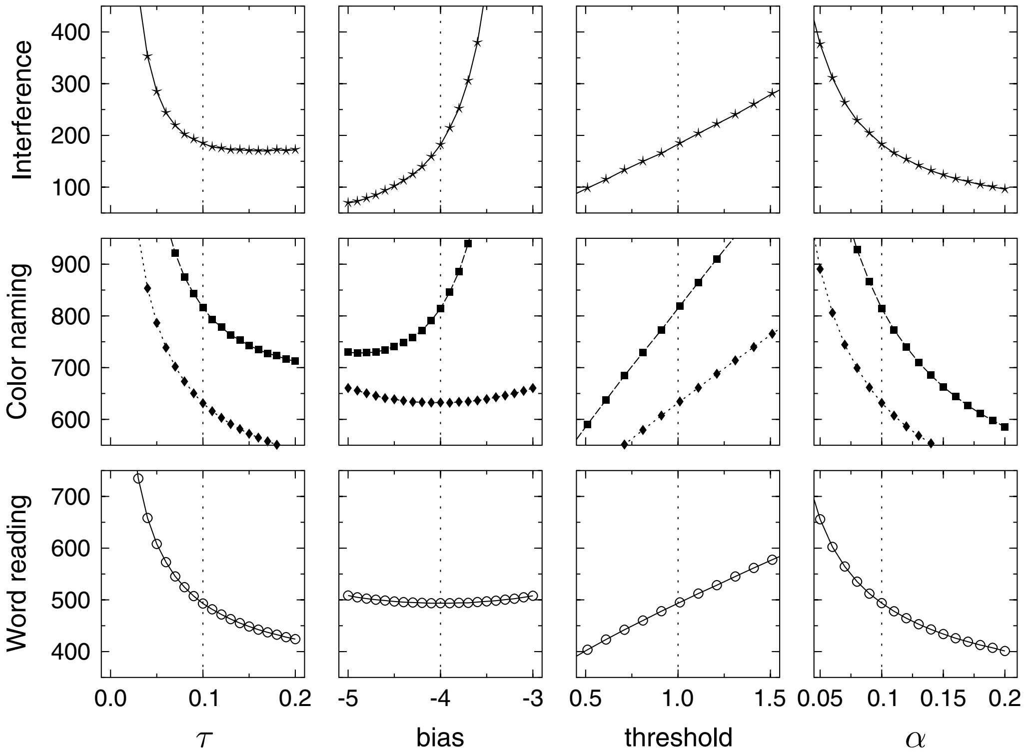

Fig. 5. Results of simulations with a reimplementation of the model of Cohen et al. (1990), in simulated time (ms) as given by the formula 12 ∗ cycles + 206. Each point plots the mean of 1000 runs. Bottom row: Model response latencies with the task demand node for reading set to 1.0, and a single stimulus node on the reading pathway input activated (to 1.0). Middle row: Results with the task demand node for color naming set to 1.0, and either only a single node on the color naming pathway input activated (baseline control condition; filled diamonds) or a node on the color naming pathway input and the opposite-label node on the reading pathway input simultaneously activated (color word interference condition; filled squares). Top row: Difference between the corresponding two middle row points, as an index of simulated Stroop interference. Each column plots the effects of varying a single model-wide parameter. From left to right: rate constant τ; hidden unit bias; accumulator response threshold; and rate of evidence accumulation α. Actual parameter values on the abscissa (common for all rows).

Fig. 6), in contrast to the empirical best-fitting curve which was estimated to be of slowly decreasing slope in the context of Simulation 1 above.

All three of these manipulations incur increased reverse-Stroop interference as a side-effect. For the ranges of simulated reading performance depicted in Fig. 6, the effect of an incongruent color stimulus on word reading ranges between 7–20 ms (τ manipulation), 7.5–10 ms (α), and 7–16 ms (threshold), always increasing with slower reading. Color-naming facilitation from a congruent word stimulus also follows the same trend, gradually increasing as reading latency, interference, and reverse interference increase. This is not surprising, as model-wide parameter settings are expected to exhibit model-wide effects. On the one hand, novel predictions are generated about the relation of all these measures to reading skill; on the other hand, the model may be brought into question if adjustment of parameters to account for one phenomenon (the effect of reading skill) results in diminished model performance in other domains, for which the parameters may have been previously optimized. This is an issue to be tested in future simulations, for which additional data on the relationship between reading and color naming should be collected.

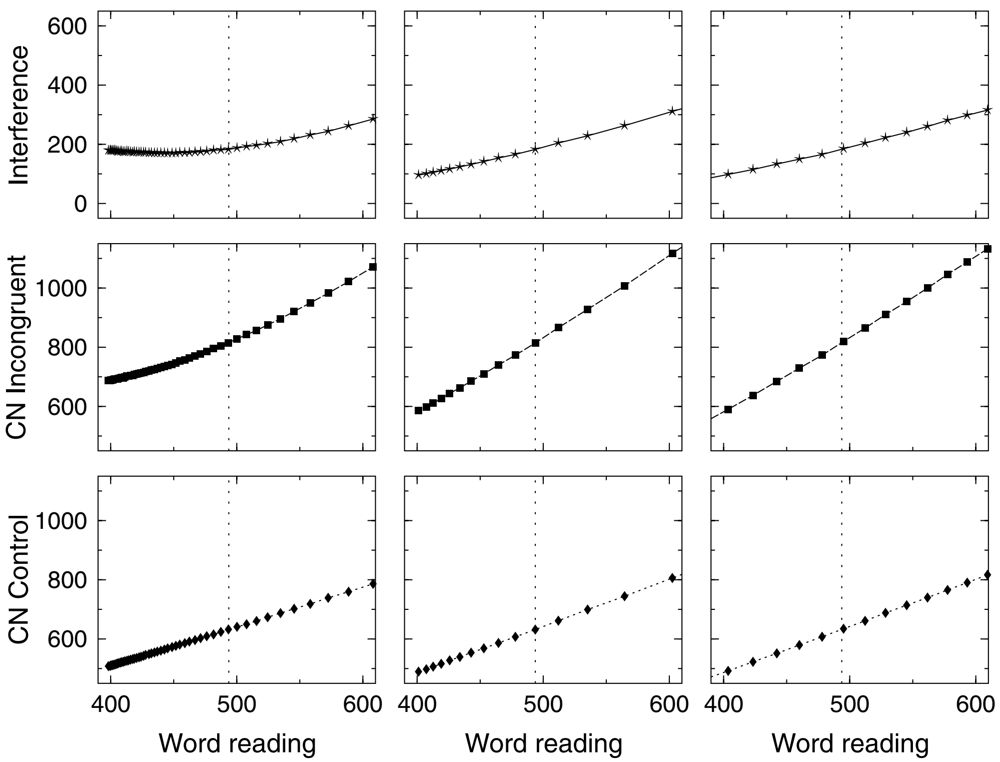

Fig. 6. Simulated color naming response latencies as a function of simulated word reading latencies in a reimplementation of the model of Cohen et al. (1990), in simulated time (ms) per item as given by the formula $12 * \text{cycles} + 206$. Each point plots the mean of 1000 runs. Each column plots the effects of varying a single model-wide parameter. From left to right: rate constant τ; accumulator response threshold; and rate of evidence accumulation α. The dotted vertical line marks performance at the default parameter settings. CN: color naming.

7. Conclusion

We have presented evidence that greater Stroop interference is associated with lower reading skill. In particular, substantial proportions of interference variance can be attributed to decoding speed while additional unique variance is taken up by reading accuracy measures, after controlling for baseline color naming speed. We have interpreted these findings in the context of automatization as supporting a dissociation between obligatoriness and resource efficiency. To account for the effect, a blocking mechanism can be invoked, preventing dominant (obligatory) responses from being produced while allowing their processing speed to determine the non-dominant response latency.

The asymmetry of reading compared to naming, with respect to word form access, may parsimoniously account for this as well as previous findings on interference. This explanation was successfully modeled by the WEAVER++ adaptation to the Stroop effect of Roelofs (2003) without modifications affecting previously reported model performance. In addition to accounting for the observed relationship between reading and Stroop interference, this model has generated a number of testable predictions regarding the relationship among reading and naming tasks, and seems to allow identification of skill components underlying successful reading and naming performance under a variety of task conditions.

It was also possible to simulate aspects of the empirical relationship by modification of parameters in the connectionist model of Cohen et al. (1990) even though the central assumptions underlying this model naturally produce the opposite pattern, in accordance with an undifferentiated account of automaticity. Because of the model-wide effects of our modifications, probably affecting the success of its performance in other cases for which it was previously optimized, we do not wish to draw strong conclusions from these simulations. However, this model highlighted the potential role of general parameters, such as general processing speed or response criterion (threshold), in a complete account of interference effects, which are likely to hinge to a large extent on executive control beyond their modulation by word processing speed.

Attentional control appears to be central in any explanation, while the role of a more specific inhibition control aspect remains to be investigated. In particular, and without underestimating the large effects of reading skill on interference, it seems likely that executive functions and other attentional and cognitive control factors account for additional variance in Stroop interference, in accordance with previous research, including studies of older and demented adults and of children with attention deficit. As the traditional notion of reading automaticity underlying interference seems no longer tenable, future theoretical refinements in our understanding of the reading process and of Stroop interference should take into account this counterintuitive finding on the relationship between the two.

References

Adams, M. J. (1990). *Beginning to read: Thinking and learning about print*. Cambridge, MA: MIT Press.

Altmann, E. M., & Davidson, D. J. (2001). An integrative approach to Stroop: combining a language model and a unified cognitive theory. In *Proceedings of the 23rd annual meeting of the Cognitive Science Society* (pp. 21–26). Hillsdale, NJ: Erlbaum.

Alwitt, L. F. (1966). Attention in a visual task among non-readers and readers. *Perceptual and Motor Skills, 23*, 361–362.

Anderson, J. R., & Lebiere, C. (1998). *The atomic components of thought*. Mahwah, NJ: Erlbaum.

Brown, T. L., Roos-Gilbert, L., & Carr, T. H. (1995). Automaticity and word perception: evidence from Stroop and Stroop dilution effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*(6), 1395–1411.

Cohen, J. D., Dunbar, K., & McClelland, J. L. (1990). On the control of automatic processes: a parallel distributed processing account of the Stroop effect. *Psychological Review, 97*, 332–361.

Cohen, J. D., Usher, M., & McClelland, J. L. (1998). A PDP approach to set size effects within the Stroop task: reply to Kahne, Balota, Spieler, and Faust (1998). *Psychological Review, 105*, 188–194.

Comalli, P. E., Wapner, S., & Werner, H. (1962). Interference effects of Stroop color-word test in childhood, adulthood and aging. *The Journal of Genetic Psychology, 100*, 47–53.

Ehri, L. C. (1995). Phases of development in learning to read words by sight. *Journal of Research in Reading, 18*, 116–125.

Ehri, L. C., & Wilce, L. S. (1979). Does word training increase or decrease interference in a Stroop task? *Journal of Experimental Child Psychology, 27*, 352–364.

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*(4), 438–468.

Everatt, J., McCorquodale, B., Smith, J., Culverwell, F., Wilks, A., Evans, D., et al. (1999). Associations between reading ability and visual processes. In J. Everatt (Ed.), *Reading and dyslexia: Visual and attentional processes* (pp. 1–39). London, UK: Routledge.

Everatt, J., Warner, J., & Miles, T. R. (1997). The incidence of Stroop interference in dyslexia. *Dyslexia, 3*, 222–228.

Georgas, D. D., Paraskevopoulos, I. N., Bezevegis, I. G., & Giannitsas, N. D. (1997). *Elliniko WISC-III: Wechsler klimakes noimosinis gia paidia* [Greek WISC-III: Wechsler intelligence scales for children]. Athens: Ellinika Grammata.

Glaser, M. O., & Glaser, W. R. (1982). Time course analysis of the Stroop phenomenon. *Journal of Experimental Psychology: Human Perception and Performance, 8*(6), 875–894.

Helland, T., & Asbjørnsen, A. (2000). Executive functions in dyslexia. *Child Neyropsychology, 6*, 37–48.

Kelly, M. S., Best, C. T., & Kirk, U. (1989). Cognitive processing deficits in reading disabilities: a prefrontal cortical hypothesis. *Brain and Cognition, 11*, 275–293.

Kuhn, M. R., & Stahl, S. A. (2003). Fluency: a review of developmental and remedial practices. *Journal of Educational Psychology, 95*, 3–21.

LaBerge, D., & Samuels, S. J. (1974). Toward a theory of automatic information processing in reading. *Cognitive Psychology, 6*, 293–323.

Lachter, J., Forster, K. I., & Ruthruff, E. (2004). Forty-five years after Broadbent (1958): still no identification without attention. *Psychological Review, 111*(4), 880–913.

Leinonen, S., Müller, K., Leppänen, P. H. T., Aro, M., Ahonen, T., & Lyytinen, H. (2001). Heterogeneity in adult dyslexic readers: relating processing skills to the speed and accuracy of oral text reading. *Reading and Writing: An Interdisciplinary Journal, 14*, 265–296.

Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences, 22*, 1–38.

Levy, B. A. (2001). Moving the bottom: improving reading fluency. In M. Wolf (Ed.), *Dyslexia, fluency and the brain* (pp. 357–379). Timonium, MD: York Press.

Logan, G. D. (1988). Toward an instance theory of automatization. *Psychological Review, 95*, 492–527.

Logan, G. D. (1997). Automaticity and reading: perspectives from the instance theory of automatization. *Reading and Writing Quarterly, 13*, 123–146.

Logan, G. D. (2002). An instance theory of attention and memory. *Psychological Review, 109*, 376–400.

Lovett, M. C. (2001). Modeling selective attention: Not just another model of Stroop (NJAMOS). In *Proceedings of the fourth international conference on cognitive modeling* (pp. 151–156). Mahwah, NJ: Erlbaum.

MacLeod, C. M. (1991). Half a century of research on the Stroop effect: an integrative review. *Psychological Bulletin, 109*, 163–203.

MacLeod, C. M., & Dunbar, K. (1988). Training and Stroop-like interference: evidence for a continuum of automaticity. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14*, 126–135.

Miozzo, M., & Caramazza, A. (2003). When more is less: a counterintuitive effect of distractor frequency in the picture–word interference paradigm. *Journal of Experimental Psychology: General, 132*, 228–252.

Misra, M., Katzir, T., Wolf, M., & Poldrack, R. A. (2004). Neural systems for rapid automatized naming in skilled readers: unraveling the RAN-reading relationship. *Scientific Studies of Reading, 8*(3), 241–256.

Perfetti, C. A., & Bolger, D. J. (2004). The brain might read that way. *Scientific Studies of Reading, 8*(3), 293–304.

Protopapas, A., & Skaloumbakas, C. (in press). Computer-based and traditional screening and diagnosis of reading disabilities in Greek. *Journal of Learning Disabilities.*

Raven, J. (1976). *Standard progressive matrices.* The Psychological Corporation.

Risko, E. F., Stolz, J. A., & Besner, D. (2005). Basic processes in reading: Is visual word recognition obligatory? *Psychonomic Bulletin & Review, 12*(1), 119–124.

Roelofs, A. (2003). Goal-referenced selection of verbal action: modeling attentional control in the Stroop task. *Psychological Review, 110*, 88–125.

Roelofs, A. (2005). From Popper to Lakatos: A case for cumulative computational modeling. In A. Cutler (Ed.), *Twenty-first century psycholinguistics: Four cornerstones* (pp. 313–330). Hillsdale, NJ: Lawrence Erlbaum.

Roelofs, A., & Hagoort, P. (2002). Control of language use: cognitive modeling of the hemodynamics of Stroop task performance. *Cognitive Brain Research, 15*, 85–97.

Salmelin, R., & Helenius, P. (2004). Functional neuroimaging of impaired reading in dyslexia. *Scientific Studies of Reading, 8*(3), 257–272.

Samuels, S. J. (1999). Developing reading fluency in learning-disabled students. In R. J. Sternberg & L. Spear-Swerling (Eds.), *Perspectives on learning disabilities: Biological, cognitive, contextual* (pp. 176–189). Boulder, Colorado: Westview.

Samuels, S. J., & Flor, R. F. (1997). The importance of automaticity for developing expertise in reading. *Reading and Writing Quarterly, 13*, 107–121.

Samuels, S. J., LaBerge, D., & Bremer, C. D. (1978). Units of word recognition: evidence for developmental changes. *Journal of Verbal Learning and Verbal Behavior, 17*, 715–720.

Savage, R. (2004). Motor skills, automaticity and developmental dyslexia: a review of the research literature. *Reading and Writing: An Interdisciplinary Journal, 17*, 301–324.

Schiller, P. H. (1966). Developmental study of color-word interference. *Journal of Experimental Psychology, 72*, 105–108.

Semrud-Clikeman, M., Guy, K., & Griffin, J. D. (2000). Rapid naming deficits in children and adolescents with reading disability and attention deficit hyperactivity disorder. *Brain and Language, 74*, 70–83.

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

Spieler, D. H., Balota, D. A., & Faust, M. E. (1996). Stroop performance in healthy younger and older adults and in individuals with dementia of the Alzheimer's type. *Journal of Experimental Psychology: Human Perception and Performance, 22*(2), 461–479.

Stanovich, K. E. (1990). Concepts in developmental theories of reading skill: Cognitive resources, automaticity, and modularity. *Developmental Review, 10*, 72–100.

Stanovich, K. E., Cunningham, A. E., & West, R. F. (1981). A longitudinal study of the development of automatic recognition skills in first graders. *Journal of Reading Behavior, 13*, 57–74.

Strayer, D. L., & Kramer, A. F. (1990). An analysis of memory-based theories of automaticity. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16*, 291–304.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology, 18*, 643–662.

Su, Y.-F. (1997). *Indicators of automaticity in word recognition*. Unpublished doctoral dissertation. Dissertation Abstracts International 58(5-A), 1589.

van den Bos, K. P., Zijlstra, B. J. H., & van den Broeck, W. (2003). Specific relations between alphanumeric-naming speed and reading speeds of monosyllabic and multisyllabic words. *Applied Psycholinguistics, 24*, 407–430.

van der Leij, A., & van Daal, V. H. P. (2000). Automatization aspects of dyslexia: speed limitations in word identification, sensitivity to increasing task demands, and orthographic compensation. *Journal of Learning Disabilities, 32*, 417–428.

van der Schoot, M., Licht, R., Jorsley, T. M., & Sergeant, J. A. (2000). Inhibitory deficits in reading disability depend on sybtype: guessers but not spellers. *Child Neuropsychology, 6*, 297–312.

Wile, T. L., & Borowsky, R. (2004). What does rapid automatized naming measure? A new RAN task compared to naming and lexical decision. *Brain and Language, 90*, 47–62.

Wiles, J., Chenery, H., Hallinan, J., Blair, A., & Naumann, D. (2000). Effects of damage to the CDM Stroop model. In *Proceedings of the fifth biennial conference of the Australasian Cognitive Science Society*. Adelaide: Causal.

Wolf, M., & Bowers, P. (1999). The "double-deficit hypothesis" for the developmental dyslexias. *Journal of Educational Psychology, 91*, 1–24.

Wolf, M., & Katzir–Cohen, T. (2001). Reading fluency and intervention. *Scientific Studies of Reading, 5*, 211–239.

Wolf, M., Miller, L., & Donnelly, K. (2000). The retrieval, automaticity, vocabulary elaboration, orthography (RAVE-O): a comprehensive fluency-based reading intervention program. *Journal of Learning Disabilities, 33*, 375–386.

Wolf, M., O'Rourke, A. G., Gidney, C., Lovett, M., Cirino, P., & Morris, R. (2002). The second deficit: an investigation of the independence of phonological and naming-speed deficits in developmental dyslexia. *Reading and Writing: An Interdisciplinary Journal, 15*, 43–72.

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