

The emergence of automaticity in reading: effects of orthographic depth and word decoding ability on an adjusted Stroop measure

Article (Accepted Version)

Megherbi, Hakima, Elbro, Carsten, Oakhill, Jane, Segui, Juan and New, Boris (2017) The emergence of automaticity in reading: effects of orthographic depth and word decoding ability on an adjusted Stroop measure. *Journal of Experimental Child Psychology*, 166. pp. 652-663. ISSN 0022-0965

This version is available from Sussex Research Online: <http://sro.sussex.ac.uk/id/eprint/70459/>

This document is made available in accordance with publisher policies and may differ from the published version or from the version of record. If you wish to cite this item you are advised to consult the publisher's version. Please see the URL above for details on accessing the published version.

Copyright and reuse:

Sussex Research Online is a digital repository of the research output of the University.

Copyright and all moral rights to the version of the paper presented here belong to the individual author(s) and/or other copyright owners. To the extent reasonable and practicable, the material made available in SRO has been checked for eligibility before being made available.

Copies of full text items generally can be reproduced, displayed or performed and given to third parties in any format or medium for personal research or study, educational, or not-for-profit purposes without prior permission or charge, provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way.

The Emergence of Automaticity in Reading:

Effects of Orthographic Depth and Word Decoding Ability on an adjusted Stroop measure

Abstract

Aims

How long does it take for word reading to become automatic? Does the appearance and development of automaticity differ as a function of orthographic depth (e.g. French vs. English)? These questions were addressed in a longitudinal study of English and French beginning readers. The study focused on automaticity as obligatory processing as measured in the Stroop test.

Method

Measures of decoding ability and the Stroop effect were taken at three time points during the first grade (and 2nd grade in the UK) in 84 children. The study was the first to adjust the classic Stroop effect for inhibition (of distracting colors).

Results

The adjusted Stroop effect was zero in the absence of reading ability, and it was found to develop in tandem with decoding ability. After a further control for decoding, no effects of age or orthography were found on the adjusted Stroop measure.

Conclusion

The results are in line with theories of the development of whole word recognition that emphasize the importance of the acquisition of the basic orthographic code.

The Emergence of Automaticity in Reading:

**Effects of Orthographic Depth and Word Decoding Ability on an adjusted Stroop
measure**

The development of reading ability takes several years. Children usually begin to learn to decode words in primary school between the ages of four and seven, depending on the country of schooling. Learning the basic grapheme-phoneme conventions ('rules') typically takes up to two years, depending on the number and complexity of the rules of the orthography (Frith, Wimmer, & Landerl, 1998; Goswami, Ziegler, & Richardson, 2005; Seymour, Aro, & Erskine, 2003; Ziegler & Goswami, 2006). Children's word decoding skills, particularly decoding speed, continue to progress throughout primary and secondary school until adult levels are reached at 200-300 words per minute for continuous texts. Importantly, word decoding gradually becomes automatic (or 'encapsulated') so that it does not interfere with comprehension processes in reading.

Facets of Automaticity

Automaticity of cognitive processing is not an all-or-none feature of mental processes but rather a set of properties that do not necessarily co-occur (Moors & de Houwer 2006; Stanovich, 1990). There are at least four discernible properties of automaticity: speed, effortlessness, autonomy, and lack of conscious awareness (Logan, 1997; Kuhn, Schwanenflugel, Meisinger, et al., 2010). First, automatic word decoding is *fast*, resulting in word identification and initiation of naming well within a second. Indeed, it is so fast that written words are recognized 'immediately' as 'sight words' without overt signs of single letter-sound decoding (e.g. Ehri, 2005). Second, automatic word decoding is also *effortless* in the sense that it allows the reader to simultaneously think of the contents of the text or even to let thoughts drift. This facet of

THE EMERGENCE OF AUTOMATICITY IN READING

automaticity attracted much attention following the influential theory of the development of automaticity of word decoding by LaBerge and Samuels (1974). The dual-task paradigm has been the most commonly used in the study of this facet of automaticity. In this paradigm, the processing penalty of decoding on simultaneous processing of other tasks has been shown to diminish with reading ability (e.g., Horn & Manis, 1987). Third, automatic word recognition is executed *obligatorily* so that once the reader has set eyes on the printed word, the identity of the word – its spoken name and meaning – *will* be available to the reader. The reader cannot decide *not* to identify the word. This feature has been widely explored in the Stroop task (named after J. Ridley Stroop, 1935). In this task, participants are asked to name *the color* in which words are printed while avoiding reading the words, e.g. the word *green* is printed in blue, and the correct response is “blue” (MacLeod, 1991).

These three facets of automaticity may all be a consequence of a fourth property: modularity or ‘encapsulation’ of word decoding from conscious thought (Fodor, 1983; Stanovich, 1990; Kahneman, 2011). If automatic processing is shielded from conscious control, it allows for the development of great speed and effortlessness – at the expense of conscious control of processing. However, the three facets of automaticity do not appear to develop in synchrony. Stroop effects – indicating obligatory processing – have been shown to occur at the end of the first grade in American school children (Ehri & Wilce, 1979; Stanovich, Cunningham, & West, 1981; West & Stanovich, 1978, 1979). This early occurrence does not entail that word recognition also occurs entirely immediate or without cognitive effort from the first grade. For example, even second graders take longer to name number words than to name the corresponding digits (Ehri & Wilce, 1983). Automaticity in the senses ‘immediate’ and ‘effortless’ continues to develop long after the emergence of the Stroop effect (e.g. Ehri & Wilce, 1979).

There may be several reasons for this asynchrony of the development of different facets of automaticity in reading. One is that some words may be recognized as wholes, so-called ‘sight words’, even though accuracy is far from perfect at the very beginning of reading development (Ehri, 2015). If words are recognized as wholes in the Stroop task, then such rapid and immediate recognition may interfere with naming of the colors in which they are presented. However, this does not mean that word recognition is effortless because, as long as orthographic representations of words are linked to the lexicon only by partial grapheme-phoneme correspondences, they are unreliable and prone to lead to recognition mistakes. So even though some words are recognized successfully as sight words, the reader has to monitor the outcome closely and be prepared to correct mistakes. This monitoring takes time and mental resources. A second reason may be that automatic *letter-sound* activation is enough to create some degree of interference in the Stroop task even though *whole word* recognition has *not* become automatic. Letter-sound interference may be possible in the Stroop task because of the very limited number of relevant (color) words. For example, the initial letter sound (e.g. “r”) plus the semantic information ‘common color’ may lead to an activation of *red* and cause interference with the naming of the printed color (e.g. blue).

A third reason may be that *for young children* the Stroop effect over-estimates the specific interference effect of mandatory processing in reading. This possibility was taken into account in the current study.

Early onset of automaticity

A small number of studies have investigated the Stroop effect in first-grade children (Ehri & Wilce, 1979; Stanovich et al., 1981; Schiller, 1966; Schadler & Thissen, 1981). These studies mostly tested children aged 6-7 years at the end of the school year (at least 8 months after beginning instruction). Some studies have reported significant Stroop interference effects at the

end of the school year (e.g. Schadler & Thissen, 1981; Stanovich et al., 1981) while others have not (e.g. Schiller, 1966). Only Stanovich et al. (1981) investigated the emergence of the Stroop effect during the course of the first grade (in late September, February, and late April). A robust Stroop effect was already detected in February grade 1, i.e. after 5 months of instruction. Somewhat surprisingly, the effect was significant even for poor readers. It seems surprising that even poor readers have developed some automaticity in word recognition after less than half a year's instruction. It may be that the first graders recognize some words as wholes, as 'sight words'. But in the case of poor readers, such sight words will be few. Hence, one may wonder whether the traditional Stroop measure provides an overestimate of automaticity.

Another set of findings may also suggest that the Stroop measure overestimates automaticity in beginning readers. Oddly, some studies have found that rather young readers (in grades 2 - 3) display Stroop effects that may be stronger than those seen in older readers (Bonino & Ciairano, 1997; Comally, Wapner, & Werner, 1962; Guttentag & Haith, 1978; Ikeda, Okuzumi & Kokubun, 2011; Peru, Faccioli, & Tassinari, 2006; Protopapas, Archonti, & Skaloumbakas, 2007; Schiller, 1966). Indeed, Schiller (1966) found an inverted U-shaped pattern of development with no interference in first-graders, maximal interference in second- and third-graders and a gradual decline beyond then and into college. In addition, other studies have found that poor readers display stronger Stroop effects than age matched controls do (Everatt, Warner, & Miles, 1997; Protopapas et al. 2007). If the Stroop effect is solely an indicator of mandatory processing, one would expect the effect to grow stronger, not to diminish with reading experience and skill.

Inhibition in the Stroop Effect

One possibility is that the Stroop effect is a compound of at least two effects. One is the obligatory decoding of the distracting words. The other is the extent to which the individual is

able to block out, to suppress, or inhibit the potential distraction (Protopapas et al., 2007). Evidence for such a blocking mechanism comes from studies in which naming of incongruent stimuli was trained with the consequence that the Stroop effect was reduced (Protopapas, Vlahou, Moirou, & Ziaka, 2014). One aim of the current study was to develop *a pure measure of the distraction effect in the Stroop task* by means of a separate control for individual inhibition.

The cognitive processes underlying the Stroop task are considered to rely at least partially on executive processes (e.g. McLeod, 1991; McDonald, Beauchamp, Crigan, & Anderson, 2014). Executive functions – including inhibition functions – are still developing at the age of school entry (and beyond) (e.g. Adleman, Menon, Blasey, White, et al., 2002; Brocki & Bohlin, 2004; Dash & Dash, 1982; Diamond & Taylor, 1996). For example, McDonald et al. (2014) reported a substantial age-related improvement in inhibition from 5 to 8 years by means of a Stroop task paradigm (see also, Schiller, 1966). Hence it appears that this development should be taken into account when the Stroop effect is interpreted in children in the early grades.

To the best of our knowledge, the present study is the first to attempt to control for inhibition in the Stroop effect. The control for inhibition was obtained by means of a separate measure of semantic distraction on *reading*, i.e. reading aloud words printed in distracting colors (Figure 1, see also the Methods section below).

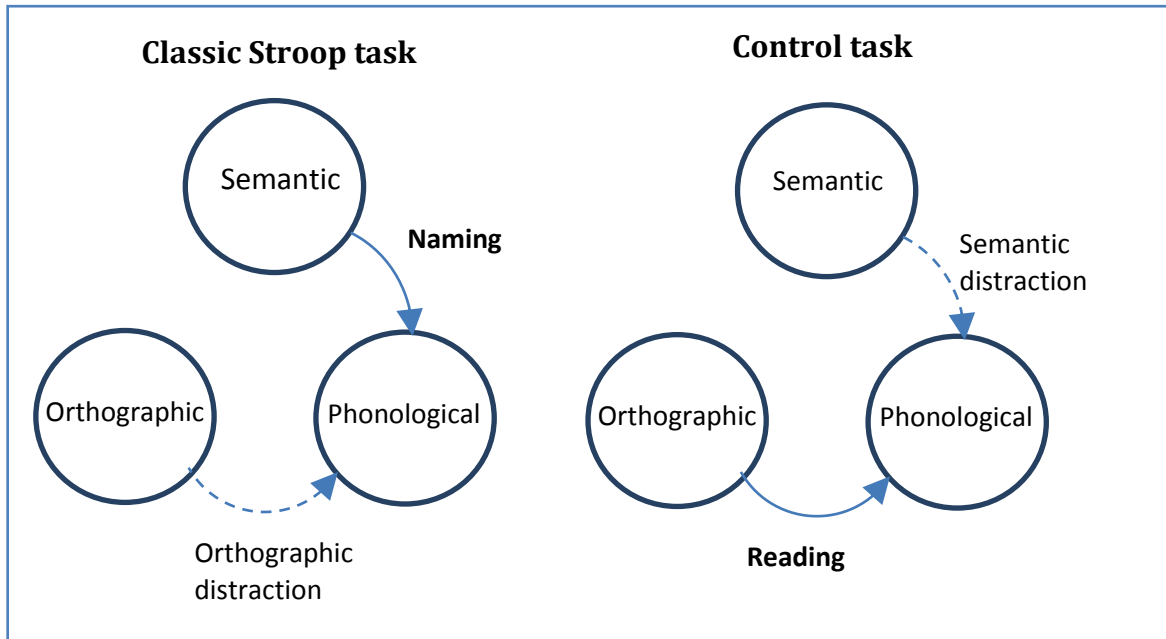


Figure 1. Activation of lexical components in the classic Stroop task (left) and in the control task (right). The classic Stroop task measures orthographic distraction on a naming task. The control task measures semantic distraction on a reading task. Both tasks required phonological activation of the same color words.

In both the classic Stroop task and in the control task, the distraction effect is the result of two opposing features: an involuntary activation and the inhibition of this activation. The control task was considered to be a relatively pure measure of inhibition because the semantic activation part of the distraction effect (seeing distracting colors) was assumed to be fairly constant across participants. This assumption was in line with the finding that individual differences in color naming (with no distraction) were much smaller than the individual differences in reading (see the results section). This and other assumptions are taken up in the discussion section below.

Orthographic Depth and the Stroop effect

In addition to the control for inhibition in the Stroop measure, the present study also explored the onset of automaticity in an orthography other than English.

There is evidence that orthographic complexity ('depth') influences initial reading acquisition (Aro & Wimmer, 2003; Ellis, Natsume, Stavropoulou, Hoxhallari, van Daal, & Petalas, 2004; Frith et al., 1998; Seymour et al., 2003; van Daal & Waas, 2016; Ziegler & Goswami, 2006). English orthography is more difficult than, say, French simply because English has a much higher number of basic grapheme-phoneme correspondences ('rules') to be acquired and more exceptions from these correspondences (e.g. Aro & Wimmer, 2003; Joshi & Aaron, 2013; Seymour et al., 2003). Consequently, it takes longer for English children to acquire the basic code and to take advantage of a self-teaching mechanism (Share, 2008) by which they continue to acquire word-specific knowledge through independent reading.

However, it is not known whether such a prolonged period of basic code acquisition is a barrier to the development of automaticity. On the one hand, English speaking children are exposed to written words from grade 1 (and sometimes before), and they might learn to recognize some of the frequent words automatically. Even if their word recognition is based on logographic or partial phonetic cues, some words might still be recognized in an immediate and obligatory manner as 'sight words' (e.g. Frith, 1986; Ehri, 2015). On the other hand, there may not be such short cuts to automaticity. It may be that automaticity develops as a function of (and following) the development of word decoding ability. A certain (modest) mastery of grapheme-phoneme connections may be necessary in order to develop fluency and automaticity with more than a few words (Ehri, 2015; Juul, Poulsen & Elbro, 2014).

The present study investigated this issue in two ways – by comparisons of the Stroop effect in English and French speaking first graders, and by comparisons of individual Stroop effects with word decoding ability.

In sum, the general aim of the present study was to explore the emergence of automaticity of word reading in beginning readers. Automaticity was here defined as obligatory recognition. It was operationalized as the Stroop effect adjusted for (degree of) inhibition. The research questions were the following:

1. When does automaticity emerge?
2. Does the emergence of automaticity vary with the orthographic depth of the language in which the child is learning to read, e.g. in English versus French?
3. Are there unique effects of orthographic depth, or can differences between English and French be fully explained by decoding ability?

Method

Design

Measures of automaticity, inhibition, and word decoding were taken in French 1st graders and in English 1st and 2nd graders. Since decoding development is considerably slower in the UK than in France, English 2nd graders were included to allow for a comparison across orthographies of children at about the same level of decoding. All children were tested at three time points across the school year, in October, February and May. The beginning of the school year is at the beginning of September in both France and the U.K.

Participants

French: The participants were 32 first-grade children from a primary school in Paris (13th arrondissement) whose parents returned permission forms. They were 14 boys and 18 girls, mean age: 6;4 years (SD 3.4 months) in session 1. All children participated in all three sessions. Children attended a school that recruits students from a wide range of socioeconomic

backgrounds. They were born in France and had all been exposed to French from birth; they were not repeating the grade and presented no behavior problems or color blindness according to their teachers. The children were of about average reading ability – both according to their teachers and by comparison with the French classes in the study by Seymour et al. (2003, table 6).

British, Grade 1: In session 1, the participants were 29 children from two primary schools in Brighton whose parents granted permission. Among them, one child moved to a new school during the year, and two had no score for reading in session 1 because they failed to read the words during the training. Thus, the analyses were conducted on 26 children. They were 13 boys, and 13 girls and the mean age was 4;11 (SD 2.8 months) at session 1. Schools were located in an area with a wide range of socioeconomic backgrounds. All children were born in England and had been exposed to English from birth. They had no behavior or color vision problems or reading disability according to their teachers. The classes were of average of reading ability and made a normal progress according to their teachers. The children's decoding scores in the May session are close to those reported by Seymour et al. (2003, table 6).

British, Grade 2: Following similar criteria as for the first graders, 36 children participated from two primary schools in Brighton in all three sessions. They were 21 boys and 14 girls, mean age 5;8 in session 1, SD=3.7 months) at session 1.

Materials and Procedures

Stroop and inhibition measures

In order to assess *automaticity* of word decoding, the participants performed the classical Stroop experiment on paper: *naming* colors with and without interfering words (Figure 2, upper panels). Pseudo-letters were used in the non-interference condition. They were constructed to match the real letters' physical properties (height, number of pixels, contiguous pixels) by re-

configuring the features of the original letters (New, Doré-Mazars, Cavézian, Pallier, & Barra, 2016).

To assess *inhibition*, the participants were also asked to *read* the color words with and without distracting print colors (Figure2, lower panels).

	Without interference				With interference					
Naming	⇨	ታላቅ	አዳጋላ	የጋጋ	ቀጥጎ	⇨	YELLOW	GREEN	RED	BLUE
colors		የጋጋ	አዳጋላ	ቀጥጎ	ታላቅ	RED	GREEN	BLUE	YELLOW	
		ቀጥጎ	ታላቅ	አዳጋላ	የጋጋ	BLUE	YELLOW	GREEN	RED	
		ታላቅ	አዳጋላ	የጋጋ	ቀጥጎ	YELLOW	GREEN	RED	BLUE	
		አዳጋላ	የጋጋ	ቀጥጎ	ታላቅ	BLUE	RED	GREEN	YELLOW	
Reading	⇨	RED	BLUE	GREEN	YELLOW	⇨	BLUE	RED	GREEN	YELLOW
words		BLUE	YELLOW	RED	GREEN	RED	GREEN	YELLOW	BLUE	
		GREEN	RED	YELLOW	BLUE	YELLOW	BLUE	RED	GREEN	
		BLUE	RED	YELLOW	GREEN	BLUE	YELLOW	GREEN	RED	
		BLUE	RED	YELLOW	GREEN	BLUE	YELLOW	GREEN	RED	

Figure 2. The four conditions for the Stroop and inhibition measures. ‘Naming colors with interference’ is the classic Stroop task.

Stimuli for each of the four conditions consisted of a matrix of four by four words or colored pseudo-letter strings. The same four colors were repeated four times in each matrix. The participants were told to name the colors of the ink (naming tasks) or to read the words (reading task) as accurately and as rapidly as possible. To exclude possible effects of task and condition order, the two tasks and two conditions (Figure 2) were completely counterbalanced across participants and within participants across the three time points. Further, the order of the items (the order of the four lines of the matrix) were also systematically counterbalanced across tasks and conditions.

Each matrix presentation was preceded by a practice session with four example stimuli. This practice was necessary to ensure that children had understood the task and were able to perform it. If a child could not read a (color) word, the experimenter helped by encouraging the child to decode the word letter by letter, or syllable by syllable, or by guessing. When a child failed to read an example word, the experimenter gave corrective feedback. The four example items in each condition were practiced until the child gave the correct responses twice in a row with a maximum of four practice rounds.

Children were corrected during the experimental conditions if they made two consecutive errors. The measure taken from the experimental conditions was the total time the child took to read or name each matrix of 16 items.

Decoding

Decoding was assessed by means of the pseudo-word reading test from Seymour et al. (2003). The test has lists of 36 one-syllable and two-syllable pseudo-words for different European languages, e.g. a French list and an English list. The lists have pseudo-words of similar structures (avoiding clusters and complex graphemes), and the pseudo-words are equally pronounceable across languages. The children were given one minute to read as many of the 36 pseudo-words as possible. The score was the number of pseudo-words correctly read within a maximum of one minute.

Results

Standard vs. adjusted Stroop effects

Participants made very few uncorrected errors, typically between zero and two per condition. The mean error rates mirrored the response times (tables 1 and 2 below) with more errors in the interference conditions, and a falling error rate in the reading condition over time. Hence, errors were not analyzed further.

THE EMERGENCE OF AUTOMATICITY IN READING

Reaction times that were more than 3 standard deviations from the mean of the participants were considered as outliers and replaced with values 3 SD from the sample mean. This procedure affected fewer than 1 % of the individual measures. The interference effects in the naming tasks (original Stroop effect) and reading tasks (inhibition measure) were calculated as the ratios between the response times with and without interference. In this way, general differences between individual response latencies were controlled. The scores were normally distributed. The adjusted Stroop effect was then calculated as the original Stroop effect minus the inhibition measure.¹ Again, scores were normally distributed.

Table 1 shows the French mean reading and the naming times with and without interference. Results for each of the three time points are shown with the calculated interference effects and the adjusted Stroop effect.

THE EMERGENCE OF AUTOMATICITY IN READING

Table 1. French 1st grade mean response times (in seconds, and SD) for reading and naming with and without interference. The standard Stroop measure is the interference effect in the naming task^a. The adjusted Stroop measure is the difference between the interference effect in the naming and the reading tasks. Decoding ability is reported in the final column (with SD and range in brackets).

Time point	Task	Without interference	With interference	Interference effect	Adjusted Stroop effect	Decoding ability
Oct.	Reading	41.6 (31.1)	47.9 (32.1)	1.37** (0.70)		4.7 (5.5; 0-22)
	Naming	17.4 (4.8)	22.6 (8.4)	1.37** (0.53) ^a	-.01 (0.86)	
Feb.	Reading	19.1 (8.8)	23.2 (11.6)	1.26** (0.42)		18.8 (8.8; 1-32)
	Naming	17.2 (7.3)	27.4 (8.3)	1.77*** (0.61) ^a	.50 (0.68)***	
May	Reading	10.7 (3.0)	13.9 (4.9)	1.31*** (0.34)		30.1 (5.5; 19-32)
	Naming	14.6 (3.2)	29.6 (10.0)	2.08*** (0.69) ^a	.77 (0.79)***	

^a The classic Stroop effect. ** $p < .01$, *** $p < .001$ (one sample t-test)

The standard Stroop effect (interference effect on color naming) was found to be significant at all three time points (one sample t-test of mean difference from 1). The emergence of a standard Stroop effect at the very first time point (October) was surprising given the fact that half of the children were practically unable to read (they read fewer than 3 of the 36 pseudo-words correctly). However, the interference effect in the reading condition (the measure of “inhibition”) was also significant at all-time points, suggesting that the traditional Stroop effect may have been inflated by the mere presence of an interfering input. In contrast, the adjusted Stroop effect was only significant at the second and third time point. A developmental trend in the adjusted Stroop effect was present (one-way ANOVA with three time points indicated a significant trend ($F(2, 92) = 8.1$ $p < .01$ partial $\eta^2 = .15$). Table 2 shows the mean scores for the English 1st and 2nd graders.

THE EMERGENCE OF AUTOMATICITY IN READING

Table 2. English 1st and 2nd grade mean response times (in seconds, and SD) for reading and naming with and without interference. The standard Stroop measure is the interference effect in the naming task. The adjusted Stroop measure is the difference between the interference effect in the naming and the reading tasks. Decoding ability is reported in the final column (with SD and range in brackets).

Time point	Task	Without interference	With interference	Interference effect	Adjusted Stroop effect	Decoding ability
Grade 1						
Oct.	Reading	44.3 (16.5)	58.8 (32.0)	1.32** (0.53)		1.2 (2.1; 0-8)
	Naming	21.0 (7.0)	22.8 (5.0)	1.15* (0.29) ^a	-.17 (0.65)	
Feb.	Reading	26.7 (11.7)	36.4 (21.8)	1.32*** (0.33)		3.7 (3.6; 0-14)
	Naming	20.1 (5.9)	24.5 (5.7)	1.30** (0.42) ^a	-.03 (0.53)	
May	Reading	18.2 (8.0)	25.3 (11.9)	1.41*** (0.47)		7.7 (5.9; 0-24)
	Naming	16.5 (3.8)	28.2 (11.1)	1.71*** (0.52) ^a	.30* (0.67)	
Grade 2						
Oct.	Reading	40.9 (26.0)	50.1 (35.6)	1.26** (0.51)		7.4 (7.2; 0-34)
	Naming	19.7 (5.7)	24.8 (8.9)	1.34** (0.57) ^a	0.08 (0.83)	
Feb.	Reading	22.9 (17.3)	29.5 (20.0)	1.33*** (0.48)		10.9 (9.1; 0-36)
	Naming	17.5 (4.6)	28.2 (8.2)	1.68*** (0.54) ^a	0.35* (0.90)	
May	Reading	18.4 (14.0)	23.0 (17.1)	1.28*** (0.42)		14.8 (9.7; 0-34)
	Naming	16.8 (3.9)	28.6 (8.4)	1.76*** (0.59) ^a	0.48*** (0.72)	

^a The classic Stroop effect, * p < .05, ** p < .01, *** p < .001 (one sample t-test)

Even though the English first graders could hardly decode words, as indicated by their very low pseudo-word reading score, they still displayed a classic Stroop effect from the beginning of the grade. Conversely, when (lack of) inhibition was controlled, the (adjusted) Stroop effect did

not show up till the end of the first grade, and it was not even present at the beginning of the second grade (in a different sample). The apparent dip in the Stroop effect between May 1st grade and October 2nd grade was not significant and should be seen in the light of the independent samples from the two grade levels.

The inhibition effect, as measured by the interference effect in the reading conditions, was constant across languages, grade levels, and time points. The effect hovered around 1.3, and no significant means differences were found.

A univariate ANOVA on the adjusted Stroop effects across all samples indicated significant developmental trends within grade levels as evidenced by a strong effect of time (test point) ($F(1,271) = 23.6, p < .001$). A separate analysis of the English samples also showed a significant effect of grade level ($F(1,176) = 5.7, p < .02$). Further, a significant effect of language was found ($F(1,271) = 17.1, p < .001$). The French first graders displayed (adjusted) Stroop effects earlier than the English students. No interaction effects were present between language and test point ($F(2,271) = 1.2, n.s.$), indicating that the developmental trends were similar across languages.

After controlling for decoding ability (pseudo-word reading) no significant effects remained of test point, grade level, or language (all F-values were below 1). No significant correlation with age remained because age was largely a function of the aforementioned variables. This pattern of results indicates that the developmental trends and language effects on the adjusted Stroop effect were entirely linked to decoding development. Figure 2 displays the correlation between pseudo-word reading ability and the adjusted Stroop effect in the French and English children separately. The lack of language (orthography) effect is indicated by the near identity of the regression lines.

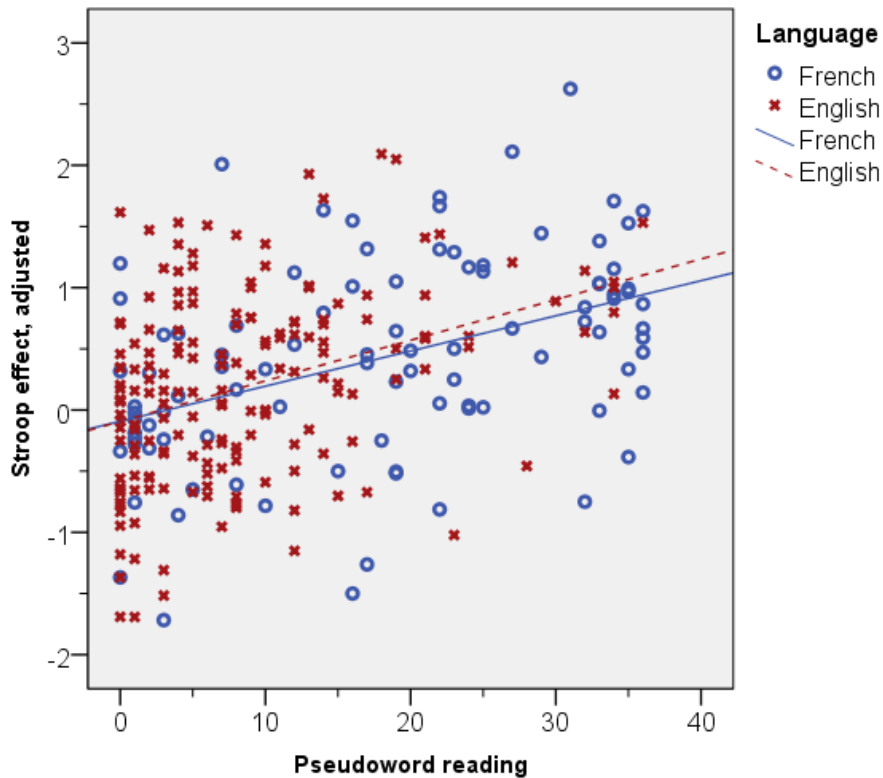


Figure 3. The (adjusted) Stroop effect develops as soon as word reading ability develops. It follows the same trajectory across language – in French- and English- speaking children.

Figure 3 shows that the onset of the adjusted Stroop effect is very close to the onset of decoding. This means that sample size will determine the time point at which Stroop effects will be significant. In the present samples, significant adjusted Stroop effects were detected in children who could correctly decode about 10 pseudo-words or more within one minute from the Seymour et al. (2003) lists.

Discussion

The general aim of the study was to investigate the emergence of automaticity of word reading. The study focused on the mandatory aspect of automaticity, that is, the degree to which written word identification is obligatory. When word identification is obligatory, words are identified

without conscious control– leaving mental resources for other aspects of reading such as text comprehension.

The results of the study indicate that automaticity develops early – typically during the course of the first year of reading instruction. This result was obtained even though the study employed an individual control for (lack of) inhibition, i.e., (lack of) suppression of alternative responses in the Stroop test. The study found that lack of inhibition may exaggerate the effects of mandatory processing in reading in 1st graders. However, significant indications of automaticity were found over and above what may have been caused by lack of inhibition. The finding of such early onset of automaticity is in line with some previous studies (e.g. Stanovich et al., 1981).

As expected, children learning to read in French developed decoding skill and automaticity more rapidly than (younger) children learning to read in English with its very irregular orthography. This difference may to some extent be related to the young age of the English 1st graders. It is impossible to know in the present study because of the confound between orthography and age at school entry. Yet, consistent effects of orthography on the initial development of word decoding have been found across a number of languages (e.g. Seymour et al., 2003).

However, when decoding development was controlled, automaticity was found to develop as fast in English as in French children. The earlier occurrence of automaticity in French first graders was thus fully explained by their earlier development of basic decoding abilities. In sum, the results support the view that automaticity develops in tandem with the development of word decoding.

The results are in line with what would be predicted by theories that link the development of word recognition automaticity to basic decoding skills. For example, in Ehri's model of sight word acquisition, the acquisition of the basic orthographic code (decoding ability) is a

foundation for the development of automaticity (e.g. Ehri, 2015). Letter-sound associations are the “glue” that connects orthographic representations to the phonological representations of whole words in the mental lexicon. Without this “glue”, sight word acquisition would only develop very slowly and would continue to be error prone. Similarly, according to Share’s self-teaching hypothesis (1999; 2008), acquisition of orthographic representations on a large scale requires that the reader has acquired the basic orthographic code. Phonological recoding (print-to-sound conversion) is the underpinning of an orthographic self-teaching mechanism that enables young readers to memorize orthographic representations and link them to known phonological word forms. One of the features of self-teaching theory is early onset. Beginning reading is beginning self-teaching (Share & Stanovich, 1995). Finally, longitudinal data support the view that the acquisition of the basic orthographic code (as indexed by pseudo-word reading ability) precedes the development of fluency in reading (Juel et al., 2014).

As mentioned in the introduction, the very early onset of automaticity as measured in the Stroop test may also be explained by an automatic activation of single letter-sounds in the specific context of color words. As soon as there are letters in the stimuli, the corresponding sounds of these letters may generate interference because the range of possible words is very limited. However, this explanation is at odds with the fact that the English children had been exposed to letters and their sounds for 1-1½ years before the joint onset of decoding ability and the Stroop effect.

Conclusions based on the present findings should be tempered by a number of limitations. These limitations may also suggest some directions for future research.

First, the control measure of inhibition was taken from Stroop’s original experiment (1935). There are a number of uncertainties related to this measure and its use in the present study. It measured the distraction caused by print colors on reading color words. As mentioned in the introduction, this distraction was considered to be composed of two opposite effects: an

activation of the names of the print colors, and an opposing *inhibition* of the activated color names (because they interfered with the correct reading response). The measure was taken to be relatively pure measure of the individual inhibition effect because the degree of semantic activation was assumed to be relatively constant across participants. This assumption was supported by the relatively small individual variation in pure color naming.

It could also be argued that the control measure of inhibition is in itself a reading measure because the amount of distraction is likely to depend on the strength of decoding ability. Stronger reading skills will more easily overrule the distracting print colors. So children with better reading ability might be less prone to distraction. However, no developmental trends were found in the inhibition measure. While this absence of a developmental trend is encouraging on the one hand, it may also be worrying on the other, because one might expect that children became better at inhibition even in the course of only one year. Perhaps the relatively small age range and the fact that inhibition was not taught (while reading was taught intensively) may explain why inhibition did not increase significantly. Future research is needed to establish the validity of the present control measure of inhibition by comparison with other measures of inhibition.

Second, neither this nor many other studies of the Stroop measure allow for generalization to other, less frequent, and less imaginable words. Stroop-like effects have been obtained with pictures across which distracting words have been written (Guttentag & Haith, 1980). However, such picture Stroop tasks have still been limited to frequent and concrete words. Neither is it clear that the picture Stroop effects are as strong as the color Stroop effects.

Third, the systematic rotation of materials and conditions may support generalization at the expense of homogeneity of the effects within groups. In particular for beginning readers, Stroop effects may have been induced when a reading condition preceded the naming condition (traditional Stroop measure). In this case, reading the color words would have been practiced

before the naming condition in which they might have interfered. Conversely, smaller Stroop effects may have been associated with the opposite order of conditions. For better readers, the order of presentation would have been less important because the better readers would have no difficulties reading the color words. So, if this source of error were important, more individual variability in the standard Stroop effect would have been present in the younger readers than in the older ones. However, there is no support for this prediction in the results. If anything, the standard deviation of the means for the traditional Stroop measure decreased with age. Nevertheless, the question remains as to whether there are better ways to assess inhibition than the one employed in the present study.

Fourth, the study attempted to assess the impact of orthographic depth on the development of decoding automaticity by comparing native French and native English readers. This comparison is, of course, prone to multiple confounds. Not only did the orthographies differ between groups, but so did language, age at school entry, instructional methods, and a great many other factors. The overlap between factors means that statistical control for one factor (such as age) would control for all the others at the same time. Future studies could control for these factors separately, e.g. by studying truly bilingual children taught to read in two orthographies. However, a more simple way ahead could be to study the development of automaticity in the longer term.

To conclude, the present study provides evidence that automaticity of word recognition – in the sense that word recognition is obligatory – develops in tandem with the development of word decoding. When the classic Stroop effect was adjusted for children's (lack of) inhibition, automaticity was found to be a significant factor in French children after only five months of reading instruction. In English first grade children, automaticity did not emerge before the end of the first grade, and it was not significant even at the beginning of second grade. Yet, the same developmental trajectories were found across orthographies when basic decoding ability was

taken into account. It remains to be seen whether an increased inhibition explains the common paradoxical finding of a *decreased* automaticity (Stroop) effect in Grade 3 and beyond.

Note

1. Alternatively, the adjusted Stroop effect was calculated as the quotient between the naming interference and the reading interference effects. However, the pattern of results was the same as the one presented below.

Acknowledgements

This work was supported by the Program of Visiting Professor of the University Paris 13 Sorbonne Paris Cité and by the Institut National de France (IUF).

We are grateful to Alexandra, Dilek, Molly and Susan for their assistance with data collection.

We are truly indebted to the participants and their teachers for their voluntary and solid efforts and to the inspectors and the directors of Paris (Baudricourt A) and Brighton for their cooperation.

References

- Adleman, N.E., Menon, V., Blasey, C.M., White, C., Warsofsky, I. S., Glover, G. H., & Reiss, A. L. (2002). A developmental fMRI study of the Stroop Color-Word task. *NeuroImage*, 16, 61–75. doi:10.1006/nimg.2001.1046
- Aro, M., & Wimmer, H. (2003). Learning to read: English in comparison to six more regular orthographies. *Applied Psycholinguistics*, 24(4), 621-635. doi: 10.1017/S0142716403000316
- Brocki, K.C., & Bohlin, G. (2004). Executive functions in children aged 6 to 13: A dimensional and developmental study. *Developmental Neuropsychology*, 28, 571-593. doi: 10.3922/j.psns.2013.1.12
- Bonino, S. & Ciairano, S. (1997). Effetto Stroop e capacita di inibizione: Una ricerca in campo evolutivo. *Giornale italiano di psicologia*, 24, 587-604. doi: 10.1421/161
- 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. doi: 10.1080/00221325.1962.10533572
- Dash, J., & Dash, A. S. (1982). Cognitive developmental studies of the Stroop phenomena: Cross-sectional and longitudinal data. *Indian Psychology*, 1, 24-33.
- Diamond, A., & Taylor, C. (1996). Development of an aspect of executive control: Development of the abilities to remember what I said and to “do as I say, not as I do.” *Developmental Psychobiology*. 29, 315–334. doi: 10.1002/(SICI)1098-2302(199605)29:4<315::AID-DEV2>3.0.CO;2-T
- Ehri, L.C. (2005). Learning to read words: Theory, findings, and issues. *Scientific Studies of reading*, 9, 167-188. doi: 10.1207/s1532799xssr0902_4
- Ehri, L. C. (2015). How children learn to read words. In A. Pollatsek & R. Treiman (Eds.), *The Oxford handbook of reading* (pp. 293-310). New York, NY: Oxford University Press.

- 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. doi: 10.1016/0022-0965(79)90055-9
- Ehri, L. C., & Wilce, L. S. (1983). Development of word identification speed in skilled and less skilled beginning readers. *Journal of Educational Psychology*, 75, 3-18. doi: 10.1037/0022-0663.75.1.3
- Ellis, N. C., Natsume, M., Stavropoulou, K., Hoxhallari, L., van Daal, V. H.P., Polyzoe, N., Tsipa, M.-L. & Petalas, M. (2004), The effects of orthographic depth on learning to read alphabetic, syllabic, and logographic scripts. *Reading Research Quarterly*, 39, 438–468. doi:10.1598/RRQ.39.4.5
- Everatt, J., Warner, J., & Miles, T.R. (1997). The incidence of Stroop interference in dyslexia. *Dyslexia*, 3, 222-228. doi:10.1002/(SICI)1099-0909(199712)3:4<222::AID-DYS12>3.0.CO;2-P
- Fodor, J. A. (1983). *Modularity of Mind: An Essay on Faculty Psychology*. Cambridge, Mass.: MIT Press.
- Frith, U. (1986). A developmental framework for developmental dyslexia. *Annals of Dyslexia*, 36, 69–81. doi: 10.1007/BF02648022
- Frith, U., Wimmer, H. & Landerl, K. (1998). Differences in phonological recoding in German- and English-speaking children. *Scientific Studies of Reading*, 2, 31-54. doi: 10.1207/s1532799xssr0201_2
- Goswami, U., Ziegler, J. C., & Richardson, U. (2005). The effects of spelling consistency on phonological awareness: A comparison of English and German. *Journal of Experimental Child Psychology*, 92, 345-365. doi: 10.1016/j.jecp.2005.06.002
- Guttentag, R. & Haith, M. (1978). Automatic processing as a function of age and reading ability. *Child Development*, 49, 707-716. doi: 10.2307/1128239

- Guttentag, R. & Haith, M. (1980). A longitudinal study of word processing by first-grade children. *Journal of Educational Psychology*, 72, 701-705. doi: 10.1037/0022-0663.72.5.701.
- Ikeda, Y., Okuzumi, H., & Kokubun, M., (2011). Age-related trends of interference control in school-age children and young adults in the Stroop color-word test. *Psychological Reports*, 108, 577-84. doi: 10.2466/04.10.22.PR0.108.2.577-584
- Horn, C. C. & Manis, F. R. (1987). Development of automatic and speeded reading of printed words. *Journal of Experimental Child Psychology*, 44, 92-108. doi: 10.1016/0022-0965(87)90024-5
- Joshi, R. M., & Aaron, P. G. (Eds.). (2013). *Handbook of orthography and literacy*. Routledge.
- Juul, H., Poulsen, M., & Elbro, C. (2014). Separating speed from accuracy in beginning reading development. *Journal of Educational Psychology*, 106 (4), 1096-1106 doi: 10.1037/a0037100
- Kahneman, D. (2011). *Thinking, Fast and Slow*. New York, NY: Farrar, Straus & Giroux.
- Kuhn, M. R., Schwanenflugel, P. J., Meisinger, E. B., Levy, B. A., & Rasinski, T. V. (2010). Aligning theory and assessment of reading fluency: Automaticity, prosody, and definitions of fluency. *Reading Research Quarterly*, 45(2), 230-251. doi: 10.1598/RRQ.45.2.4
- Laberge, D., & Samuels, S. J. (1974). Towards a theory of automatic information processing in reading. *Cognitive Psychology*, 6(2), 293-323. doi: 10.1016/0010-0285(74)90015-2
- Logan, G. D. (1997). Automaticity and reading: Perspectives from the instance theory of automatization. *Reading & Writing Quarterly: Overcoming Learning Difficulties*, 13(2), 123-146.
- McDonald, J.A., Beauchamp, M.H., Crigan, J.A., & Anderson, P.J. (2014). Age-related differences in inhibitory control in early school years. *Child Neuropsychology: A Journal*

- on Normal and Abnormal Development in Childhood and Adolescence*, 20, 509-526. doi: 10.1080/09297049.2013.822060
- MacLeod, C. M. (1991). Half a century of research on the Stroop effect: An integrative review. *Psychological Bulletin*, 109, 163-203. doi: 10.1037/0033-2909.109.2.163
- Moors, A., & De Houwer, J. (2006). Automaticity: A Theoretical and Conceptual Analysis. *Psychological Bulletin*, 132(2), 297-326. doi: 10.1037/0033-2909.132.2.297
- New, B., Doré-Mazars, K., Cavézian, C., Pallier, C., & Barra, J. (2016). The letter height superiority illusion. *Psychonomic Bulletin & Review*, 23(1), 291-298. doi: 10.3758/s13423-014-0753-8
- Peru, A., Faccioli, C., & Tassinari, G. (2006). Stroop effects from 3 to 10 years: The critical role of reading acquisition. *Archives Italiennes De Biologie*, 144(1), 45-62.
- Protopapas, A., Archonti, A., & Skaloumbakas, C. (2007). Reading ability is negatively related to Stroop interference. *Cognitive Psychology*, 54(3), 251-282. doi: 10.1016/j.cogpsych.2006.07.003
- Protopapas, A., Vlahou, E. L., Moirou, D., & Ziaka, L. (2014). Word reading practice reduces Stroop interference in children. *Acta Psychologica*, 148, 204-208. doi: 10.1016/j.actpsy.2014.02.006.
- Schadler, M., & Thissen, D.M. (1981). The development of automatic word recognition and reading skill. *Memory and Cognition*, 9, 132-141. doi: 10.3758/BF03202327
- Schiller, P. H. (1966). Developmental study of color-word interference. *Journal of Experimental Psychology*, 72, 105-108. doi: 10.1037/h0023358
- Seymour, P. H. K., Aro, M., & Erskine, J. M. (2003) Foundation literacy acquisition in European orthographies. *British Journal of Psychology*, 94, 143-174. doi: 10.1348/000712603321661859

- Share, D. L. (1999). Phonological Recoding and Orthographic Learning: A Direct Test of the Self-Teaching Hypothesis. *Journal of Experimental Child Psychology, 72*, 95-129. doi: 10.1006/jecp.1998.2481
- Share, D. L. (2008). On the Anglocentrism of current reading research and practice: The perils of overreliance on an “outlier orthography.” *Psychological Bulletin, 134*, 584–615. doi: 10.1037/0033-2909.134.4.584
- Share, D. L., & Stanovich, K. E. (1995). Cognitive processes in early reading development: Accommodating individual differences into a model of acquisition. *Issues in Education: Contributions from Educational Psychology, 1*, 1-57.
- Stanovich, K. E. (1990). Concepts in developmental theories of reading skill: Cognitive resources, automaticity, and modularity. *Developmental Review, 10(1)*, 72-100. doi: 10.1016/0273-2297(90)90005-O
- 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.
- Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology, 18*, 643-662. doi: 10.1037/h0054651
- van Daal, V. H. P. & Wass, M. (e-pub ahead of print, 29 Nov 2016). First- and Second-Language Learnability Explained by Orthographic Depth and Orthographic Learning: A “Natural” Scandinavian Experiment. *Scientific Studies of Reading*. doi: 10.1080/10888438.2016.1251437
- West, R. F., & Stanovich, K. E. (1978). Automatic contextual facilitation in readers of three ages. *Child Development, 49*, 717-727. doi: 10.2307/1128240
- West, R. F., & Stanovich, K. E. (1979). The development of automatic word recognition skills. *Journal of Reading Behavior, 11*, 211-219.

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