Research in Developmental Disabilities 64 (2017) 152–167

Contents lists available at ScienceDirect

Research in Developmental Disabilities

Research paper

(Non-)symbolic magnitude processing in children with mathematical difficulties: a meta-analysis

Christin Schwenk[a,*], Delphine Sasanguie[b,c], Jörg-Tobias Kuhn[a], Sophia Kempe[a], Philipp Doebler[d], Heinz Holling[a]

[a] Institute of Psychology, University of Münster, Fliednerstraße 21, 48149 Münster, Germany
[b] Brain and Cognition, KU Leuven, Tiensestraat 102 - box 3711, 3000 Leuven, Belgium
[c] Faculty of Psychology and Educational Sciences@Kulak, KU Leuven Kulak, Etienne Sabbelaan 53, 8500 Kortrijk, Belgium
[d] TU Dortmund University, Faculty of Statistics, Technische Universität Dortmund, 44221 Dortmund, Germany

ARTICLE INFO

Article history:
Received 1 August 2016
Received in revised form 21 February 2017
Accepted 6 March 2017
Available online 19 April 2017
Number of reviews: 2

Keywords:
Mathematical difficulties
Developmental dyscalculia
Symbolic magnitude comparison
Non-symbolic magnitude comparison
Distance effect
Meta-analysis

ABSTRACT

Symbolic and non-symbolic magnitude representations, measured by digit or dot comparison tasks, are assumed to underlie the development of arithmetic skills. The comparison distance effect (CDE) has been suggested as a hallmark of the preciseness of mental magnitude representations. It implies that two magnitudes are harder to discriminate when the numerical distance between them is small, and may therefore differ in children with mathematical difficulties (MD), i.e., low mathematical achievement or dyscalculia. However, empirical findings on the CDE in children with MD are heterogeneous, and only few studies assess both symbolic and non-symbolic skills. This meta-analysis therefore integrates 44 symbolic and 48 non-symbolic response time (RT) outcomes reported in 19 studies (N = 1,630 subjects, aged 6–14 years). Independent of age, children with MD show significantly longer mean RTs than typically achieving controls, particularly on symbolic (Hedges' g = 0.75; 95% CI [0.51; 0.99]), but to a significantly lower extent also on non-symbolic (g = 0.24; 95% CI [0.13; 0.36]) tasks. However, no group differences were found for the CDE. Extending recent work, these meta-analytical findings on children with MD corroborate the diagnostic importance of magnitude comparison speed in symbolic tasks. By contrast, the validity of CDE measures in assessing MD is questioned.

What this paper adds

This meta-analysis adds substantially to the existing body of research on numerical cognition in children with low mathematical achievement and dyscalculia (i.e., children with mathematical difficulties, MD). Being the first quantitative meta-analysis that explicitly focuses on this clinically relevant population, it sheds light on the diagnostic meaning of different measures of magnitude processing. Most importantly, it corroborates the significance of mainly symbolic (i.e., digit) comparison speed as a measure that identifies children with MD compared to typical achievers. This result is in line with the access-deficit hypothesis rather than with the assumption that MD up to dyscalculia arise from problems with the innate Approximate Number System (ANS) per se. Moreover, the meta-analytical results support the recent discussion

* Corresponding author.
E-mail addresses: christin.schwenk@uni-muenster.de (C. Schwenk), delphine.sasanguie@kuleuven.be (D. Sasanguie), t.kuhn@uni-muenster.de (J.-T. Kuhn), sophia.kempe@googlemail.com (S. Kempe), doebler@statistik.tu-dortmund.de (P. Doebler), holling@uni-muenster.de (H. Holling).

http://dx.doi.org/10.1016/j.ridd.2017.03.003
0891-4222/

criticizing the CDE as an index of the symbolic magnitude representation (Lyons, Nuerk, & Ansari, 2015). The meta-analytical model explicitly controlled for statistical dependencies between multiple effects derived from the same study by using robust variance estimation. A main benefit of this procedure, compared to other approaches such as stratification, within-study pooling or selecting only one outcome per study, is that all the available outcomes are taken into account (Hedges, Tipton, & Johnson, 2010). Taken together, this study extends the current state of research on numerical cognition in school-aged children with MD, both in respect of the clinically relevant population it focuses on and the meta-analytical methodology employed.

1. Introduction

Numerical processing skills and broader mathematical competencies help children deal with many everyday tasks and future professional activities (e.g., Ancker & Kaufman, 2007). In contrast, low mathematical skills negatively impact quality of life (Parsons & Bynner, 2005) and economic well-being (Ritchie & Bates, 2013). Importantly, a substantial number of primary school children experience learning difficulties in mathematics, which are referred to as developmental dyscalculia (DD) when causing an atypical numerical development despite normal intelligence and educational opportunities. Prevalence estimates of DD vary between 3 and 7% (Butterworth, 2005; Reigosa et al., 2008; Rubinsten & Henik, 2009).

During the past two decades, an increasing number of studies aimed at unravelling the cognitive mechanisms behind the development of mathematical difficulties, and consistently revealed that children with MD are impaired in numerical magnitude processing tasks (see De Smedt, Noël, Gilmore, & Ansari, 2013 for a literature review). However, it remains unclear to what extent the processing of symbolic (i.e., digits) or non-symbolic (i.e., arrays of dots or other objects) magnitudes or both, is affected. Heterogeneous results in this regard have led to two different etiological hypotheses: the ANS deficit hypothesis (Wilson & Dehaene, 2007) versus the access deficit hypothesis (Rousselle & Noël, 2007).

According to the *ANS deficit hypothesis*, the impairments originate from deficits in the Approximate Number System (ANS), an internal analogue magnitude system which allows humans to represent and manipulate approximate numerosities (Feigenson, Dehaene, & Spelke, 2004). Evidence for this hypothesis has been provided by studies demonstrating that children with MD have problems with non-symbolic magnitude processing (e.g., Mazzocco, Feigenson, & Halberda, 2011; Piazza et al., 2010) or both non-symbolic and symbolic magnitude processing (e.g., Landerl, Fussenegger, Moll, & Willburger, 2009; Mussolin, Mejias, & Noël, 2010), as symbolic magnitudes are assumed to be mapped onto the ANS (Mundy & Gilmore, 2009; for an alternative view see Noël & Rousselle, 2011; Sasanguie, De Smedt & Reynvoet, 2017). By contrast, the *access deficit hypothesis* assumes that children with MD do not have an ANS dysfunction per se, but rather a problem with accessing the ANS when magnitudes are expressed symbolically (Rousselle & Noël, 2007). This idea emerged from studies reporting deficient symbolic, but intact non-symbolic, magnitude processing in children with MD (e.g., Andersson & Östergren, 2012; De Smedt & Gilmore, 2011; Landerl & Kölle, 2009).

To integrate the findings mentioned above, a quantitative meta-analysis is necessary. Recently, three meta-analyses reviewed the associations between magnitude processing and mathematical competencies in unselected populations. Because some authors argue that MD up to DD form part of a continuum of ability (e.g., Dowker, 2009), we here briefly summarize these meta-analyses: Chen and Li (2014) and Fazio, Bailey, Thompson and Siegler (2014) included non-symbolic outcomes only. Both meta-analyses report a weak but reliable association with mathematical competence (i.e., a correlation of $r = .20$ and $r = .22$, respectively). Schneider et al. (2016) extended these findings by also including symbolic magnitudes. Based on 284 effect sizes, their analyses showed a significantly larger effect for symbolic ($r = .30$) than for non-symbolic ($r = .24$) magnitude processing, which decreased slightly with age. Furthermore, they observed the highest correlations for response times (RT) and Weber fractions (i.e., the smallest ratio of two numerosities that one can reliably judge as larger or smaller; Halberda, Mazzocco, & Feigenson, 2008). However, the abovementioned meta-analyses do not offer satisfying evidence to evaluate the two etiological hypotheses about the magnitude processing impairments observed in children with MD. The current meta-analysis therefore closes this gap for this clinically relevant group.

Numerical magnitude processing is most frequently assessed by comparison tasks (Ansari, 2008; Lyons et al., 2015). In such tasks, participants are instructed to select as quickly and accurately as possible which of two visually presented magnitudes is numerically larger. Visual stimuli can either be symbolic or non-symbolic. Typically, a comparison distance effect (CDE), or a conceptually similar ratio effect, is observed: Error rates and RT decrease with increasing distance between the magnitudes at comparison (or a ratio between the magnitudes that substantially differs from 1). This has traditionally been explained by assuming a cognitive magnitude representation on a mental number line, with small magnitudes on the left and larger magnitudes on the right. Each magnitude is represented with certain noise, expressed as a Gaussian distribution around the corresponding quantity (i.e., the mental number line hypothesis, Dehaene, 1997). Consequently, the CDE is thought to reflect the activation of magnitude representations on the mental number line, or in other words, the ANS (Price & Ansari, 2013; but see van Opstal, Gevers, de Moor & Verguts, 2008, modelling the CDE as a decisional process). The size of the CDE has even been assumed to index ANS precision: A smaller CDE was regarded as a more precise, and a larger effect as a less precise underlying representation (cf. Lyons et al., 2015). This assumption has led to several major theoretical claims, one of which is that children with MD should show a larger CDE because they have more noisy mental magnitude representations than typically achieving peers (e.g., Mussolin et al., 2010).

Against this background, we meta-analyzed the CDE of children with MD on comparison tasks. In line with Lyons et al. (2015), we chose RT instead of the popular Weber fractions, which focus exclusively on error rates, as measure of performance because we were primarily interested in commonalities and differences between symbolic and non-symbolic comparison performance, and error rates tend to be very low for symbolic comparisons (Lyons et al., 2015).

Several potential moderators of the performance of children with MD on symbolic and non-symbolic magnitude processing tasks have been suggested. First, Noël and Rousselle (2011) observed a dissociation concerning participants' age: Only studies examining older children (10-year-olds and above) showed significant differences between dyscalculic and control children in non-symbolic magnitude processing, whereas studies with younger dyscalculic children (6–9 years) only revealed different performances in symbolic, but not non-symbolic magnitude processing. Second, the cut-off score on the math test used to identify children with MD needs to be taken into account (Butterworth, 2005; Geary, 2013): Some studies concentrate on subjects performing at or below the 10th percentile, which are commonly referred to as dyscalculic. Others set less strict criteria up to the 25th percentile rank (PR), which includes a group of children (i.e., PR 11–25) that is sometimes considered as 'low achievers' (e.g., Geary, 2013). We included studies with both kinds of samples and refer to them as *children with MD* hereinafter. A maximum of PR 25 was chosen to both account for etiological variety of the disorder (Mazzocco et al., 2011) and actual prevalence rates (Swanson & Jerman, 2006). Third, while most non-symbolic tasks are designed to control for non-numerical visual parameters (e.g., item size or total dot surface) which carry heuristic information of the magnitude, the degree of control varies (De Smedt et al., 2013). This has an impact on the extent to which visual cues can be used when performing on comparison tasks, which might translate to the relative impairments we observe in MD children when compared to their typically achieving (TA) peers.

Consequently, we addressed two main research questions with the current meta-analysis: First, do children with MD differ from TA controls in terms of (a) response times and (b) the CDE on response times in either symbolic or non-symbolic magnitude comparison tasks or both? Second, to what extent are potential group differences moderated by (a) sample characteristics (age, math test cut-off criterion) and (b) task characteristics (number range, use of stimuli in the subitizing range, or visual properties of non-symbolic magnitude stimuli)? By statistically integrating the evidence in this regard, we can provide a more profound answer to the abovementioned claims: (1) Is it correct that children with MD show a larger CDE than control children and can we therefore conclude that they have a more imprecise representation of numbers?, (2) Is there indeed a developmental trajectory in the performance of children with MD on symbolic versus non-symbolic tasks as suggested by Noël and Rousselle (2011)?, and (3) Which hypothesis is more statistically robust: the ANS deficit hypothesis or the access deficit hypothesis?

2. Methods

2.1. Identification of studies

Literature search started in December 2013 (see flow chart, Fig. 1). First, the search term combining sample, tasks and outcome key words, was entered into the ERIC and EBSCOhost online databases, producing almost 3,000 hits. Reference lists of key studies and reviews were screened for further relevant publications (i.e., snowball search, producing two further hits). Moreover, 71 experts within the domain of numerical cognition and math learning difficulties were asked for unpublished data, using a standardized email form, which resulted in six data sets provided by them.

All abstracts of the identified studies were screened for the following eight inclusion criteria: (1) The publication language is English, German, or Dutch. (2) At least one sample of children with mathematical difficulties, i.e., low mathematical achievement or dyscalculia, is compared with a typically achieving control group. Studies examining only general learning disabilities were excluded. (3) We focused on an age range between six and 14 years (cf. Noël & Rousselle, 2011). The population of adolescents and adults was excluded because of differences in mathematical motivation and cognitive aspects like verbal and non-verbal intelligence. (4) At least one symbolic or non-symbolic magnitude comparison task had to be reported, that (5) captured RT data. More specifically, we included all comparison tasks with the instruction 'indicate the larger', while excluding comparison tasks making use of other types of instructions, such as 'indicate whether the magnitudes are numerically the same or different' (cf. Schneider et al., 2016). Moreover, symbolic tasks were only considered if the numbers to be compared did not differ in physical size (thus studies on physical Stroop effects were excluded). (6) To make sure that mathematical deficits could not be accounted for by general intellectual impairments (DSM-5, American Psychiatric Association, 2013), samples had to have an at least average $IQ \geq 80$[1]. (7) Further, the MD diagnosis had to be the result of systematic testing, with scores below a cut-off of PR 25 on standardized mathematical achievement tests, or below a defined criterion on criterion-based tests. (8) For statistical power reasons, only samples with $N \geq 10$ subjects per group were selected (Morris, 2008).

In case that a study met all criteria, but data on RT – or a split of data in small versus large distance trials for determining the CDE – was missing, the authors were contacted. If the MD group was selected according to a PR larger than 25, the authors were asked to provide us with a subset of the sample of interest. In total, 18 published and one unpublished study (1,630 subjects) were included, with 44 outcomes for symbolic and 48 outcomes for non-symbolic magnitude comparisons (see Table 1).

[1] The common definition of below average intelligence applies the –1SD criterion, i.e., $IQ \leq 85$. However, as two of the samples included in our meta-analysis (Kucian et al., 2011; Kuhn et al., 2013) contained few children with $80 \leq IQ < 85$, we set the overall criterion to $IQ \geq 80$.

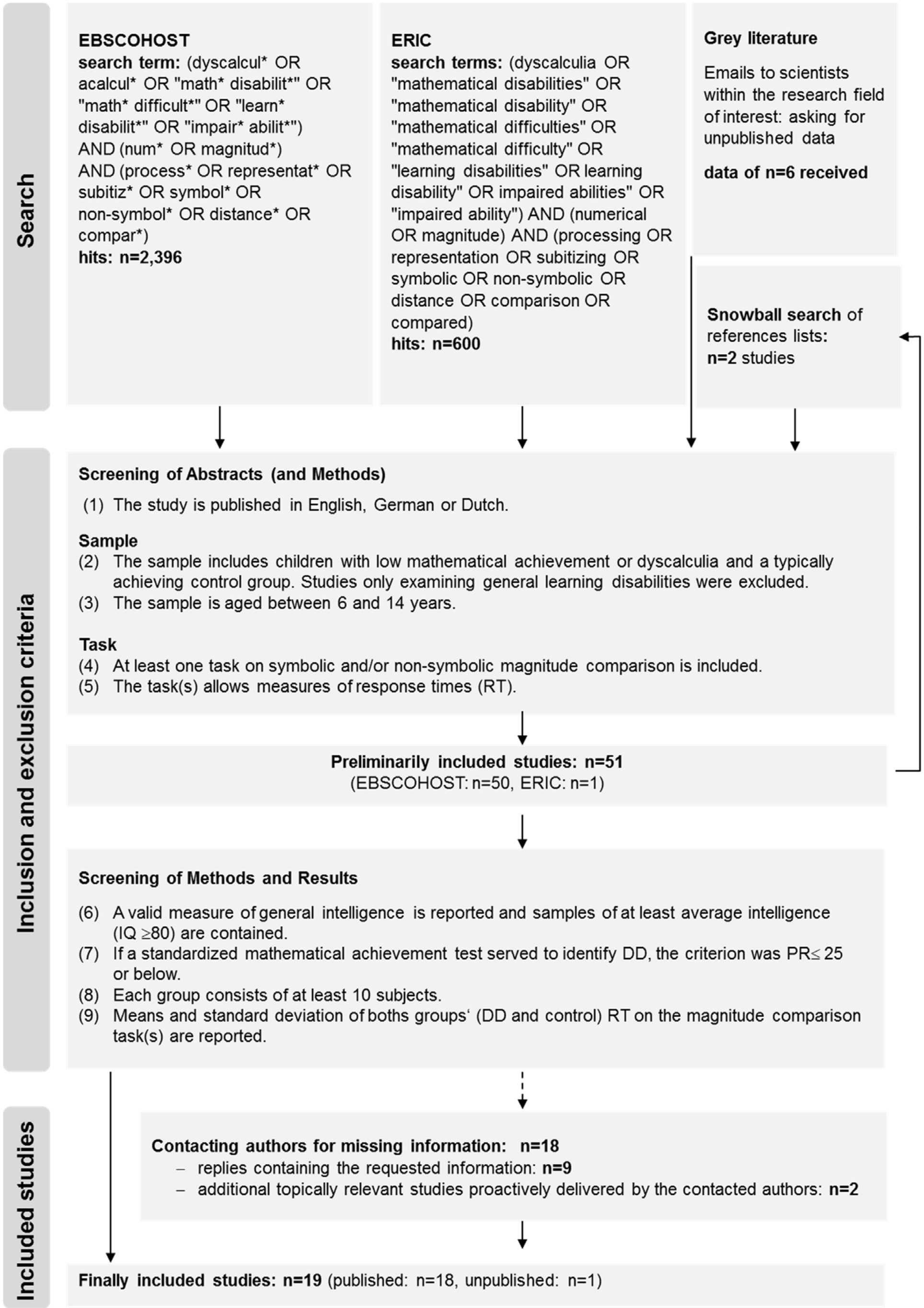

Fig. 1. Flow chart of search process and inclusion criteria.

2.2. Coding

The coding sheet captured study and sample characteristics, tasks details and relevant outcomes, i.e., mean and standard deviation (SD) of response times per group, also separated for small versus large numerical distances, if available. All studies were coded by the first author, a doctoral student, and the fourth author, a master student with experience in numerical cognition research. Initial interrater agreement was satisfactory (>77% agreement or consistency Intra-Class-Correlation ICC ≥. 97 for the numerical codes, and > 97% agreement or Cohen's K > .94 for the categorical codes). All discordances could be resolved after consultation with the second and third author.

Table 1
Included studies.

Study	Outcomes		Sample			
	Format	k	Math criterion	Baseline reaction speed	IQ	Reading speed
Andersson and Östergren (2012)	Symbolic	2 (2)	unstandardized	MD < CON	MD < CON	NA
	Non-sym.	1 (0)				
Ashkenazi et al. (2009)	Symbolic	1 (0)	unstandardized	MD = CON	MD = CON	MD = CON
Brankaer et al. (2014)	Symbolic	1 (1)	standardized	MD < CON	MD = CON	MD < CON
	Non-sym.	1 (1)				
Chan, Au, and Tang (2013)	Symbolic	1 (0)	unstandardized	NA	NA	NA
de Oliveira Ferreira et al. (2012)	Symbolic	2 (0)	standardized	MD = CON	MD = CON	MD*: MD = CON
	Non-sym.	2 (0)				MD + L*: MD < CON
De Smedt and Gilmore (2011)	Symbolic	2 (2)	standardized	NA	NA	NA
	Non-sym.	2 (2)				
Defever et al. (2013)	Non-sym.	1 (1)	standardized	MD = CON	MD = CON	
Dinkel et al. (2013)	Non-sym.	1 (0)	standardized	NA	NA	NA
Grond, Kucian, O'Gorman, Martin, and von Aster (Unpublished)	Non-sym.	1 (1)	standardized	MD = CON	MD < CON	NA
Heine et al. (2013)	Non-sym.	4 (4)	standardized	NA	MD = CON	NA
Kucian et al. (2011)	Non-sym.	1 (1)	standardized	NA	NA	NA
Kuhn et al. (2013)	Symbolic	2 (2)	standardized	MD = CON	RS*: MD < CON	MD < CON
					DYS*: MD = CON	
Landerl (2013)	Symbolic	2 (2)	standardized	NA	MD < CON	MD < CON
	Non-sym.	1 (1)				
Landerl and Kölle (2009)	Symbolic	4 (4)	standardized	MD = CON	MD < CON	MD < CON
	Non-sym.	2 (2)				
Mussolin et al. (2010)	Symbolic	1 (1)	unstandardized	NA	MD = CON	MD = CON
	Non-sym.	3 (3)				
Piazza et al. (2010)	Non-sym.	1 (0)	standardized	NA	MD = CON	MD < CON
Rousselle and Noël (2007)	Symbolic	2 (2)	unstandardized	NA	MD = CON	MD = CON
	Non-sym.	2 (2)				
Skagerlund and Träff (2014)	Symbolic	1 (1)	unstandardized	MD < CON	MD < CON	NA
	Non-sym.	1 (0)				
Vanbinst et al. (2014)	Symbolic	3 (3)	standardized	NA	MD = CON	MD = CON
	Non-sym	3 (3)				

Note: k: number of effect sizes on mean response time group difference (and on the comparison distance effect).
*Sample names as used in the original studies: *MD*: sample with math difficulties, *MD + L*: sample with math difficulties associated with language difficulties; *MLD*: sample with mathematics learning disabilities, *LA*: sample with low achievement; *RS*: sample with math difficulties not fulfilling the IQ discrepancy criterion, *DYS*: DD sample fulfilling the IQ discrepancy criterion; MD = CON: no significant differences between MD and control group, MD<CON: significantly lower baseline for reaction speed (i.e., higher response times), lower IQ or lower reading speed of the MD compared to the control group.

2.2.1. *Sample characteristics*

Each sample was more closely defined by its mean age and the PR used as cut-off for MD, which were assessed as potential moderators. Four control criteria were considered to further judge the quality of the sample: First, the source of MD diagnosis (standardized vs. unstandardized test) as expression of diagnostic validity; second, it was coded whether the MD and control groups differed in their performance on simple reaction tasks, since such general reaction speed differences could partly account for differences in magnitude comparison speed. Third and fourth, it was coded whether the MD and the control group differed in reading speed or IQ (Tables 1 and 2).

2.2.2. *Task characteristics*

Symbolic comparison outcomes were specified by their number range (single versus two-digit). In non-symbolic outcomes, we differentiated between tasks containing or not containing items in the subitizing range (i.e., magnitudes up to four items which can readily be extracted from a set without counting). Moreover, following Gebuis and Reynvoet (2012), we coded whether the following visual parameters were controlled for: (1) diameter, (2) density, (3) total surface and (4) convex hull, i.e., the shortest possible contour line around the entire stimulus set. A composite score was calculated, reflecting the number of controlled properties. For the computation of the CDE, small and large distance trials were assigned as defined by the authors of each study.

2.3. *Statistical analyses*

All analyses were conducted separately for (a) symbolic and (b) non-symbolic effects, using the R packages *metafor* (Viechtbauer, 2010) and *robumeta* (Fisher & Tipton, 2015). To assess overall group differences in mean RT, 44 symbolic and 48 non-symbolic outcomes could be included (Table 1). To assess group differences in CDE, we could only consider studies

Table 2
Variables used in meta-regression.

Attribute	Variable	Metrics	
Sample	Age	Continuous	in years
Sample	Cut-off	Categorical	1 = 10 < PR ≤ 25 0 = PR ≤ 10
Sample	Baseline reaction speed*	Categorical	0 = no group difference 1 = group difference
Sample	Math criterion*	Categorical	1 = standardized test 0 = unstandardized test
Sample	Intelligence*	Categorical	0 = no group difference 1 = group difference
Sample	Reading speed*	Categorical	0 = no group difference 1 = group difference
Task: Symbolic	Number range	Categorical	0 = single digit 1 = two-digit
Task: Non-symbolic	Subitizing	Categorical	0 = no magnitudes in subitizing range 1 = magnitudes in subitizing range
Task: Non-symbolic	Visual properties	Continuous	0-4: number of visual properties (diameter, density, total surface and convex hull) controlled for

Note: * Control variables.

in which a split of RT data for small versus large trials was provided in the paper or on request (20 symbolic and 21 non-symbolic outcomes). In the majority of cases 'small' referred to numerical differences of one or two, and 'large' to differences of three and larger[2].

Moderating effects of the sample and task characteristics were assessed by simple meta-regression, trying to explain the heterogeneity in group differences across outcomes by each potential moderator.

2.3.1. *Calculation of effect sizes*

This meta-analysis compares two distinct groups, a MD and a TA control group. Contrary to studies that investigate the full spectrum of mathematical competence by means of correlational analysis (e.g., Schneider et al., 2016), we use standardized group differences. To assess whether children with MD differ from their TA peers, the groupwise means and SD of each symbolic and non-symbolic outcome were integrated to a standardized mean difference, assuming heteroscedastic population variances in the two groups and corrected for positive bias, i.e., a variant of Hedges' *g* (Hedges, 1981; Bonett, 2009).

To assess whether children with MD and TA children differ in their CDE, this interaction was modelled as the group difference in distance effects, adapting a procedure by Morris (2008). For each group, the CDE was operationalized as the difference of RT for small and large distances, divided by the SD of the large distance trials.

2.3.2. *Meta-analytical model*

To account for differences in sample and task characteristics between studies, and statistical dependencies of multiple endpoints from the same study, a random effects model with robust variance estimation (RVE; Hedges et al., 2010; Tipton, 2015) was adopted. RVE deals with dependencies between multiple outcomes without requiring effect selection or integration (Hedges et al., 2010; Fisher & Tipton, 2015). It operates by correcting standard errors and requires fewer distributional assumptions and statistical power than multilevel modelling. RVE needs an estimate of the correlation ρ between all pairs of observed effect sizes within each study. In this meta-analysis, ρ was set to the default value of .80 (Tanner-Smith & Tipton, 2014). Sensitivity analyses with ρ varied between 0 and 1 were conducted to assess the impact of this decision on the estimated overall effect ($\hat{g}$), the standard error of its estimation (*SE*) and the between-study variance (τ^2). The assumption of heterogeneity was assessed by the RVE-version of the *Q*-test, and the I^2 value informs on the amount of between-study variance in excess of variance due to sampling error (Higgins, Thompson, Deeks, & Altman, 2003).

[2] For the non-symbolic tasks, three out of 11 studies (Defever et al., 2013; Grond et al., unpublished; Rousselle & Noël, 2007) with four out of 21 outcomes reported ratios instead of differences to define small or large.

3. Results

3.1. Group differences in response time

The overall RT difference of children with MD and controls was significantly different from zero ($p < .001$) for both symbolic ($\hat{g} = 0.75$, Fig. 2, Table 3) and non-symbolic tasks ($\hat{g} = 0.24$, Fig. 3, Table 3). However, the effect on symbolic outcomes was significantly larger ($t(15.3) = 4.06$, $p < .001$). Sensitivity analysis showed that the impact of the assumed within-study correlation between multiple effects (ρ) on $\hat{g}$, SE and τ^2 was negligible for all outcomes.

The Q-test indicated significant heterogeneity only for symbolic outcomes ($Q \sim \chi^2(12.08) = 31.10$, $p < .002$; $\tau^2 = 0.10$; $I^2 = 61.14\%$; Table 3). This heterogeneity could not be explained by any of the tested moderator variables (all $ps > .22$, Table 4). For example, although the effect was slightly stronger in samples with the stricter cut-off criterion of PR ≤ 10 ($\Delta\hat{g} = 0.23$ when compared to the PR ≤ 25 samples, Table 4), it failed to reach statistical significance. This was also the case when using the IQ discrepancy criterion[3] as a signature of clinically relevant dyscalculia, instead. On non-symbolic outcomes, where no significant heterogeneity was detected ($Q \sim \chi^2(9.88) = 9.88$, $p = .44$), moderator analyses revealed no significant influences either.

Table 3
Random effects models – group difference in mean RT and CDE.

Effect	k	$\hat{g}$	SE	t	CI 95%	Heterogeneity τ^2	I^2	Q
Mean RT								
Symbolic	44	0.75***	0.11	6.85	[0.51; 0.99]	0.10	61.14	31.10**
Non-sym.	48	0.24***	0.05	4.82	[0.13; 0.35]	<0.001	<0.001	9.88
CDE on RT								
Symbolic	20	−0.05	0.07	−0.69	[−0.21; 0.11]	0.02	30.31	13.06
Non-sym.	21	−0.12	0.08	−1.49	[−0.31; 0.07]	0.04	36.75	15.93

Note: RT: response time.
CDE: comparison distance effect.
$\hat{g}$: estimated average effect across all outcomes
*p < .05, **p < .01, ***p < .001.

Table 4
Meta-regression – mean RT symbolic.

Effect	Category	k	$\hat{g}$	SE	t	$p(t)$	CI 95%
Age	intercept b_0	44	0.39	0.47	0.83		[−0.81; 1.60]
	slope b_1		0.04	0.06	0.71	.51	[−0.11; 0.19]
Cut-off	10 < PR ≤ 25	43	0.60	0.11	5.74		[0.34; 0.86]
	PR ≤ 10		0.23	0.18	1.30	.25	[−0.22; 0.68]
Baseline RT*	no difference	23	0.80	0.34	2.35		[−0.31; 1.90]
	difference		0.31	0.38	0.83	.45	[−0.71; 1.33]
Intelligence*	no difference	39	0.86	0.24	3.66		[0.27; 1.45]
	difference		−0.09	0.27	−0.31	.76	[−0.71; 0.54]
Reading speed*	no difference	33	0.89	0.28	3.16		[0.09; 1.68]
	difference		−0.20	0.29	−0.67	.52	[−0.89; 0.50]
Math criterion*	unstandardized	44	0.91	0.19	4.73		[0.40; 1.42]
	standardized		−0.30	0.22	−1.32	.22	[−0.80; 0.21]

Note: see Table 3.
For all dummy-coded moderators (i.e. all except age), the estimate $\hat{g}$ in the second category indicates the increment or decrement of the effect compared to the reference category above (cf. Table 2).
* control variables.

[3] Only two of the studies included in this meta-analysis (Brankaer et al., 2014; Kuhn et al., 2013) reported explicitly that the IQ discrepancy criterion was set at individual level. For all the other studies, we could only compute a proxy at group level when group means of both math and IQ test results were available.

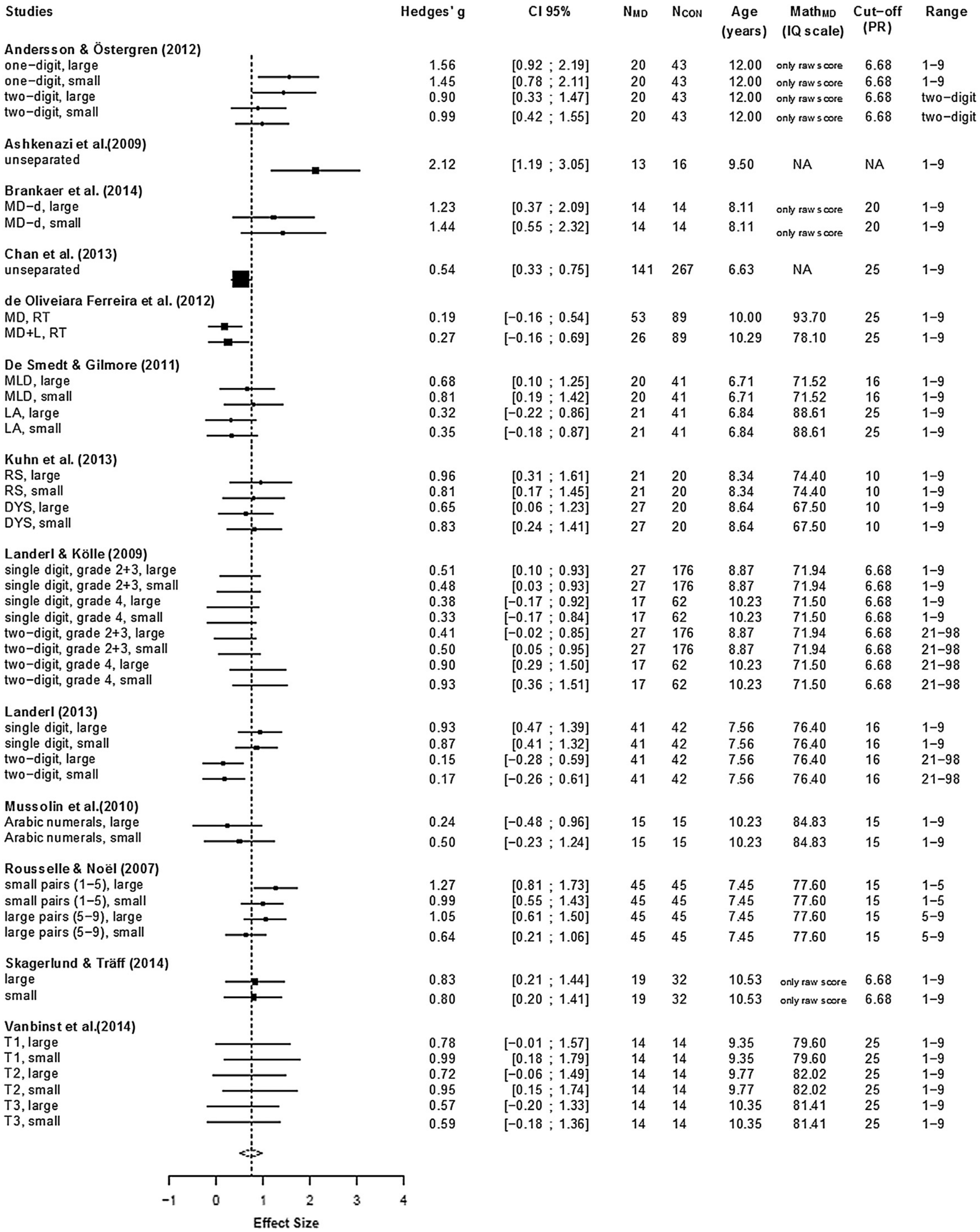

Fig. 2. Group difference in response times – symbolic.

Note: Sample names as used in the original studies: *MD-d*: IQ discrepant mathematical difficulties, *MD*: sample with math difficulties, *MD+L*: sample with math difficulties associated with language difficulties; *MLD*: children with mathematics learning disabilities, *LA*: children with low achievement; *RS*: sample with math difficulties not fulfilling the IQ discrepancy criterion, *DYS*: DD sample fulfilling the IQ discrepancy criterion.

Note that Skagerlund and Träff (2014) report group effects adjusted for various cognitive measures. As such adjusted estimates are (i) neither directly comparable to Hedges'*g* values from other studies and (ii) Fischer (2015) argues convincingly against the use of ANCOVA in this data, we calculated Hedges' g from raw means and standard deviations. Consequently, the findings summarized here are not in line with the results reported by Skagerlund and Träff (2014).

Studies	Hedges' g	CI 95%	N_{MD}	N_{CON}	Age (years)	$Math_{MD}$ (IQ scale)	Cut-off (PR)	Subitizing	#visual parameters controlled
Andersson & Östergren (2012)									
unseparated	0.17	[−0.40 ; 0.74]	20	43	12.00	NA	6.68	yes	1
Brankaer et al. (2014)									
MD−d, large	1.17	[0.32 ; 2.02]	14	14	8.11	only raw score	20	yes	3
MD−d, small	0.81	[0.03 ; 1.60]	14	14	8.11	only raw score	20	yes	3
de Oliveiara Ferreira et al. (2012)									
MD, RT	0.14	[−0.22 ; 0.50]	53	89	10.00	93.70	25	no	2
MD+L, RT	0.03	[−0.43 ; 0.49]	26	89	10.29	78.10	25	no	2
De Smedt & Gilmore (2011)									
MLD, large	0.17	[−0.39 ; 0.73]	20	41	6.71	71.52	16	yes	3
MLD, small	0.20	[−0.38 ; 0.79]	20	41	6.71	71.52	16	yes	3
LA, large	-0.06	[−0.60 ; 0.49]	21	41	6.84	88.61	25	yes	3
LA, small	0.07	[−0.48 ; 0.63]	21	41	6.84	88.61	25	yes	3
Defever et al. (2013) – Exp. 2									
comparison, large	0.58	[−0.04 ; 1.20]	21	21	11.19	only raw score	10	no	4
comparison, small	0.46	[−0.15 ; 1.08]	21	21	11.19	only raw score	10	no	4
Dinkel et al. (2013)									
unseparated	0.34	[−0.37 ; 1.04]	16	16	8.17	88.00	10	yes	2
Grond et al. (unpublished)									
large	0.37	[−0.39 ; 1.14]	14	13	13.43	only raw score	NA	no	3
small	-0.07	[−0.83 ; 0.69]	14	13	13.43	only raw score	NA	no	3
Heine et al. (2013)									
subitizing, large	0.23	[−0.40 ; 0.86]	20	20	8.23	76.60	5	yes	3
subitizing, small	0.42	[−0.21 ; 1.05]	20	20	8.23	76.60	5	yes	3
large number estimation, large	0.48	[−0.15 ; 1.12]	20	20	8.23	76.60	5	no	3
large number estimation, small	0.59	[−0.05 ; 1.22]	20	20	8.23	76.60	5	no	3
canonical patterns, large	0.35	[−0.28 ; 0.98]	20	20	8.23	76.60	5	yes	3
canonical patterns, small	0.53	[−0.10 ; 1.17]	20	20	8.23	76.60	5	yes	3
small−number estimation, large	0.26	[−0.37 ; 0.88]	20	20	8.23	76.60	5	yes	3
small−number estimation,small	-0.02	[−0.64 ; 0.61]	20	20	8.23	76.60	5	yes	3
Kucian et al. (2011)									
large	0.27	[−0.46 ; 0.99]	15	15	10.95	NA	NA	yes	1
small	0.42	[−0.31 ; 1.14]	15	15	10.95	NA	NA	yes	1
Landerl & Kölle (2009)									
grade 2+3, large	0.14	[−0.28 ; 0.57]	27	176	8.87	71.94	6.68	no	3
grade 2+3, small	0.02	[−0.38 ; 0.41]	27	176	8.87	71.94	6.68	no	3
grade 4, large	0.26	[−0.23 ; 0.75]	17	62	10.23	71.50	6.68	no	3
grade 4, small	0.06	[−0.44 ; 0.56]	17	62	10.23	71.50	6.68	no	3
Landerl (2013)									
large	0.37	[−0.06 ; 0.81]	41	42	7.56	76.40	16	no	2
small	0.32	[−0.11 ; 0.76]	41	42	7.56	76.40	16	no	2
Mussolin et al.(2010)									
canonical dot patterns, large	0.23	[−0.49 ; 0.95]	15	15	10.23	84.83	15	yes	1
canonical dot patterns, small	0.17	[−0.56 ; 0.89]	15	15	10.23	84.83	15	yes	1
noncanonical dot patterns, large	0.23	[−0.49 ; 0.96]	15	15	10.23	84.83	15	yes	1
noncanonical dot patterns, small	0.30	[−0.42 ; 1.03]	15	15	10.23	84.83	15	yes	1
random stick patterns, large	0.15	[−0.57 ; 0.87]	15	15	10.23	84.83	15	yes	3
random stick patterns, small	0.23	[−0.49 ; 0.95]	15	15	10.23	84.83	15	yes	3
Piazza et al. (2010)									
unseparated	-0.16	[−0.73 ; 0.41]	23	26	10.55	60.90	2.28	no	2
Rousselle & Noël (2007)									
ratio set, density condition, large	0.45	[0.03 ; 0.87]	45	45	7.45	77.60	15	no	3
ratio set, density condition, small	0.49	[0.06 ; 0.91]	45	45	7.45	77.60	15	no	3
ratio set, surface condition, large	0.38	[−0.04 ; 0.79]	45	45	7.45	77.60	15	no	3
ratio set, surface condition, small	0.18	[−0.24 ; 0.59]	45	45	7.45	77.60	15	no	3
Skagerlund & Träff (2014)									
unseparated	0.19	[−0.40 ; 0.78]	19	32	10.53	only raw score	6.68	no	2
Vanbinst et al.(2014)									
T1, large	0.38	[−0.38 ; 1.13]	14	14	9.35	79.60	25	yes	3
T1, small	0.44	[−0.31 ; 1.20]	14	14	9.35	79.60	25	yes	3
T2, large	0.24	[−0.51 ; 0.99]	14	14	9.77	82.02	25	yes	3
T2, small	-0.13	[−0.88 ; 0.62]	14	14	9.77	82.02	25	yes	3
T3, large	0.34	[−0.42 ; 1.09]	14	14	10.35	81.41	25	yes	3
T3, small	0.31	[−0.44 ; 1.06]	14	14	10.35	81.41	25	yes	3

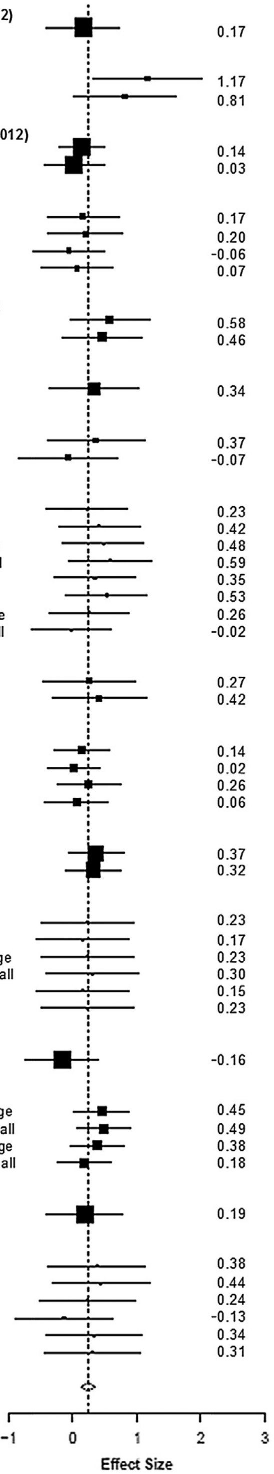

Fig. 3. Group difference in response times – non-symbolic. *Note*: see Fig. 2.

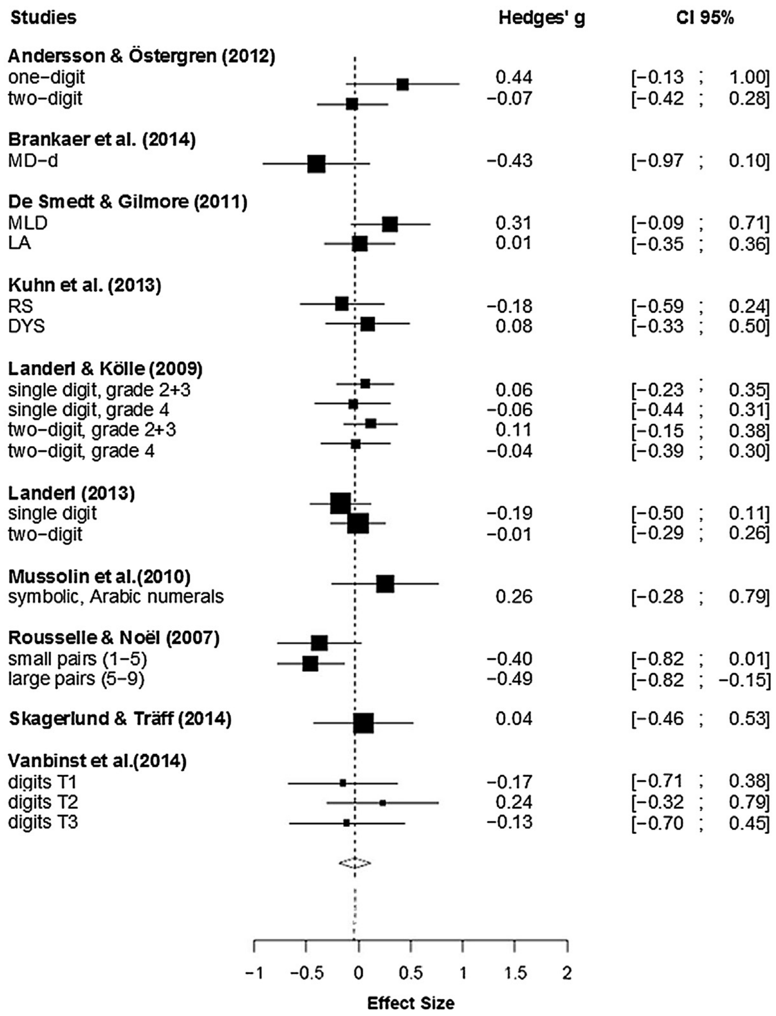

Fig. 4. Group difference in distance effects – symbolic. *Note*: see Fig. 2.

3.2. Group differences in the comparison distance effect

The average group differences on the CDE (Table 3) for both symbolic ($\hat{g} = -0.05$, Fig. 4) and non-symbolic stimuli ($\hat{g} = -0.12$, Fig. 5) did neither differ from zero ($p > .17$) nor from each other ($t(11.21) = 0.88$, $p = .39$). The impact of different values of ρ on $\hat{g}$, SE and τ^2 was negligible. There was no significant amount of heterogeneity across the studies' effects (Q statistics with *ps* > .10, Table 3). Likewise, there were no significant moderator effects.

3.3. Publication bias

This meta-analysis was a priori designed to minimize influences of publication bias by including only studies with group sized $N \geq 10$ (Morris, 2008), and by searching for unpublished data. To examine whether there were still clues of bias due to missing studies, funnel plots were generated to visually inspect asymmetries in the distribution of effects on symbolic and non-symbolic tasks (see Fig. A1 in the Appendix A). Statistically, the RVE version of Egger's regression test of the standard error's influence on the estimated effect size (Egger, Smith, Schneider, & Minder, 1997) did not corroborate asymmetry for any of the outcomes (all *ps* > .10) so that publication bias is unlikely.

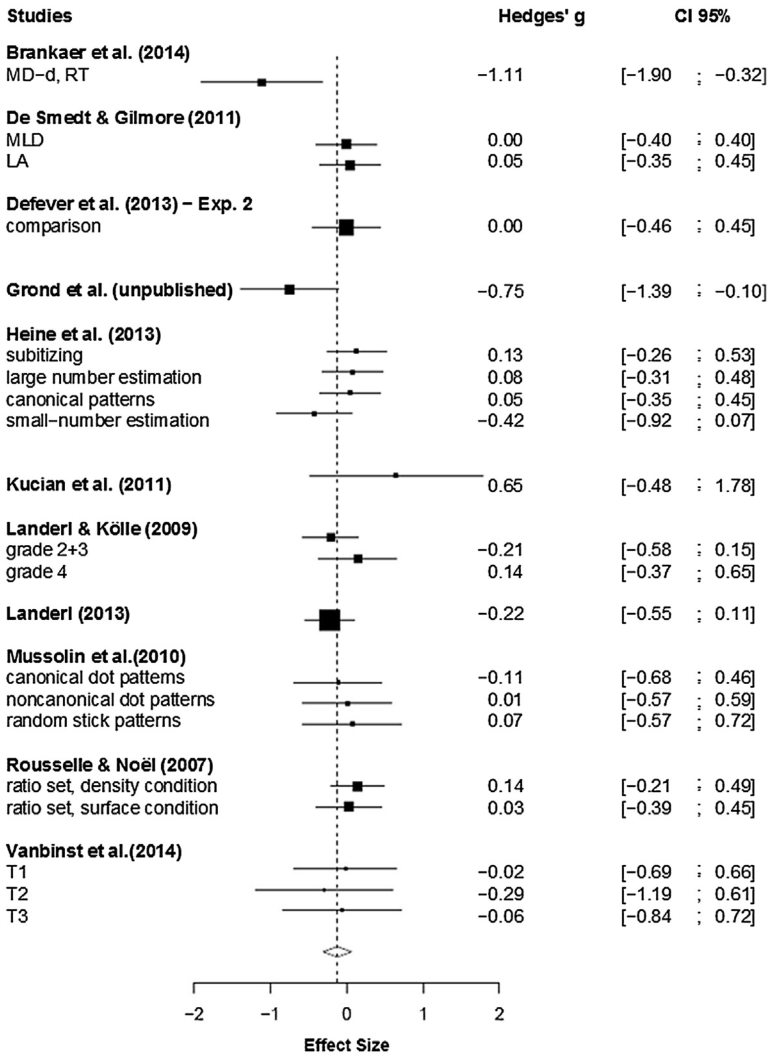

Fig. 5. Group difference in distance effects – non-symbolic. *Note*: see Fig. 2.

4. Discussion

This meta-analysis compared children with and without mathematical difficulties, i.e., low achievement up to dyscalculia, in their performance on symbolic and non-symbolic magnitude comparison tasks. The key results were that (a) children with MD displayed stronger impairments in their response speed on symbolic compared to non-symbolic tasks, (b) neither the sample characteristics *age* nor the *cut-off* set to define MD nor the task variables *number range*, the use of *stimuli in the subitizing range* nor the *number of visual task properties controlled for* moderated group differences and (c) on a meta-analytical level, no group differences were corroborated with respect to the comparison distance effect.

In line with several empirical studies (e.g., De Smedt & Gilmore, 2011; Landerl & Kölle, 2009; Rousselle & Noel, 2007; Olsson, Östergren, & Träff, 2016), our results support the finding that in children with MD, symbolic magnitude processing is more strongly impaired than non-symbolic processing. These results conform with the access-deficit hypothesis (Rousselle & Noël, 2007), which implies that impairments observed in children with MD arise from accessing numerical meaning from symbols (i.e., numbers) rather than from difficulties in the innate ANS per se. Consistent with this, longitudinal studies have shown that in predicting future mathematical performance in young children, symbolic rather than non-symbolic magnitude processing is crucial (Bartelet, Vaessen, Blomert, & Ansari, 2014; Sasanguie, Göbel, Moll, Smets, & Reynvoet, 2013). Further, the relation between non-symbolic magnitude processing and math achievement that has been found in some studies (e.g., Mazzocco et al., 2011) may be mediated by symbolic magnitude processing (van Marle, Chu, Li, & Geary, 2014) or more general skills like inhibitory control (Gilmore et al., 2013). In a recent study, Vanbinst, Ansari, Ghesquière and De Smedt (2016) even found that symbolic magnitude processing is as important for predicting arithmetic skills in elementary school as phonological awareness is for predicting future reading skills. The performance on symbolic magnitude comparison therefore appears well-suited as a core marker for predicting the development of arithmetic skills and especially for an early identification of MD. However, in order to use this knowledge within a diagnostic framework, future studies should focus on cognitive processes underlying the performance in symbolic magnitude comparison (e.g., number identification, cardinality knowledge, ordinality knowledge; Merkley & Ansari, 2016). In this regard, Sasanguie, Lyons, De Smedt and Reynvoet (2016) recently unraveled adults' performance on a symbolic comparison task and demonstrated that of several cognitive candidate processes, ordering ability was most predictive and fully mediated the relationship between digit comparison performance and arithmetic.

Interestingly, this meta-analysis suggests that the difference in symbolic and non-symbolic magnitude comparison between children with MD and controls is not moderated by age, i.e., follows a stable developmental trajectory: Children with MD have difficulties in processing symbolic numbers and, to a lesser degree, non-symbolic numerosities from the start, and this difference remains stable across childhood. An important implication of this finding is that the difficulties of children with MD appear to be mainly related to the acquisition of a culturally transmitted rather than an innate skill: the fluent processing and understanding of numerical symbols. This skill can be trained. For example, early interventions that focus on understanding and processing numbers have been shown to result in substantial and broad improvements in early math skills of preschoolers (Ramani, Siegler, & Hitti, 2012; Sella, Tressoldi, Lucangeli, & Zorzi, 2016; Maertens, De Smedt, Sasanguie, Elen, & Reynvoet, 2016). However, although this meta-analysis showed that the impairments of children with MD are significantly smaller on non-symbolic – when compared to symbolic– outcomes, they are not negligible and it is possible that other indices based on accuracy (such as the Weber fraction) would have been even more sensitive to non-symbolic effects than response time. This is interesting in the context of a recent study by Wang, Odic, Halberda and Feigenson (2016) who trained preschoolers on non-symbolic magnitude processing and observed short-term transfer to (symbolic) arithmetic skills. However, the causal evidence of such training studies is discussed (Merkley, Matejko, & Ansari, 2017), such that additional research is needed to clarify the causal pathways between ANS and symbolic performance up to arithmetic improvements.

Meta-analytical results necessarily reflect a highly aggregate level that deals with the common denominator of comparable, but still different studies. While meta-analyses help to keep an eye on the big picture, they do not inform us about important effects at the (inter)individual level. For example, children tend to show different profiles of the association between their non-symbolic and symbolic magnitude processing competences (Chew, Forte, & Reeve, 2016). Accordingly, non-symbolic processing is likely to only impact the mathematical development of some subgroups, so that the relatively weaker effect that turns out at group level does not generalize to each cluster of children with MD up to dyscalculia.

We also found that neither reading skills nor IQ, which were treated as control variables, explained the group differences in magnitude comparison speed. This is in line with a growing body of studies finding that reading skills do not affect tasks without verbal content (Raddatz, Kuhn, Holling, Moll, & Dobel, 2016). It further conforms to results showing that when MD or even dyscalculia is present, IQ does not substantially affect numerical processing (Brankaer et al., 2014).

Regarding stimulus properties, several studies have shown that non-symbolic magnitudes are processed faster when visual and numerical cues are compatible, suggesting that visually controlled tasks require inhibition of the non-meaningful visual cues (e.g., Clayton & Gilmore, 2015). Further, a replication study showed that when visual cues are controlled in non-symbolic tasks, no relationship of performance on this task and exact number knowledge could be found in 3–6 year old children (Negen & Sarnecka, 2014). In the original study, by contrast, a relationship was observed in the case that visual cues were not controlled (Mussolin, Nys, Leybaert, & Content, 2012). In our meta-analysis, the number of visual parameters (out of stimulus diameter, density, total surface and convex hull) controlled for was not a significant moderator of the relative impairments in children with MD. A closer look at the different studies reveals that both the definitions of the visual parameters (especially in the distinction of convex hull and total surface) and the technical scripts used for their controls vary. Consequently, this effect might not (yet) be assessable on a meta-analytical level, but requires experimental studies and a consensus in operationalization before.

The comparison distance effect has been described by several authors as a measure of the precision of mental number representation. Smaller distance effects have been related to a more precise number representation and were consequently regarded as indicators of better mathematical skills (although response inhibition rather than representational overlap has been suggested as an alternative explanation; van Opstal et al., 2008). Some studies have indeed found a relationship between individual distance effects and mathematical achievement (e.g., Holloway & Ansari, 2009; Mussolin et al., 2010), whereas others have not (e.g., Lyons et al., 2015; Rousselle & Noel, 2007). In this meta-analysis, we did not find qualitative differences in response speed based distance effects between children with and without MD. This finding, applying to both symbolic and non-symbolic tasks, has two important implications: First, it undercuts the idea that the distance effect should be used for diagnostic purposes. Second, it does not support the assumption that children with MD have less precise numerical representations as a result of a noisier mental number line. This is reflected in the recent debate on whether the innate ANS System in fact exists at all. Indeed, recently Leibovich, Katzin, Harel and Henik (2016) reviewed studies investigating the ANS and concluded that the ability to extract discrete number originates from the ability to process continuous magnitudes, thus rather suggesting an Approximate Magnitude System (AMS) than an Approximate Number System. Sasanguie and Reynvoet (2016) commented on this review by emphasizing the implications of the AMS instead of an ANS for the symbol grounding problem: If the ANS does exist, how are then numerical symbols learned, how do they acquire their meaning? By alternatively explaining the arguments previously put forward in favor of the ANS-mapping hypothesis, Reynvoet and Sasanguie (2016) argued for an account supporting the development of an exact symbolic system next to the ANS/AMS.

Following Carey (2009), these authors suggest that, in a first step, numerical symbols are mapped on the object tracking system, a system that allows to keep track up to four items (or for an alternative up to 10 fingers, see Siegler, 2016). In a second phase, knowledge about the counting list may be used to infer critical principles of the symbolic number system, such as the principle that numbers later in the counting list are larger (Davidson, Eng, & Barner, 2012). As a result, gradually, symbolic numbers (in terms of digits) are primarily represented through (order) associations with other symbolic numbers (number words) in a separate semantic network of symbolic numbers (Krajci, Lengyel, & Kojouharova, 2016; Sasanguie et al., 2017). Together, the absence of qualitative differences in response speed-based distance effects between children with and without MD and the recent alternative account for symbol learning suggest that it is unlikely that mathematical difficulties can be explained by less precise numerical representations.

In this meta-analysis, we chose responses times as a criterion that is suitable to express commonalities and differences between symbolic and non-symbolic magnitude comparison performance. By contrast, accuracy information is characterized by considerable limitations: While Weber fractions are the state-of-the-art measure for non-symbolic, but not symbolic, tasks, simple accuracy data (error rate or 1-error rate) are prone to ceiling effects and very low variance, which in turn tends to inflate the effect sizes (Hedges'g) based thereon. However, speed is just one expression of performance, of course, and ignoring accuracy information would leave a blind spot to the picture of evidence, e.g., regarding the speed-accuracy trade-off. That is why we provide an overview of the accuracy data that was accessible for 13 of the 19 studies included in this meta-analysis (cf. supplementary material). In summary, all the overall effects on accuracy (symbolic and non-symbolic group difference, CDE) are highly similar to the ones reported based on the response time data. Crucially, there are no discordances between the two measures in any of the included outcomes, i.e., no cases where the MD group appeared to be slower but more accurate than the typically achieving control group or vice versa. Thus, we find no qualitative difference in speed-accuracy trade-off between children with and without MD on a meta-analytical level.

5. Conclusion

To conclude, we found that irrespective of age, IQ and reading skills, response times for symbolic magnitude comparison tasks were better able to discriminate between children with and without MD than for non-symbolic tasks, supporting the access-deficit hypothesis. Further, the CDE did not differ substantially between children with MD and typically achieving controls. This result, together with the finding that the distance effect is generally no reliable indicator of mathematical skills in representative samples (Lyons et al., 2015), supports the notion that the diagnostic and theoretical value of the CDE is uncertain. Rather, speed of symbolic magnitude processing proves to be a substantial indicator of mathematical difficulties and should in turn be a core aspect of interventions designed to support children who are impaired in their mathematical development.

Appendix A.

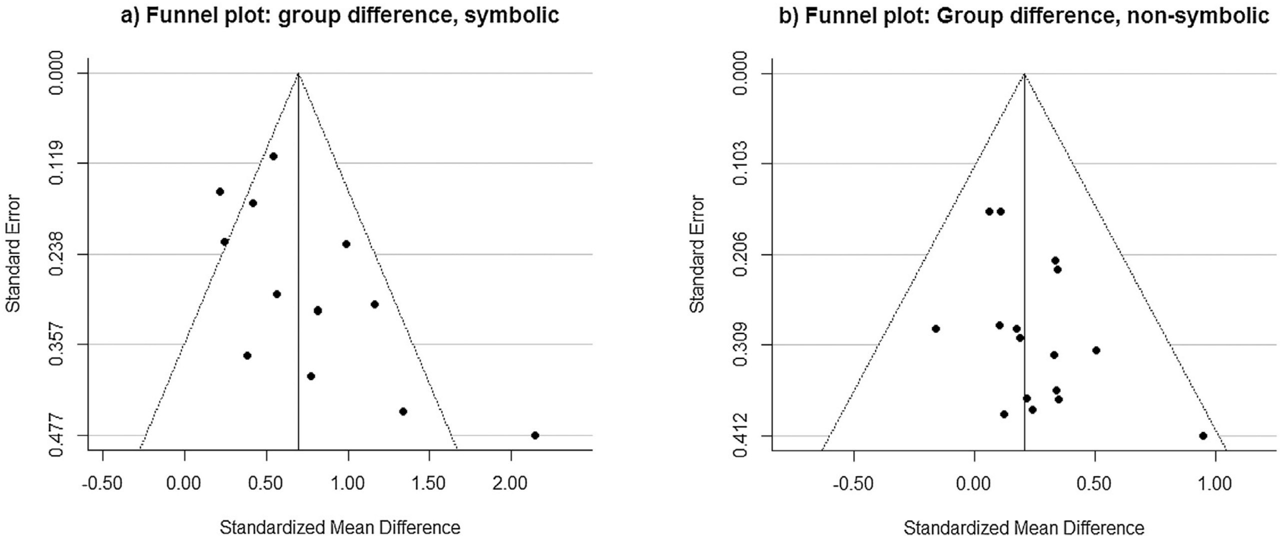

Fig. A1. Funnel plots for (a) group difference in RT on symbolic and (b) non-symbolic magnitude comparison tasks. Funnel plots display the distribution of outcomes included in a meta-analysis and serve as a visual tool to check for publication bias. Each effect size is plotted against its sampling error, which is a function of sample size. The absence of publication bias would entail a symmetric distribution of effect sizes around the mean and a variance of estimated effects that decreases with the standard error (i.e. with increasing sample size). Such funnel plots are biased when several effects per study are treated as if they were independent. Therefore, effects were integrated to one mean per study, weighted by the number of trials per task or subjects per group before generating the plots. Here, only the plot for the symbolic outcomes includes few effect sizes outside the expected funnel. However, this asymmetry is not corroborated by the RVE version of Egger's regression test of the standard error's influence on the estimated effect size (Egger, Smith, Schneider, & Minder, 1997; all *ps* > .10).

Appendix B. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.ridd.2017.03.003.

References[4]

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Arlington, VA: American Psychiatric Publishing. http://dx.doi.org/10.1176/appi.books.9780890425596.

Ancker, J. S., & Kaufman, D. (2007). Rethinking health numeracy: A multidisciplinary literature review. *Journal of the American Medical Informatics Association*, *14*, 713–721. http://dx.doi.org/10.1197/jamia.M2464

*Andersson, U., & Östergren, R. (2012). Number magnitude processing and basic cognitive functions in children with mathematical learning disabilities. *Learning and Individual Differences*, *22*, 701–714. http://dx.doi.org/10.1016/j.lindif.2012.05.004

Ansari, D. (2008). Effects of development and enculturation on number representation in the brain. *Nature Reviews Neuroscience*, *9*, 278–291. http://dx.doi.org/10.1038/nrn2334

*Ashkenazi, S., Mark-Zigdon, N., & Henik, A. (2009). Numerical distance effect in developmental dyscalculia. *Cognitive Development*, *24*, 387–400. http://dx.doi.org/10.1016/j.cogdev.2009.09.006

Bartelet, D., Vaessen, A., Blomert, L., & Ansari, D. (2014). What basic number processing measures in kindergarten explain unique variability in first-grade arithmetic proficiency? *Journal of Experimental Child Psychology*, *117*, 12–28. http://dx.doi.org/10.1016/j.jecp.2013.08.010

Bonett, D. G. (2009). Meta-analytic interval estimation for standardized and unstandardized mean differences. *Psychological Methods*, *14*, 225–238. http://dx.doi.org/10.1037/a0016619

*Brankaer, C., Ghesquière, P., & De Smedt, B. (2014). Numerical magnitude processing deficits in children with mathematical difficulties are independent of intelligence. *Research in Developmental Disabilities*, *35*, 2603–2613. http://dx.doi.org/10.1016/j.ridd.2014.06.022

Butterworth, B. (2005). The development of arithmetical abilities. *Journal of Child Psychology and Psychiatry*, *46*, 3–18. http://dx.doi.org/10.1111/j.1469-7610.2005.00374.x

Carey, S. (2009). Where our number concepts come from. *Journal of Philosophy*, *106*, 220–254. http://dx.doi.org/jphil2009106418

*Chan, W. W. L., Au, T. K., & Tang, J. (2013). Developmental dyscalculia and low numeracy in Chinese children. *Research in Developmental Disabilities*, *34*, 1613–1622. http://dx.doi.org/10.1016/j.ridd.2013.01.030

Chen, Q., & Li, J. (2014). Association between individual differences in non-symbolic number acuity and math performance: A meta-analysis. *Acta Psychologica*, *148*, 163–172. http://dx.doi.org/10.1016/j.actpsy.2014.01.016

Chew, C. S., Forte, J. D., & Reeve, R. A. (2016). Cognitive factors affecting children's nonsymbolic and symbolic magnitude judgment abilities: A latent profile analysis. *Journal of Experimental Child Psychology*, *152*, 173–191. http://dx.doi.org/10.1016/j.jecp.2016.07.001

Clayton, S., & Gilmore, C. (2015). Inhibition in dot comparison tasks. *ZDM*, *47*, 759–770. http://dx.doi.org/10.1007/s11858-014-0655-2

Davidson, K., Eng, K., & Barner, D. (2012). Does learning to count involve a semantic induction ?. *Cognition*, *123*, 162–173. http://dx.doi.org/10.1016/j.cognition.2011.12.013

[4] Asterisk denoting that articles were part of the meta-analysis.

*de Oliveira Ferreira, F., Wood, G., Pinheiro-Chagas, P., Lonnemann, J., Krinzinger, H., Willmes, K., & Haase, V. G. (2012). Explaining school mathematics performance from symbolic and nonsymbolic magnitude processing: Similarities and differences between typical and low-achieving children. *Psychology & Neuroscience*, *5*, 37–46. http://dx.doi.org/10.3922/j.psns.2012.1.06
*De Smedt, B., & Gilmore, C. K. (2011). Defective number module or impaired access? numerical magnitude processing in first graders with mathematical difficulties. *Journal of Experimental Child Psychology*, *108*, 278–292. http://dx.doi.org/10.1016/j.jecp.2010.09.003
De Smedt, B., Noël, M.-P., Gilmore, C., & Ansari, D. (2013). How do symbolic and non-symbolic numerical magnitude processing skills relate to individual differences in children's mathematical skills? A review of evidence from brain and behavior. *Trends in Neuroscience and Education*, *2*, 48–55. http://dx.doi.org/10.1016/j.tine.2013.06.001
*Defever, E., Reynvoet, B., & Gebuis, T. (2013). Task- and age-dependent effects of visual stimulus properties on children's explicit numerosity judgments. *Journal of Experimental Child Psychology*, *116*, 216–233. http://dx.doi.org/10.1016/j.jecp.2013.04.006
Dehaene, S. (1997). *The number sense: How the mind creates mathematics*. New York: Oxford University press.
*Dinkel, P. J., Willmes, K., Krinzinger, H., Konrad, K., & Koten, J. W., Jr. (2013). Diagnosing developmental dyscalculia on the basis of reliable single case FMRI methods: promises and limitations. *PLoS ONE*, *8*, e83722. http://dx.doi.org/10.1371/journal.pone.0083722
Dowker, A. (2009). *What Works for Children with Mathematical Difficulties?* London: DfES.
Egger, M., Smith, G. D., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *Bmj*, *315*(7109), 629–634. http://dx.doi.org/10.1136/bmj.315.7109.629
Fazio, L. K., Bailey, D. H., Thompson, C. A., & Siegler, R. S. (2014). Relations of different types of numerical magnitude representations to each other and to mathematics achievement. *Journal of Experimental Child Psychology*, *123*, 53–72. http://dx.doi.org/10.1016/j.jecp.2014.01.013
Feigenson, L., Dehaene, S., & Spelke, E. (2004). Core systems of number. *Trends in Cognitive Sciences*, *8*, 307–314. http://dx.doi.org/10.1016/j.tics.2004.05.002
Fischer, J. P. (2015). Commentary: Development of magnitude processing in children with developmental dyscalculia: space, time, and number. *Frontiers in Psychology*, *6*. http://dx.doi.org/10.3389/fpsyg.2015.00804
Fisher, Z., & Tipton, E. (2015). robumeta: Robust variance meta regression. R package version 1.6. https://cran.r-project.org/web/packages/robumeta/.
Geary, D. C. (2013). Early foundations for mathematics learning and their relations to learning disabilities. *Current Directions in Psychological Science*, *22*, 23–27. http://dx.doi.org/10.1177/0963721412469398
Gebuis, T., & Reynvoet, B. (2012). The interplay between nonsymbolic number and its continuous visual properties. *Journal of Experimental Psychology: General*, *141*, 642–648. http://dx.doi.org/10.1037/a0026218
Gilmore, C., Attridge, N., Clayton, S., Cragg, L., Johnson, S., Marlow, N., Simms, V., & Inglis, M. (2013). Individual differences in inhibitory control, not non-verbal number acuity, correlate with mathematics achievement. *PLoS ONE*, *8*, e67374. http://dx.doi.org/10.1371/journal.pone.0067374
*Grond, U., Kucian, K., O'Gorman, R., Martin, E., & von Aster, M. Is the general magnitude system impaired in developmental dyscalculia? Behavioural evaluation of an fMRI paradigm. 24.-26.5.2012, EARLI, SIG-22-Meeting Neuroscience and Education, London.
Halberda, J., Mazzocco, M. M., & Feigenson, L. (2008). Individual differences in non-verbal number acuity correlate with maths achievement. *Nature*, *455*, 665–668. http://dx.doi.org/10.1038/nature07246
Hedges, L. V. (1981). Distribution theory for glass's estimator of effect size and related estimators. *Journal of Educational Statistics*, *6*, 107–128. http://dx.doi.org/10.2307/1164588
Hedges, L. V., Tipton, E., & Johnson, M. C. (2010). Robust variance estimation in meta-regression with dependent effect size estimates. *Research Synthesis Methods*, *1*, 39–65. http://dx.doi.org/10.1002/jrsm.5
Heine, A., Wißmann, J., Tamm, S., De Smedt, B., Schneider, M., Stern, E., & Jacobs, A. M. (2013). An electrophysiological investigation of non-symbolic magnitude processing: Numerical distance effects in children with and without mathematical learning disabilities. *Cortex A: Journal Devoted to the Study of the Nervous System and Behavior*, *49*, 2162–2177. http://dx.doi.org/10.1016/j.cortex.2012.11.009
Higgins, J. P., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *Bmj*, *327*(7414), 557–560. http://dx.doi.org/10.1136/bmj.327.7414.557
Holloway, I. D., & Ansari, D. (2009). Mapping numerical magnitudes onto symbols: The numerical distance effect and individual differences in children's mathematics achievement. *Journal of Experimental Child Psychology*, *103*, 17–29. http://dx.doi.org/10.1016/j.jecp.2008.04.001
Krajci, A., Lengyel, G., & Kojouharova, P. (2016). The source of the symbolic numerical distance and size effects. *Frontiers in Psychology*, *7*, 1795. http://dx.doi.org/10.3389/fpsyg.2016.01795
Kucian, K., Loenneker, T., Martin, E., & von Aster, M. (2011). Non-symbolic numerical distance effect in children with and without developmental dyscalculia: A parametric fMRI study. *Developmental Neuropsychology*, *36*, 741–762. http://dx.doi.org/10.1080/87565641.2010.549867
Kuhn, J.-T., Raddatz, J., Holling, H., & Dobel, C. (2013). Dyskalkulie vs. Rechenschwäche: Basisnumerische Verarbeitung in der Grundschule. *Lernen und Lernstörungen*, *2*, 229–247. http://dx.doi.org/10.1024/2235-0977/a000044
Landerl, K. (2013). Development of numerical processing in children with typical and dyscalculic arithmetic skills - A longitudinal study. *Frontiers in Psychology*, *4*. http://dx.doi.org/10.3389/fpsyg.2013.00459
Landerl, K., Fussenegger, B., Moll, K., & Willburger, E. (2009). Dyslexia and dyscalculia: Two learning disorders with different cognitive profiles. *Journal of Experimental Child Psychology*, *103*, 309–324. http://dx.doi.org/10.1016/j.jecp.2009.03.006
Landerl, K., & Kölle, C. (2009). Typical and atypical development of basic numerical skills in elementary school. *Journal of Experimental Child Psychology*, *103*, 546–565. http://dx.doi.org/10.1016/j.jecp.2008.12.006
Leibovich, T., Katzin, N., Harel, M., & Henik, A. (2016). From 'sense of number' to 'sense of magnitude'?The role of continuous magnitudes in numerical cognition. *Behavioral and Brain Sciences*, 1–62. http://dx.doi.org/10.1017/S0140525X16000960
Lyons, I. M., Nuerk, H.-C., & Ansari, D. (2015). Rethinking the implications of numerical ratio effects for understanding the development of representational precision and numerical processing across formats. *Journal of Experimental Psychology: General*, *144*, 1021–1035. http://dx.doi.org/10.1162/jocn.2009.21124
Maertens, B., De Smedt, B., Sasanguie, D., Elen, J., & Reynvoet, B. (2016). Enhancing arithmetic in preschoolers with comparison or number line estimation training: does it matter? *Learning and Instruction*, *46*, 1–11. http://dx.doi.org/10.1016/j.learninstruc.2016.08.004
Mazzocco, M. M. M., Feigenson, L., & Halberda, J. (2011). Impaired acuity of the approximate number system underlies mathematical learning disability (dyscalculia). *Child Development*, *82*, 1224–1237. http://dx.doi.org/10.1111/j.1467-8624.2011.01608.x
Merkley, R., & Ansari, D. (2016). Why numerical symbols count in the development of mathematical skills: evidence from brain and behavior. *Current Opinion in Behavioral Sciences*, *10*, 14–20. http://dx.doi.org/10.1016/j.cobeha.2016.04.006
Merkley, R., Matejko, A. A., & Ansari, D. (2017). Strong causal claims require strong evidence: A commentary on Wang and colleagues. *Journal of Experimental Child Psychology*, *153*, 163–167. http://dx.doi.org/10.1016/j.jecp.2016.07.008
Morris, S. B. (2008). Estimating effect sizes from pretest-posttest-control group designs. *Organizational Research Methods*, *11*, 364–386. http://dx.doi.org/10.1177/1094428106291059
Mundy, E., & Gilmore, C. K. (2009). Children's mapping between symbolic and nonsymbolic representations of number. *Journal of Experimental Child Psychology*, *103*, 490–502. http://dx.doi.org/10.1016/j.jecp.2009.02.003
Mussolin, C., Mejias, S., & Noël, M. (2010). Symbolic and nonsymbolic number comparison in children with and without dyscalculia. *Cognition*, *115*, 10–25. http://dx.doi.org/10.1016/j.cognition.2009.10.006
Mussolin, C., Nys, J., Leybaert, J., & Content, A. (2012). Relationships between approximate number system acuity and early symbolic number abilities. *Trends in Neuroscience and Education*, *1*, 21–31. http://dx.doi.org/10.1016/j.tine.2012.09.003
Negen, J., & Sarnecka, B. W. (2014). Is there really a link between exact-number knowledge and approximate number system acuity in young children? *British Journal of Developmental Psychology*, *33*, 92–105. http://dx.doi.org/10.1111/bjdp.12071

Noël, M., & Rousselle, L. (2011). Developmental changes in the profiles of dyscalculia: An explanation based on a double exact-and-approximate number representation model. *Frontiers in Human Neuroscience*, *5*, 165. http://dx.doi.org/10.3389/fnhum.2011.00165

Olsson, L., Östergren, R., & Träff, U. (2016). Developmental dyscalculia: A deficit in the approximate number system or an access deficit? *Cognitive Development*, *39*, 154–167. http://dx.doi.org/10.1016/j.cogdev.2016.04.006

Parsons, S., & Bynner, J. (2005). *Does numeracy matter more?* London: National Research and Development, Centre for Adult Literacy and Numeracy.

Piazza, M., Facoetti, A., Trussardi, A. N., Berteletti, I., Conte, S., Lucangeli, D., Dehaene, S., & Zorzi, M. (2010). Developmental trajectory of number acuity reveals a severe impairment in developmental dyscalculia. *Cognition*, *116*, 33–41. http://dx.doi.org/10.1016/j.cognition.2010.03.012

Price, G. R., & Ansari, D. (2013). Dyscalculia: Characteristics, causes, and treatments. *Numeracy*, *6*(1), 2. http://dx.doi.org/10.5038/1936-4660.6.1.2

Raddatz, J., Kuhn, J. T., Holling, H., Moll, K., & Dobel, C. (2016). Comorbidity of arithmetic and reading disorder: Basic number processing and calculation in children with learning impairments. *Journal of Learning Disabilities*,. 0022219415620899.

Ramani, G. B., Siegler, R. S., & Hitti, A. (2012). Taking it to the classroom: Number board games as a small group learning activity. *Journal of Educational Psychology*, *104*, 661. http://dx.doi.org/10.1037/a0028995

Reigosa, V., Sosa, M. V., Butterworth, B., Torres, P., Santos, E., Suárez, R., . . . & Hernández, D. (2008). 48. Large-scale prevalence studies of learning disabilities in Cuban school-children population. *Clinical Neurophysiology*, *119*(9), e111. http://dx.doi.org/10.1016/j.clinph.2008.04.064

Reynvoet, B., & Sasanguie, D. (2016). The Symbol Grounding Problem Revisited: A Thorough Evaluation of the ANS Mapping Account and the Proposal of an Alternative Account Based on Symbol-Symbol associations. *Frontiers in Psychology*, 7. http://dx.doi.org/10.3389/fpsyg.2016.01581, art.nr. 1581

Ritchie, S. J., & Bates, T. C. (2013). Enduring links from childhood mathematics and reading achievement to adult socioeconomic status. *Psychological Science*, *24*, 1301–1308. http://dx.doi.org/10.1177/0956797612466268

Rousselle, L., & Noël, M. (2007). Basic numerical skills in children with mathematics learning disabilities: A comparison of symbolic vs non-symbolic number magnitude processing. *Cognition*, *102*, 361–395. http://dx.doi.org/10.1016/j.cognition.2006.01.005

Rubinsten, O., & Henik, A. (2009). Developmental dyscalculia: Heterogeneity may not mean different mechanisms. *Trends in Cognitive Sciences*, *13*, 92–99. http://dx.doi.org/10.1016/j.tics.2008.11.002

Sasanguie, D., De Smedt, B., & Reynvoet, B. (2017). Evidence for distinct magnitude systems for symbolic and non-symbolic number. *Psychological Research*, *81*, 231–242. http://dx.doi.org/10.1007/s00426-015-0734-1

Sasanguie, D., Göbel, S. M., Moll, K., Smets, K., & Reynvoet, B. (2013). Approximate number sense, symbolic number processing, or number?space mappings: What underlies mathematics achievement? *Journal of Experimental Child Psychology*, *114*, 418–431. http://dx.doi.org/10.1016/j.jecp.2012.10.012

Sasanguie, D., Lyons, I.M., De Smedt, B., Reynvoet, B. (2016). Unpacking symbolic number comparison and its relation with arithmetic in adults. Manuscript under revision.

Sasanguie, D., & Reynvoet, B. (2016). A 'sense of magnitude' requires a new alternative for learning numerical symbols. Commentary on "From 'sense of number'to 'sense of magnitude'? The role of continuous magnitudes in numerical cognition" by Leibovich, Katzin, Harel, & Henik (2016)". *Behavioral and Brain Sciences*.

Schneider, M., Beeres, K., Coban, L., Merz, S., Schmidt, S. S., Stricker, J., & De Smedt, B. (2016). Associations of non-symbolic and symbolic numerical magnitude processing with mathematical competence: a meta-analysis. *Developmental Science*, http://dx.doi.org/10.1111/desc.12372

Sella, F., Tressoldi, P., Lucangeli, D., & Zorzi, M. (2016). Training numerical skills with the adaptive videogame "The Number Race": A randomized controlled trial on preschoolers. *Trends in Neuroscience and Education*, *5*, 20–29. http://dx.doi.org/10.1016/j.tine.2016.02.002

Siegler, R. S. (2016). Magnitude knowledge: The common core of numerical development. *Developmental Science*, *19*, 341–361. http://dx.doi.org/10.1111/desc.12395

Skagerlund, K., & Träff, U. (2014). Development of magnitude processing in children with developmental dyscalculia: space, time, and number. *Frontiers in Psychology*, *5*, 675. http://dx.doi.org/10.3389/fpsyg.2014.00675

Swanson, H. L., & Jerman, O. (2006). Math disabilities: A selective meta-analysis of the literature. *Review of Educational Research*, *76*, 249–274. http://dx.doi.org/10.3102/00346543076002249

Tanner-Smith, E. E., & Tipton, E. (2014). Robust variance estimation with dependent effect sizes: practical considerations including a software tutorial in Stata and SPSS. *Research Synthesis Methods*, *5*, 13–30. http://dx.doi.org/10.1002/jrsm.1091

Tipton, E. (2015). Small sample adjustments for robust variance estimation with meta-regression. *Psychological Methods*, *20*, 375. http://dx.doi.org/10.1037/met0000011

Vanbinst, K., Ansari, D., Ghesquière, P., & De Smedt, B. (2016). Symbolic numerical magnitude processing is as important to arithmetic as phonological awareness is to reading. *PloS one*, *11*, e0151045. http://dx.doi.org/10.1371/journal.pone.0151045

Vanbinst, K., Ghesquière, P., & De Smedt, B. (2014). Arithmetic strategy development and its domain-specific and domain-general cognitive correlates: A longitudinal study in children with persistent mathematical learning difficulties. *Research in Developmental Disabilities*, *35*, 3001–3013. http://dx.doi.org/10.1016/j.ridd.2014.06.023

van Marle, K., Chu, F. W., Li, Y., & Geary, D. C. (2014). Acuity of the approximate number system and preschoolers' quantitative development. *Developmental Science*, *17*, 492–505. http://dx.doi.org/10.1111/desc.12143

van Opstal, F., Gevers, W., de Moor, W., & Verguts, T. (2008). Dissecting the symbolic distance effect: Comparison and priming effects in numerical and nonnumerical orders. *Psychonomic Bulletin & Review*, *15*, 419–425. http://dx.doi.org/10.3758/PBR.15.2.419

Viechtbauer, W. (2010). Conducting meta-analysis in R with the metafor package. *Journal of Statistical Software*, *36*, 1–48. http://dx.doi.org/10.18637/jss.v036.i03

Wang, J. J., Odic, D., Halberda, J., & Feigenson, L. (2016). Changing the precision of preschoolers' approximate number system representations changes their symbolic math performance. *Journal of Experimental Child Psychology*, *147*, 82–99. http://dx.doi.org/10.1016/j.jecp.2016.03.002

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