Child Neuropsychology, 16: 461–477, 2010
http://www.psypress.com/childneuropsych
ISSN: 0929-7049 print / 1744-4136 online
DOI: 10.1080/09297041003689780

Psychology Press
Taylor & Francis Group

THE NUMERICAL STROOP EFFECT IN PRIMARY SCHOOL CHILDREN: A COMPARISON OF LOW, NORMAL, AND HIGH ACHIEVERS

Angela Heine,[1] Sascha Tamm,[1] Bert De Smedt,[2] Michael Schneider,[3] Verena Thaler,[1] Joke Torbeyns,[4] Elsbeth Stern,[3] Lieven Verschaffel,[4] and Arthur Jacobs[1]

[1]Freie Universität Berlin, Dept. of Psychology, Berlin, Germany, [2]K. U. Leuven, Centre for Parenting, Child Welfare and Disabilities, Leuven, Belgium, [3]ETH Zurich, Zurich, Switzerland, and [4]K. U. Leuven, Centre for Instructional Psychology and Technology, Leuven, Belgium

Sixty-six primary school children were selected, of which 21 scored low on a standardized math achievement test, 23 were normal, and 22 high achievers. In a numerical Stroop experiment, children were asked to make numerical and physical size comparisons on digit pairs. The effects of congruity and numerical distance were determined. All children exhibited congruity and distance effects in the numerical comparison. In the physical comparison, children of all performance groups showed Stroop effects when the numerical distance between the digits was large but failed to show them when the distance was small. Numerical distance effects depended on the congruity condition, with a typical effect of distance in the congruent, and a reversed distance effect in the incongruent condition. Our results are hard to reconcile with theories that suggest that deficits in the automaticity of numerical processing can be related to differential math achievement levels. Immaturity in the precision of mappings between numbers and their numerical magnitudes might be better suited to explain the Stroop effects in children. However, as the results for the high achievers demonstrate, in addition to numerical processing capacity per se, domain-general functions might play a crucial role in Stroop performance, too.

Keywords: ***Numerical Stroop paradigm; Mathematical disabilities; Automatized numerical magnitude processing; Size congruity effect; Numerical distance effect; Reverse distance effect.***

INTRODUCTION

In the numerical Stroop paradigm, participants are asked to make comparative judgments about pairs of numbers that vary with respect to two independent stimulus dimensions

This work was supported by a grant to A. Jacobs and E. Stern from the German Federal Ministry of Education and Research (BMBF), funded under the interdisciplinary research initiative “NIL Neuroscience, Instruction, Learning: A Program for the Promotion of Scientific Collaboration between the Neurosciences and Research on Learning and Instruction.” Additional support was provided by the GOA grant 2006/1 from the Research Fund of the Katholieke Universiteit Leuven, Belgium. Bert De Smedt is a postdoctoral fellow of the Research Foundation Flanders (FWO), Belgium.

We thank Nadja Rosental for proofreading the manuscript.

Address correspondence to Angela Heine, Freie Universität Berlin, Dept. of Psychology, Habelschwerdter Allee 45, Berlin, 14195 Germany. E-mail: aheine@zedat.fu-berlin.de

—i.e., the numerical magnitude and the physical size of the digits, of which one is task relevant and the other is not. The often-replicated finding that in Stroop contexts the congruence of relevant and irrelevant stimulus features affects participants' task performance is generally taken to be evidence of an automatic processing of the respective task-irrelevant stimulus dimension (Stroop, 1935). In this vein, the so-called *numerical Stroop effect* or *size congruity effect* in the numerical Stroop task refers to shorter reaction times for congruent trials —i.e., trials where numerical and physical information is consistent, compared to incongruent trials.

While research on numerical Stroop effects initially focused on the impact of physical size information on numerical comparisons (Besner & Coltheart, 1979), Henik and Tzelgov (1982) were among the first to investigate the influence of numerical information on subjects' performance in the physical comparison task. In their study, participants compared the physical size of digits while ignoring their numerical magnitudes. The finding that subjects' performance varied systematically with congruity between the two dimensions in the physical comparison task led the authors to conclude that in competent adults, numerical magnitude information is accessed and processed automatically, whether it is of relevance for task execution or not. A later study by the authors showed that this effect is also present when no direct comparison between items is possible and subjects have to revert to memorized standards (Tzelgov, Meyer, & Henik, 1992).

On the neurocognitive level, results from neuroimaging studies suggest that in judgments of continuous quantities such as numerical and physical magnitude, distributed but potentially overlapping neuronal structures in the parietal cortex are activated (Cohen Kadosh, Cohen Kadosh, Linden, et al., 2007; Kaufmann et al., 2005; Pinel, Piazza, LeBihan, & Dehaene 2004). According to Walsh's ATOM-model (*A Theory of Magnitude*, 2003), the posterior parietal cortex might be related to a number of different cognitive functions such as the processing of time, quantity, and space, which all have in common that they draw on magnitude representations in a more general sense. This cross-domain representational overlap might be one reason for the interaction of numerical and physical size information that can be observed in numerical Stroop tasks. However, apart from an interaction at the level of perceptual and conceptual stimulus evaluation, three recent studies demonstrate that the resolution of conflict between the two dimensions of numerical and physical size might not be limited to the early processing stages (Cohen Kadosh, Cohen Kadosh, Linden, et al., 2007; Szücs & Soltész, 2007, 2008). Using the high temporal resolution of EEG parameters, these studies showed that effects of congruity might actually be observable as late as during the stage of response execution.

In addition to studies on the numerical Stroop effect in general, two more specialized lines of research have emerged during the past few years, both yielding somewhat inconclusive outcomes. The first line of research addresses the question of the point in individual development when children start to show effects of congruity in numerical and physical size comparison tasks. Two of those studies showed that while even beginning first graders show congruity effects in the numerical comparison task, in physical size comparisons the effects of congruity between relevant and irrelevant stimulus dimension are not observable until later in children's math development (Girelli, Lucangeli, & Butterworth, 2000; Rubinsten, Henik, Berger, & Shahar-Shalev, 2002). The results of these studies suggest that during the first few years of schooling, the automatization of numerical magnitude processing develops gradually and should become fully manifest in the form of congruity effects in the physical comparison task when children reach Grade 3. However, this conclusion was challenged by a study on fourth graders that did not find a congruity effect in the physical size comparison task for that age group (Landerl, Bevan, & Butterworth, 2004). To further complicate the issue, two recent studies that used neurophysiological measures to investigate effects of congruence in numerical Stroop tasks demonstrate that even when children's behavioral Stroop effects resemble mature performance patterns, on the level of cerebral activation, the differences between children and adults might be substantial (Kaufmann et al., 2006; Szücs, Soltész, Jármi, & Csépe, 2007). Apart from the ability to process numerical meaning automatically, differences in executive functioning appear to be another factor that determines specific patterns of behavioral Stroop effects in different populations. Overall, it can be concluded that while the automatic activation of numerical information might have precursors that are present as early as 3 years of age (Rousselle & Noël, 2008), the developmental pathway that ultimately leads to adult-like processing patterns in numerical Stroop tasks appears to be influenced by more than one underlying cognitive function, that each has its own specific developmental course, which might not be completed much earlier than age 10.

A second line of research concerns the question of possible deficits in the automatic processing of numerical information in individuals suffering from neuropsychological impairments such as attentional deficits (e.g., Kaufmann & Nuerk, 2006) or deficits in the domain of mathematics (e.g., Rubinsten & Henik, 2005). In the first experimental study that investigated the numerical Stroop effect in adults with developmental math disabilities (MD), Rubinsten and Henik demonstrated a considerably reduced size congruity effect in two physical comparison tasks in the MD group compared to normal achievers, while no differences were found between the groups for the numerical comparison task. Furthermore, the congruity effect found for the physical comparison tasks consisted of the interference component only (i.e., the relative disadvantage of incongruent compared to neutral stimuli), while the facilitatory component—i.e., the advantage of congruent compared to neutral trials [MacLeod, 1991], was missing. In a more recent study, Cohen Kadosh, Cohen Kadosh, Schumann, et al. (2007) further corroborated these results. The authors were able to mimic these specific patterns of performance—i.e., overall reduced congruity effects in the physical comparison task with no advantage for congruent over neutral stimuli—by disrupting right intraparietal regions in originally well-performing adults using transcranial magnetic stimulation.

Combining these two lines of research—i.e., the developmental and the neuropsychological perspective on the numerical Stroop effect—a small number of studies addressed the question of how children with mathematical disabilities perform in the numerical and physical Stroop tasks compared to their normally developing peers. Koontz and Berch (1996) showed that while 10-year-old, normally developing children exhibited effects of interference of numerical information in a physical identity judgment task consisting of stimuli in the subitizing range, children with MD did not show such an impact of the task-irrelevant numerical information. However, these findings are contradicted by results reported by two other studies that investigated Stroop performance in learning disabled children. Comparing groups of children with and without mathematical disabilities, Landerl et al. (2004) reported that while in the numerical comparison task a congruity effect was observed in all of the fourth graders involved in their study, in the physical comparison task, size congruity effects were not found in either group. In contrast to these results, Rousselle and Noël (2007) demonstrated an influence of numerical information in size comparisons in groups of normal and of learning disabled second graders in an experimental setting where the display of physical size information lagged behind the numerical information. However, the considerable differences between experimental approaches make it difficult to compare the results of these three studies.

Overall, the somewhat inconsistent results of the available developmental and neuropsychological studies into the numerical Stroop effect call for further research into performance patterns of children with and without deficits in the mathematical domain. The question whether different math achievement levels in children are related to differences in the automatization of numerical magnitude processing is yet to be answered conclusively. This issue becomes even more relevant in view of the results of recent neurophysiological studies into the numerical Stroop effect. Szücs and Soltész (2007; see also Szücs et al., 2007) point out that caution is needed in interpreting size congruity effects in children as a measure of automatic processing of numerical magnitude information because immature processes of response organization and execution in this population might also be of relevance.

What further complicates the discussion on the relationship between children's math achievement levels and their performance under numerical Stroop conditions is that Tang, Critchley, Glaser, Dolan, and Butterworth (2006), in line with Szücs and colleagues, suggest that in numerical Stroop tasks not so much the *size congruity effect* but rather the *numerical distance effect* might actually "prove to be a more sensitive measure" (p. 2060) for assessing automatized magnitude processing. Originally, the numerical distance effect refers to the fact that performance measures in numerical comparison tasks vary systematically with the numerical distance between the numbers to be compared (Moyer & Landauer, 1967); i.e., the smaller the distance between two numbers, the harder it is to decide which is smaller or larger. These effects of numerical distance are generally interpreted as evidence for an ordered numerical magnitude representation in the sense of a mental number line (Dehaene & Cohen, 1995). The higher representational overlap between close numbers can be assumed to result in a higher level of difficulty for numerical comparisons.

In contrast to the typical distance effect, in the context of the physical comparison of the numerical Stroop paradigm a number of studies found a *reversed numerical distance effect* (Girelli et al., 2000; Henik & Tzelgov, 1982; Szücs & Soltész, 2007; Szücs et al., 2007; Tang et al., 2006; Tzelgov et al., 1992). It was shown that in the incongruent condition of the physical comparison, i.e., when numerical magnitude is the task-irrelevant dimension, larger numerical distances result in larger reaction times compared to pairs of numbers where the numerical distance is small. These findings seem to be caused by a stronger interference of numerical information for larger numerical distances compared to smaller ones (Szücs et al., 2007). Following Henik and Tzelgov (1982) and Tzelgov et al. (1992), Tang and colleagues (2006) emphasize the importance of looking for a reversed distance effect in Stroop tasks by pointing out that finding such an effect "would indicate that exact numerical values had been computed despite task irrelevance" (p. 2052). For the question of whether differential levels of performance in the domain of mathematics are related to differences in the development of automatic numerical magnitude processing, this means that apart from effects of congruity per se, the finding of a reversed numerical distance effect in children diagnosed with MD would be positive evidence for automatized magnitude processing in this group of children.

By exploring the interaction of size congruity and numerical distance effects under Stroop conditions in three groups of primary school children with different math achievement levels—i.e. low, normal, and high achievers—the present study aims to provide further insights into the development of automatized numerical processing in these populations. In line with studies that suggest that focusing on the interaction of both effects might be a promising approach instead of solely looking for size congruity effects

(Szücs & Soltész, 2007; Szücs et al., 2007; Tang et al., 2006), we implemented a Stroop task that allows for separate in-depths analyses of the respective influences of congruity and numerical distance on children's Stroop performance.

In view of the findings of studies that implemented more than one numerical distance (e.g., Girelli et al., 2000, Henik & Tzelgov, 1982), we expect that the occurrence of size congruity effects in the physical comparison task depends on the numerical distance between the two digits to be compared; i.e., the congruity effect can be assumed to be generally more pronounced when the distance between the two digits is large. An occurrence of both congruity effects and effects of numerical distance that in turn are dependent on congruity conditions would be indicative of automaticity in the processing of numerical magnitude information in our groups of children. However, not only automaticity of processing but also differential executive control levels might have an influence on Stroop processing. An impact of the latter can be assumed to come into effect mainly when disadvantageous influences of task irrelevant information have to be inhibited, i.e., in the incongruent trials.

METHOD

Participants

Sixty-six second and third graders out of a pool of 1,242 children from six public primary schools in Berlin took part in the present study. Participants were selected on the basis of their achievement levels in several diagnostic tests. Only children whose parents had given their written informed consent participated in the study.

In a first step, all children from the original pool were administered the *Heidelberger Rechentest 1-4* (HRT 1-4; Haffner, Baro, Parzer, & Resch, 2005). The HRT 1-4 is a standardized math achievement test for Grades 1 to 4 and comprises two subscales, such as an arithmetic scale that includes timed tests of arithmetic skills, e.g., addition, subtraction, multiplication, and division, and a second scale that consists of tests of children's visuospatial abilities, such as 2D length estimation, estimation of set size.

After the initial testing with the HRT 1-4, a group of 382 children entered the second diagnostic phase. Children's general cognitive performance levels were determined using a standardized intelligence test for primary school children (*Kognitiver Fähigkeitstest* [KFT 1-3]; Heller & Geisler, 1983). An assessment for possible attentional problems was carried out using a standardized diagnostic tool (*Children's Color Trails Test* [CCTT]; Llorente, Williams, Satz, & D'Elia, 2003) and a short questionnaire for teachers (cf. Conners, 1973). Additionally, low achievers' performance in reading and spelling was measured by standardized tests, i.e., the SLS 1-4 (*Salzburger Lese-Screening für die klassenstufen 1-4*; Mayringer & Wimmer, 2003) and the spelling subtest of the SLRT (*Salzburger Leseund Rechtschreibtest*; Landerl, Wimmer, & Moser, 1997). After exclusion of children with IQs more than 1 *SD* below the standard mean, children with either reading and/or spelling problems (i.e., scores 1.65 *SD*s below the standard mean for the SLS and SLRT) and children suspected to suffer from attentional deficits according to CCTT results and/or teachers' assessments, a group of 66 children was selected for the present study on the basis of their scores on the arithmetic subscale of the math achievement test. Of the children that were chosen to take part in the experimental study, 21 scored low (≤ 1.65 *SD* below the standard mean) on the arithmetic subscale of the HRT 1-4 (low achievers [LA]), 23 children reached normal scores (normal achievers, NA), and 22 children completed the subtests with exceptionally high scores (≥ 2 *SD* above the standard mean; high achievers [HA]).

Finally, in order to test for differences in domain-general processing speed (Catts, Gillespie, Leonard, Kail, & Miller, 2002), children had to perform three rapid automatized naming tasks (RAN) that required them to rapidly name either objects—i.e., one-syllable and three-syllable animals—or single digits. Naming times and errors were recorded.

Materials and Procedure

Pairs of one-digit numbers were presented to the children on a computer screen with one digit displayed on the left and the other on the right side of the screen (distance: 3 cm). The digits of each pair varied with respect to two dimensions, i.e., numerical size and physical size. Children were instructed to select the numerically larger digit in one task (numerical comparison) and the physically larger digit in the other (size comparison). The order of presentation of the two subtasks was counterbalanced for each group of children in order to avoid task-order effects.

The stimulus sets were composed of the numbers 1 to 9, with number 5 excluded. The numerically different digit pairs were separated by the numerical distances of either 1 or 5. Each digit was allowed to occur in the same position (left/right) only twice, which resulted in 16 different digit pairs (i.e., 1-2; 2-1; 3-4; 4-3; 6-7; 7-6; 8-9; 9-8; 1-6; 6-1; 2-7; 7-2; 3-8; 8-3; 4-9; and 9-4). The physical size of the digits varied between either 1.2 cm (font: 48 pt) or 0.6 cm (font: 24 pt) for the physically different digit pairs. Combining each digit pair with each of the two size conditions—i.e., small digit on the left, large digit on the right and vice versa—yielded 16 congruent (the numerical larger digit was also physically larger) and 16 incongruent stimuli (the numerical larger digit was the physically smaller one) for both comparison tasks. In the numerical comparison, neutral stimuli consisted of the 16 digit pairs that were displayed in the same physical size (0.9 cm, font: 36 pt). In the size comparison, tied pairs were included (i.e., 1-1; 2-2; 3-3; 4-4; 6-6; 7-7; 8-8; and 9-9) and combined with the two size conditions, which added up to 16 neutral digit pairs. Each subtask was thus composed of 48 different stimuli. Each stimulus was displayed two times, which resulted in 96 trials per task.

Digit pairs were presented pseudorandomized in order to avoid habituation effects in the children's responses. That meant that for each task (a) the same digit did not occur in the same position in consecutive trials, (b) the correct response side was changed at least every other trial, (c) no more than two stimuli of the same congruity condition appeared consecutively, and (d) the numerical splits between the displayed digits were the same in no more than two consecutive trials.

Children were seated in front of a 17" monitor (resolution: 1024 x 768, distance from the screen: 60 cm). Responses were given using a button box with only two buttons. Children were instructed to press the button on the same side where the numerically larger digit appeared in the numerical comparison or on the side where the physically larger digit appeared in the size comparison. Children were asked to respond as fast as possible without sacrificing accuracy.

Each digit pair was preceded by a fixation cross in the middle of the screen for 300 ms and a blank screen for 500 ms. After each trial, a blank screen was displayed for 2000 ms before the next fixation cross was shown.

Children were familiarized with the tasks during a practice period before each part of the experiment. In these practice trials, children were asked to comment on their responses to make sure they understood the tasks correctly. After 10 practice trials, the main experiments started. Response times and button clicks were recorded automatically for each trial.

Data Analysis

In a first step, Tukey's (1977) fence method was used to remove outliers. The lower Tukey fence is the first quartile minus 1.5 times the interquartile range, while the upper Tukey fence is the third quartile plus 1.5 times the interquartile range, i.e., Q1-1.5 (Q3–Q1) and Q3+1.5 (Q3–Q1). Values outside this range were considered to be outliers. For the Stroop data, trials with RTs that exceeded Tukey's criterion were excluded from the analyses. After the elimination of outliers, an average of 99.4% and 99.8% of the trials entered the analyses for the numerical comparison and for the size comparison, respectively.

For each participant, average reaction times (RT) were calculated on the basis of correct trials only. Mean RTs and error rates were determined for each participant per task (numerical comparison, size comparison), congruity (congruent, neutral, incongruent), and numerical distance (1 and 5). In a first step, mean error rates and RTs were subjected to repeated-measures analysis of covariance (ANCOVAs) with task, congruity, and numerical distance as within-subjects factors, and group (LA, NA, HA) as between-subjects factors. Intelligence was entered as a covariate after mean centering the KFT scores (Delaney & Maxwell, 1981). A number of follow-up analyses were carried out in order to allow for further insights into the specific impacts of the factors of task, congruity, and distance on our performance parameters. Across all levels of analysis, Bonferroni-corrected post hoc tests were used to test for differences between groups and factor levels. Greenhouse-Geisser corrections were applied to the degrees of freedom where necessary.

RESULTS

Diagnostic Data

Table 1 provides a detailed overview of the diagnostic data. Most importantly, the groups of low, normal, and high achievers differed with respect to their results on the math achievement test; i.e., the mean composite scores for the arithmetic subscale of the HRT 1-4 differed significantly between the groups, $F(2, 63) = 479.93$, $p < .001$. Post hoc tests yielded highly significant pairwise comparisons (ps < .001).

Furthermore, the groups of children obtained significantly different scores on the intelligence test, $F(2, 63) = 5.86$, $p = .005$. Bonferroni-corrected post hoc tests revealed this effect to be due to significant differences between NA and HA ($p = .003$), while both other comparisons were not significant (ps > .230). A higher mean IQ score for the group of high achievers could not be avoided because exceptional scores in the math achievement test were invariably related with higher scores in the intelligence test in our group of 382 children.

Apart from the observed IQ variations, the groups also differed with respect to working memory functions that were assessed using a standardized test battery (*Working Memory Test Battery for Children* [WMTB-C]; Pickering & Gathercole, 2001). The subtests of the WMTB-C load on three different subscales: i.e., phonological loop functions, $F(2, 63) = 10.64$, $p < .001$; visuospatial sketchpad functions, $F(2, 63) = 8.63$, $p < .001$; and executive functions, $F(2, 63) = 3.61$, $p = .033$. Post hoc comparisons showed that while the groups of NA and LA did not differ significantly with respect to any of the three WMTB-C subscales (ps > .101), the performance differences between HA and both other groups were significant (ps < .050) with the exception of the visuospatial sketchpad scores of HA and NA, which did not differ significantly ($p > .099$).

Table 1 Subject Details and Mean Performance for the Diagnostic Tests.

	Achievement Group		
	LA	NA	HA
n	21	23	22
Gender (m / f)	9 / 12	10 / 13	14 / 8
Grade (2nd / 3rd)	11 / 10	12 / 11	10 / 12
Age (years)	8.0 (0.9)	8.1 (0.8)	8.0 (0.9)
Intelligence*	55.3 (9.1)	51.5 (9.0)	59.8 (5.8)
*Mathematics**			
Arithmetic operations	32.00 (4.47)	51.57 (4.82)	75.09 (4.39)
Visuospatial abilities	39.05 (8.91)	50.96 (7.52)	64.36 (9.37)
*Working Memory***			
Central executive	87.86 (12.54)	95.87(10.48)	106.50 (19.74)
Phonological loop	88.48 (8.54)	93.61 (9.94)	104.05 (15.41)
Visuospatial sketchpad	83.05 (10.70)	92.61(11.02)	94.09 (20.08)
RAN ***			
1-digit numbers	23.67 (8.42)	23.68 (6.76)	23.40 (3.73)
1-syllable animals	60.92 (13.90)	59.16(12.30)	51.63 (7.86)
3-syllable animals	69.04 (18.58)	70.76(20.94)	57.45 (12.12)

Note. Standard deviations are shown in parentheses.
LA, low achievers; NA, normal achievers; HA, high achievers.
*Standard score: Mean = 50, *SD* = 10. **Standard score: Mean = 100, *SD* = 15.
***Mean reaction times in seconds.

A repeated-measures analysis of variance (ANOVA) on naming times with RAN condition as within- subjects factor and group as between-subjects factor revealed significant main effects of condition, $F(2, 126) = 283.62$, $p < .001$, and group, $F(2, 63) = 4.17$, $p = .020$, but no significant Task × Group interaction, $F(4, 126) = 2.59$, $p = .088$. While the post hoc analyses yielded no significant naming time differences between NA and LA across all three RAN conditions ($p > .999$), HA were significantly faster than both other groups in both object naming tasks (*p*s > .43), whereas their naming times were not different from those of the two other groups in the digit-naming condition (*p*s > .999).

Analyses of Overall Experimental Task Performance

In a first step, repeated-measures ANCOVAs with task (numerical comparison, size comparison), condition (congruent, incongruent), and numerical distance (small, large) as within-subject factors, group as between-subject factor, and intelligence as covariate were conducted on error rates and reaction times. Table 2 lists mean overall reaction times and error rates for the different levels of the factors task, condition, and distance.

For the errors rates, the ANCOVA yielded highly significant main effects of task, $F(1, 62) = 71.67$, $p < .001$, congruity, $F(1, 62) = 78.26$, $p < .001$, and distance, $F(1, 62) = 35.54$, $p < .001$, but no main effect of group, $F(2, 62) = 0.38$, $p = .687$. Of the two-way interactions, Task × Congruity, $F(1, 62) = 96.15$, $p < .001$, Task × Distance, $F(1, 62) = 53.11$, $p < .001$, and Congruity × Distance, $F(1, 62) = 35.86$, $p < .001$, were significant. However, no two-way interaction with the factor group reached significance (*p*s > .305). The three-way interactions of Task × Congruity × Distance, $F(1, 62) = 50.14$, $p < .001$,

Table 2 Mean Overall Reaction Times and Error Rates for the Levels of the Factors Task, Congruity, and Distance.

	RTs in ms	% Errors
Task		
Numerical comparison	945 (215)	8.9 (7.2)
Physical size comparison	598 (131)	1.8 (2.4)
Congruity		
Incongruent	823 (172)	9.0 (6.8)
Congruent	719 (145)	1.8 (3.0)
Numerical distance		
Distance 1	787 (160)	6.7 (4.8)
Distance 5	755 (155)	4.1 (4.3)

Note. Standard deviations are shown in parentheses.

and of Congruity × Distance × Group, $F(1, 62) = 3.25$, $p = .045$, were significant. The analyses yielded neither a main effect of intelligence, $F(1, 62) = 0.12$, $p = .736$, nor significant interaction effects between intelligence and any other factor (ps > .147). A Bonferroni-adjusted post hoc test revealed no differences between the groups of children for this overall analysis (ps = 1.000).

The repeated-measures ANCOVA on reaction times revealed highly significant main effects of task, $F(1, 62) = 371.50$, $p < .001$, congruity, $F(1, 62) = 282.31$, $p < .001$, distance, $F(1, 62) = 63.80$, $p < .001$, and group, $F(2, 62) = 21.09$, $p < .001$. With the exception of Group × Distance, $F(2, 62) = 0.52$, $p = .595$, all other two-way interactions were significant, i.e., Task × Congruity, $F(1, 62) = 219.39$, $p < .001$, Task × Distance, $F(1, 62) = 78.40$, $p < .001$, Congruity × Distance, $F(1, 62) = 7.09$, $p < .010$, Task × Group, $F(2, 62) = 11.13$, $p < .001$, and Congruity × Group, $F(2, 62) = 5.77$, $p = .005$. Of the three-way interactions, Task × Congruity × Distance, $F(1, 62) = 29.38$, $p < .001$, and Task × Distance × Group, $F(1, 62) = 3.70$, $p = .030$, were significant. The interaction between congruity, distance, and group reached marginal significance, $F(1, 62) = 2.93$, $p = .061$. There was neither a main effect of intelligence, $F(1, 62) = 0.15$, $p = .697$, nor did the interactions between intelligence and the factors task, congruity, or distance reach significance (ps > .118). Bonferroni-corrected post hoc tests showed that over all in tasks and conditions LA were significantly slower that both other groups, while HA were significantly faster than LA and NA (ps < .017).

Children's Performance Data Separated by Task

For the numerical comparison task, a repeated-measures ANOVA on error rates with congruity and numerical distance as within-subjects factors and group as between-subjects factor yielded highly significant main effects of congruity, $F(1, 63) = 99.96$, $p < .001$, and distance, $F(1, 63) = 52.34$, $p < .001$, but no effect of group, $F(2, 63) = 0.47$, $p = .625$. No interaction effects were found (ps > .153). The analysis of the reaction times revealed highly significant main effects of congruity, $F(1, 63) = 312.47$, $p < .001$, distance, $F(1, 63) = 84.82$, $p < .001$, and group, $F(2, 63) = 25.14$, $p < .001$. Of the interactions only Congruity × Group was significant, $F(2, 63) = 4.32$, $p = .017$. Post hoc tests showed that LA were significantly slower than NA and HA (ps < .001), and HA were faster than both other groups (ps < .022).

For the size comparison task, the repeated-measures ANOVA on error rates yielded no effects of congruity, distance, group, or intelligence ($ps > .166$). Of the interactions, only Congruity × Distance was significant, $F(1, 63) = 8.06$, $p = .006$. The analysis of the reaction times showed highly significant main effects of congruity, $F(1, 63) = 13.23$, $p < .001$, and group, $F(2, 63) = 9.45$, $p = .001$. Of the interactions Congruity × Distance, $F(1, 63) = 39.47$, $p < .001$, and Distance × Group, $F(2, 63) = 3.87$, $p = .027$, were significant. Post hoc tests showed that while the reaction times of LA and NA did not differ ($p = .587$), HA were significantly faster than both other groups ($ps < .013$).

Effects of the Interaction of Congruity and Numerical Distance

Figure 1 illustrates the specifics of the interactions of task, numerical distance and congruency for children's reaction times. For the numerical comparison (Figure 1, left side), the graphs shows the typical steep drop of reaction times across the three congruency conditions for all groups, irrespective of numerical distance. For the physical comparison task, the picture changes (Figure 1, right side). Apart from a reversal of the influence of numerical distance for the incongruent compared to the congruent trials, a second effect can be gathered from the reaction time curves. While for the small numerical distance the classical effect of congruity is not evident in any of the three groups, the reaction times for trials where the numerical distance is five show the classical Stroop pattern, i.e., longer reaction times for incongruent compared to congruent trials.

In order to test the different congruity and distance effects, several two-way analyses of variance were carried out on reaction times and error rates. Table 3 lists the ANOVA results for all congruity (congruent, neutral, incongruent) and distance combinations of the numerical comparison task. Post hoc pairwise comparisons for the effects of congruity showed that for both numerical distances, reaction times were significantly shorter and error rates significantly lower for congruent than for neutral trials ($ps < .001$). The differences in reaction times and error rates between neutral and incongruent trials were significant too ($ps < .001$), with the only exception of the error rates for the large distance ($p = .539$). Post hoc tests for group differences in mean reaction times showed that for all congruity and distance conditions HA were significantly faster than NA, whereas

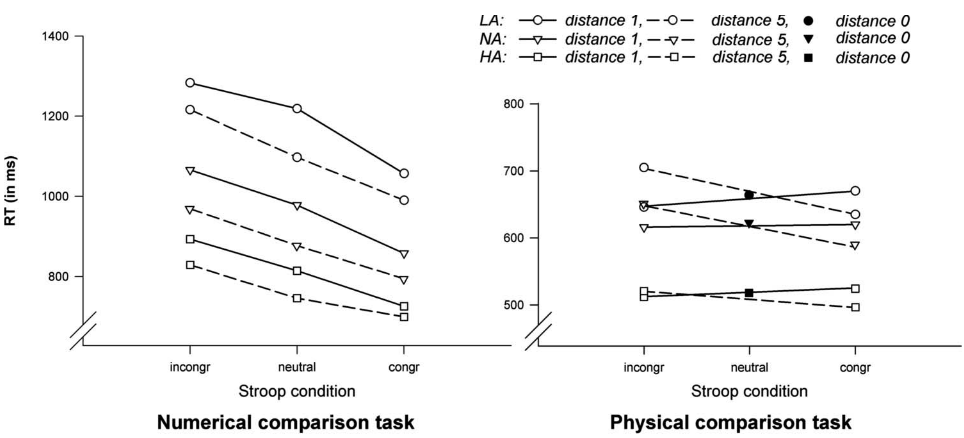

Figure 1 The congruity effect as a function of task and numerical distance for the three groups of children.

Table 3 Results of the Two-Way Repeated Measures ANOVAs for the Numerical Comparison Task.

	RTs			Errors		
Factor	*df* *	*F*	*p*	*df* *	*F*	*p*
Congruity effect for factor numerical distance = 1						
		(ε_{GG} = 0.93)			(ε_{GG} = 0.71)	
Congruity	(2, 126)	138.71	< .001	(2, 126)	95.73	< .001
Group	(2, 63)	26.41	< .001	(2, 63)	0.42	.657
Congruity × Group	(4, 126)	1.83	.128	(4, 126)	0.70	.595
Congruity effect for factor numerical distance = 5						
		(ε_{GG} = 0.82)			(ε_{GG} = 0.54)	
Congruity	(2, 126)	142.44	< .001	(2, 126)	43.49	< .001
Group	(2, 63)	25.21	< .001	(2, 63)	1.01	.406
Congruity × Group	(4, 126)	3.53	.009	(4, 126)	1.06	.353
Distance effect for factor congruity = incongruent						
Distance	(1, 63)	64.43	< .001	(1, 63)	57.83	< .001
Group	(2, 63)	25.14	< .001	(2, 63)	0.52	.599
Distance × Group	(2, 63)	1.27	.289	(2, 63)	1.71	.188
Distance effect for factor congruity = neutral						
Distance	(1, 63)	95.82	< .001	(1, 63)	24.34	< .001
Group	(2, 63)	26.77	< .001	(2, 63)	0.07	.930
Distance × Group	(2, 63)	2.46	.093	(2, 63)	0.34	.716
Distance effect for factor congruity = congruent						
Distance	(1, 63)	33.52	< .001	(1, 63)	0.02	.879
Group	(2, 63)	21.90	< .001	(2, 63)	0.04	.958
Distance × Group	(2, 63)	1.97	.147	(2, 63)	0.73	.485

*Greenhouse-Geisser adjusted where appropriate.

NA were significantly faster than LA (*p*s < .039). No group differences were found for the error rates (*p*s > .964).

The results of the two-way ANOVAs for all congruity (congruent, incongruent) and distance combinations on reaction times and error rates for the physical comparison task can be gathered from Table 4. All post hoc pairwise comparisons yielded significant reaction time differences between HA and both other groups (*p*s < .018), while the differences between LA and NA were not significant (*p*s > .413). Again, no group differences were found for the error rates (*p*s > .481).

Follow-up one-way ANOVAs were conducted for each group separately to further investigate the two significant interactions for the reaction times. While the overall effects of congruity for the large distance in the numerical task as well as all post hoc comparisons of the congruity conditions were highly significant in each of the groups (*p*s < .001), the groups showed differential susceptibility to the influence of numerical distance for the incongruent condition of the physical comparison. While the analyses yielded significant reversed distance effects for LA, $F(1, 20) = 17.33$, $p < .001$, and NA, $F(1, 22) = 10.45$, $p < .004$, no such effect was found for the high achievers, $F(1, 21) = 1.38$, $p = .253$.

DISCUSSION

The main objective of the present study was to determine whether primary school children of different math achievement levels show differences in measures of controlled

Table 4 Results of the Two-Way Repeated Measures ANOVAs for the Size Comparison Task.

	RTs			Errors		
Factor	*df*	*F*	*p*	*df*	*F*	*p*
Congruity effect for factor numerical distance = 1						
Congruity	(1, 63)	6.26	.015	(1, 63)	10.88	.002
Group	(2, 63)	8.13	.001	(2, 63)	1.01	.343
Congruity × Group	(2, 63)	1.19	.310	(2, 63)	0.83	.443
Congruity effect for factor numerical distance = 5						
Congruity	(1, 63)	33.21	< .001	(1, 63)	0.25	.618
Group	(2, 63)	10.63	< .001	(2, 63)	0.34	.711
Congruity × Group	(2, 63)	2.48	.092	(2, 63)	0.06	.938
Distance effect for factor congruity = incongruent						
Distance	(1, 63)	28.95	< .001	(1, 63)	12.23	.001
Group	(2, 63)	8.82	< .001	(2, 63)	0.08	.923
Distance × Group	(2, 63)	5.59	.006	(2, 63)	0.05	.950
Distance effect for factor congruity = congruent						
Distance	(1, 63)	27.91	< .001	(1, 63)	1.32	.254
Group	(2, 63)	9.75	< .001	(2, 63)	0.95	.393
Distance × Group	(2, 63)	0.12	.888	(2, 63)	0.36	.698

and automatized processing of numerical magnitude information. The overall analyses showed, across our three groups, typical patterns of longer reaction times and higher error rates for the more demanding conditions, i.e., numerical compared to physical size comparisons, incongruent compared to congruent conditions, and comparisons that involved smaller compared to larger numerical distances. These findings are well in line with previous studies on number comparison, in general, and numerical Stroop tasks, in particular (Cohen Kadosh, Cohen Kadosh, Linden, et al., 2007; Girelli et al., 2000; Henik & Tzelgov, 1982; Kaufmann et al., 2005; Kaufmann et al, 2006; Moyer & Landauer, 1967; Rubinsten et al., 2002; Szücs et al., 2007; Szücs & Soltész, 2007; Tang et al., 2006; Tzelgov et al., 1992).

However, apart from similar findings of variation in performance measures that depend on the different dimensions of task difficulty, a number of studies on developmental dyscalculia reported the absence of congruity effects in physical size comparison tasks in children and adults suffering from MD (Ashkenazi, Rubinsten, & Henik, 2009; Koontz & Berch, 1996; Rubinsten & Henik, 2005). Differential effects of size congruity in the context of physical comparisons for different math achievement groups were taken as evidence for automatization deficits in numerical magnitude processing in populations with MD (Rubinsten & Henik, 2005). However, in the light of two recent studies that emphasize the necessity to not only investigate the effects of congruity per se but to also look into the influence of numerical distance as a second relevant measure of automaticity in number processing (Szücs et al., 2007; Tang et al., 2006), the present study used an experimental paradigm that implemented two numerical distances for the digits pairs to be compared. This allowed us to investigate the size congruity effect as a function of small and large numerical distance conditions.

For the numerical comparison task, where *controlled* access to and processing of numerical magnitude plays the main role, the results for all congruency levels, and in particular for the neutral condition, demonstrate that the groups of children operate on different performance levels, with their performance being related to the math achievement scores.

However, apart from quantitative differences between the achievement groups, their patterns of behavior are rather similar from a qualitative point of view. These results are compatible with models of the development of normal and impaired mathematical processing that assume math achievement to be based on functions on the level of basic numerical representations (Butterworth, 1999; Wilson & Dehaene, 2007).

More relevant for the question of whether MD is related to or concomitant with automatization deficits in numerical magnitude processing are the results from the physical size comparison task. However, we will focus this part of the discussion on the analyses of reaction time data only. As overall error rates were very low for the physical size comparison task with almost half of the children making no errors at all, accuracy does not seem to be a sufficiently sensitive measure of performance for our specific experimental setting.

To begin with, the fact that the interaction of congruity and distance was significant for the physical comparison while no main effect of distance was found already suggests that the effects of congruity should probably not be interpreted without taking the effects of distance into account. Following Girelli et al. (2000), we separated the effects of size congruity for the two distance conditions. While such an analysis yielded no additional insights on the case of the numerical comparison task, for the physical comparison separate analyses of congruity effects for the two numerical distances revealed a whole new picture. When the numerical distance between the digits is small, neither of the groups shows a size congruity effect. By contrast, when the numerical distance is large, effects of congruity between task-relevant and task-irrelevant stimulus dimensions can be found in all three groups. That means that not only normal and high achievers show a numerical Stroop effect but low achievers do too. Interestingly, Girelli and colleagues (2000) reported similar interactions of numerical distance and size congruity effects in a physical comparison task. One plausible interpretation may be that for small numerical distances the magnitude information available to the child is not fine grained enough in order to influence processing under Stroop conditions in a context where the numerical magnitude is task irrelevant. Such an explanation that focuses on representational acuity functions would be in line with studies on developmental changes in numerical distance effects. It has been shown repeatedly that younger children typically show larger numerical distance effects than older children or adults (Duncan & McFarland, 1980; Holloway & Ansari, 2008). According to Holloway and Ansari, this decrease in numerical distance effects over developmental time might be related to a noisier representation of numerical and/or nonnumerical magnitude per se, or to noisier mappings between symbolic and nonsymbolic representations in earlier phases of development.

By comparing the effects of numerical distance for the congruent and the incongruent condition in the physical comparison task, further insight into children's specific performance patterns can be obtained. The combination of a classical distance effect for the congruent condition with a reversed distance effect for the incongruent condition in the groups of low and normal achievers is in line with previous studies that report similar performance patterns in studies on the number Stroop task (Henik & Tzelgov, 1992; Szücs & Soltész, 2007). Following Tang and colleagues (2006), the finding of a reversed distance effect in the incongruent condition of the physical comparison task can be interpreted as evidence of automatized processing of numerical magnitude information. All in all, the demonstration of both a congruity effect, at least for the large distance, and a reversed distance effect for the group of children with math difficulties does not support the notion of automatization deficits as necessarily tied to the development of MD.

While the performance patterns of normal children and children with problems in the domain of mathematical processing do not differ on the whole for the physical comparison, it becomes apparent on this deeper level of analysis that the performance of the high achievers diverges substantially. Similar to the performance of low and normal achievers for the large distance condition, high achievers exhibit an effect of facilitation when stimulus information is congruent. However, what is different is the apparent lack of the massively detrimental effect of incongruity that can be gathered from the absence of a reversed distance effect for the large distance condition. These results can be explained by classical theories on Stroop processing. It was suggested that the overall Stroop effect can actually be divided into two separate components, i.e., *facilitation*, which is reflected in the performance gain for the congruent compared to the neutral condition, and *inhibition*, which is reflected by the loss in performance for the incongruent compared to the neutral condition. It is assumed that these two Stroop components are based on somewhat different underlying cognitive mechanisms, with the facilitatory component relying more on automaticity of processing (MacLeod & Dunbar, 1988), and interference drawing mainly on attentional or inhibitory control resources (Posner, 1978). Such an explanation of the performance in the case of the incongruent condition links the results for the high achievers to their considerably higher scores in the KFT. In an early developmental study on children's performance in the classical color Stroop paradigm, Friedman (1971) showed that, at least in children of our age group, general intelligence and small interference effects under Stroop conditions are correlated.

This means that an explanation of the missing reversed distance effect should probably focus on the fact that the group of HA reached significantly higher scores for the executive control measures of the working memory test (WMTB-C; Pickering & Gathercole, 2001) than both other groups. Our results might be an indicator of differences in the ability to control attentional resources when the children are confronted with conflicting information. Such an explanation would fit the findings of an fMRI study on the numerical Stroop task, where higher activation in the dorsolateral prefrontal (DLPFC) and anterior cingulate (ACC) cortices were found for incongruent compared to congruent trials (Kaufmann et al., 2005). According to MacDonald, Cohen, Stenger, and Carter (2000), activations in these regions can be associated with top-down control of task-appropriate behavior (DLPFC) and evaluatory processes (ACC) under conflict. Such an interpretation of the results of the high achievers in the physical comparison task is well in line with the idea originally proposed by Dempster and Corkill (1999), who assume that resistance to interference is an important factor in general cognitive ability; i.e., children with higher IQ should be less prone to intrusions from irrelevant sources of information due to more efficient inhibitory control (cf. Engle, Kane, & Tuholski, 1999 for an extensive discussion of the relationship between intelligence and executive control functioning). The specific relevance of inhibitory control for children's performance in numerical Stroop tasks is corroborated by a study on children with attention deficit/hyperactivity disorder (ADHD), which found larger interference effects in children with selective attention deficits compared to their normal peers (Kaufmann & Nuerk, 2006). And finally, the finding that the reaction times of HA in the digit-naming condition of the RAN tasks were not different from those of the other two groups allows for the conclusion that differences in general numerical processing speed cannot underlie these effects (Denckla & Rudel, 1974).

In conclusion, we found that when the influence of numerical distance is taken into consideration, children of different achievement groups—i.e., low, normal, and high achievers in the domain of mathematics—show similar congruity effects in each of the two subtasks of the number Stroop test. For the numerical comparison we found congruity effects for both numerical distances in all groups as well as classical distance effects for each of the congruity conditions. In the physical size comparison, all children show an effect of congruity for the large numerical distance, while this effect is missing in all groups for the comparisons involving small numerical distances. The finding of considerably large reversed numerical distance effects not only for the group of normal achievers but also for the children with MD is not easily reconciled with theories that propose automatization deficits to be causally related to developmental math disabilities. We suggest that future research on differential congruity effects in numerical Stroop tasks across age or achievement groups should not disregard differences in representational acuity as an explanatory factor. Furthermore, the lack of a significant reversed distance effect in the group of high achievers for the physical comparison task, which becomes apparent only when the impacts of congruity and distance are separated, corroborates the findings of previous studies that have shown that apart from differences on the level of numerical magnitude representations, differences even at the stage of response execution—i.e., related to the inhibition of prepotent reactions—may influence Stroop performance.

Overall, our results suggest that models that focus on different cognitive subdomains and processing levels (e.g., Szücs & Soltész, 2007) may be more appropriate to explain the complex performance patterns in numerical Stroop tasks than approaches that focus on one explanatory factor only. Furthermore, apart from conclusions related directly to the questions of automatized numerical processing in different achievement groups, on a methodological level these results underline the necessity to investigate the interaction of the two separable stimulus dimensions in the context of numerical Stroop tasks. Only by separating the influences of the dimensions of congruity and numerical distance is it possible to identify specific performance patterns and to explain them in a consistent manner.

Original manuscript received October 17, 2008
Revised manuscript accepted February 7, 2010
First published online April 29, 2010

REFERENCES

Ashkenazi, S., Rubinsten, O., & Henik, A. (2009). Attention, automaticity, and developmental dyscalculia. *Neuropsychology*, *23*, 535–540.

Besner, D., & Coltheart, M. (1979). Ideographic and alphabetic processing in skilled reading of English. *Neuropsychologia*, *17*, 467–472.

Butterworth, B. (1999). *The mathematical brain*. London: Macmillan.

Catts, H. W., Gillespie, M., Leonard, L. B., Kail, R. V., & Miller, C. A. (2002). The role of speed of processing, rapid naming, and phonological awareness in reading achievement. *Journal of Learning Disabilities*, *35*, 509–524.

Cohen Kadosh, R., Cohen Kadosh, K., Linden, D. E. J., Gevers, W., Berger, A., & Henik, A. (2007). The brain locus of interaction between number and size: A combined functional magnetic resonance imaging and event-related potential study. *Journal of Cognitive Neuroscience*, *19*, 957–970.

Cohen Kadosh, R., Cohen Kadosh, K., Schumann, T., Kaas, A., Goebel, R., Henik, A., et al. (2007). Virtual dyscalculia induced by parietal lobe TMS impairs automatic magnitude processing. *Current Biology*, *17*, 1–5.

Conners, C. K. (1973). Rating scales for use in drug studies with children. *Psychopharmacology Bulletin*, *9*, 24–84.

Dehaene, S., & Cohen, L. (1995). Towards an anatomical and functional model of number processing. *Mathematical Cognition*, *1*, 83–120.

Delaney, H. D., & Maxwell, S. E. (1981). On using analysis of covariance in repeated measures designs. *Multivariate Behavioral Research*, *16*, 105–123.

Dempster, F. N., & Corkill, A. J. (1999). Individual differences in susceptibility to interference and general cognitive ability. *Acta Psychologica*, *101*, 395–416.

Denckla, M. B., & Rudel, R. G. (1974). Rapid automatized naming of pictured objects, colors, letters and numbers by normal children. *Cortex*, *10*, 186–202.

Duncan, E. M., & McFarland, C. E., Jr. (1980). Isolating the effects of symbolic distance and semantic congruity in comparative judgments: An additive-factors analysis. *Memory & Cognition*, *8*, 612–622.

Engle, R. W., Kane, M. J., & Tuholski, S. W. (1999). Individual differences in working memory capacity and what they tell us about controlled attention, general fluid intelligence and functions of the prefrontal cortex. In A. Miyake & P. Shah, (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control*. pp. 102–134. London: Cambridge Press.

Friedman, R. (1971). The relationship between intelligence and performance on the Stroop color-word test in second- and fifth-grade children. *The Journal of Genetic Psychology*, *118*, 147–148.

Girelli, L., Lucangeli, D., & Butterworth, B. (2000). The development of automaticity in accessing number magnitude. *Journal of Experimental Child Psychology*, *76*,104–122.

Haffner, J., Baro, K., Parzer, P., & Resch, F. (2005). *Heidelberger Rechentest: Erfassung mathematischer Basiskompetenzen im Grundschulalter, HRT 1-4*. Göttingen: Hogrefe.

Heller, K., & Geisler, H. J. (1983). *Kognitiver Fähigkeitstest – Grundschulform, KFT 1-3*. Weinheim: Beltz.

Henik, A., & Tzelgov, J. (1982). Is three greater than five: The relation between physical and semantic size in comparison tasks. *Memory & Cognition*, *10*, 389–395.

Holloway, I. D., & Ansari, D. (2008). Domain-specific and domain-general changes in children's development of number comparison. *Developmental Science*, *11*, 644–649.

Kaufmann, L., Koppelstätter, F., Delazer, M., Siedentopf, C., Rhomberg, P., Golaszewski, S., et al. (2005). Neural correlates of distance and congruity effects in a numerical Stroop task: An event-related fMRI study. *Neuroimage*, *25*, 888–898.

Kaufmann, L., Koppelstätter, F., Siedentopf, C., Haala, I., Haberlandt, E., Zimmerhackl, L. B., et al. (2006). Neural correlates of the number-size interference task in children. *Neuroreport*, *17*, 587–591.

Kaufmann, L., & Nuerk, H. C. (2006). Interference effects in a numerical Stroop paradigm in 9- to 12-year-old children with ADHD-C. *Child Neuropsychology*, *12*, 223–243.

Koontz, K. L., & Berch, D. B. (1996). Identifying simple numerical stimuli: Processing inefficiencies exhibited by arithmetic learning disabled children. *Mathematical Cognition*, *2*, 1–23.

Landerl, K., Bevan, A., & Butterworth, B. (2004). Developmental dyscalculia and basic numerical capacities: A study of 8- to 9-year-old students. *Cognition*, *93*, 99–125.

Landerl, K., Wimmer, H., & Moser, E. (1997). *Salzburger Lese- und Rechtschreibtest: Verfahren zur Differentialdiagnose von Störungen des Lesens und Schreibens für die 1. bis 4. Schulstufe, SLRT 1-4*. Bern: Hans Huber.

Llorente, A. M., Williams, J., Satz, P., & D`Elia, L. F. (2003). *Children's color trails test*. Lutz, FL: Psychological Assessment Resources.

MacDonald, A. W., Cohen, J. D., Stenger, V. A., & Carter, C. S. (2000). Dissociating the role of dorsolateral prefrontal cortex and anterior cingulate cortex in cognitive control. *Science*, *288*, 1835–1837.

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.

Mayringer, H., & Wimmer, H. (2003). *Salzburger Lese-Screening für die Klassenstufen 1-4, SLS 1-4*. Bern: Hans Huber.

Moyer, R. S., & Landauer, T. K. (1967). The time required for numerical inequality. *Nature, 215*, 1519–1520.

Pickering, S., & Gathercole, S. (2001). *WMTB-C: Working Memory Test Battery for Children*. London: The Psychological Corporation.

Pinel, P., Piazza, M., LeBihan, D., & Dehaene, S. (2004). Distributed and overlapping cerebral representations of number, size, and luminance during comparative judgments. *Neuron, 41*, 983–993.

Posner, M. (1978). *Chronometric explorations of mind*. Hillsdale: Erlbaum.

Rousselle, L., & Noël, M.-P. (2007). Basic numerical skills in children with mathematics learning disabilities: A comparison of symbolic vs. non-symbolic number magnitude processing. *Cognition, 102*, 361–395.

Rousselle, L., & Noël, M.-P. (2008). The development of automatic numerosity processing in preschoolers: Evidence for numerosity-perceptual interference. *Developmental Psychology, 44*, 544–560.

Rubinsten, O., & Henik, A. (2005). Automatic activation of internal magnitudes: A study of developmental dyscalculia. *Neuropsychology, 19*, 641–648.

Rubinsten, O., Henik, A., Berger, A., & Shahar-Shalev, S. (2002). The development of internal representations of magnitude and their association with Arabic numerals. *Journal of Experimental Child Psychology, 81*, 74–92.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology, 18*, 613–622.

Szücs, D., & Soltész, F. (2007). Event-related potentials dissociate facilitation and interference effects in the numerical Stroop paradigm. *Neuropsychologia, 45*, 3190–3202.

Szücs, D., & Soltész, F. (2008). The interaction of task-relevant and task-irrelevant stimulus features in the number/size congruency paradigm: An ERP study. *Brain Research*. *1190*, 143–158.

Szücs, D., Soltész, F., Jármi, É., & Csépe, V. (2007). The speed of magnitude processing and executive functions in controlled and automatic number comparison in children: An electro-encephalography study. *Behavioural and Brain Functions, 3/23*. doi: 10.1186/1744-9081-3-23

Tang, J., Critchley, H. D., Glaser, D. E. Dolan, R. J., & Butterworth, B. (2006). Imaging informational conflict: A functional magnetic resonance imaging study of numerical Stroop. *Journal of Cognitive Neuroscience, 15*, 2049–2062.

Tukey, J. W. (1977). *Exploratory data analysis*. Reading, MA: Addison-Wesley.

Tzelgov, J., Meyer, J., & Henik, A. (1992). Automatic and intentional processing of numerical information. *Journal of Experimental Psychology: Learning, Memory and Cognition, 18*, 166–179.

Walsh, V. (2003). A theory of magnitude: Common cortical metrics of time, space and quantity. *Trends in Cognitive Sciences, 7*, 483–488.

Wilson, A. J., & Dehaene, S. (2007). Number sense and developmental dyscalculia. In D. Coch, G. Dawson, & K. Fischer (Eds.), *Human behavior, learning and the developing brain: Atypical development*. pp. 212–238. New York: Guilford Press.