Research Article

Severe Developmental Dyscalculia Is Characterized by Core Deficits in Both Symbolic and Nonsymbolic Number Sense

Gisella Decarli[1]⬤, Francesco Sella[2], Silvia Lanfranchi[3], Giulia Gerotto[4], Silvia Gerola[5], Giuseppe Cossu[5], and Marco Zorzi[4,6]⬤
[1]Integrative Neuroscience and Cognition Center, Université Paris Cité, CNRS; [2]Centre for Mathematical Cognition, Loughborough University; [3]Department of Developmental Psychology and Socialisation, University of Padova; [4]Department of General Psychology, University of Padova; [5]Children's Cognitive Neurorehabilitation Unit, Centro Medico di Foniatria, Padova; and [6]IRCCS San Camillo Hospital, Venice-Lido, Italy

Abstract

A long-standing debate concerns whether developmental dyscalculia is characterized by core deficits in processing nonsymbolic or symbolic numerical information as well as the role of domain-general difficulties. Heterogeneity in recruitment and diagnostic criteria make it difficult to disentangle this issue. Here, we selected children (n = 58) with severely compromised mathematical skills (2 SD below average) but average domain-general skills from a large sample referred for clinical assessment of learning disabilities. From the same sample, we selected a control group of children (n = 42) matched for IQ, age, and visuospatial memory but with average mathematical skills. Children with dyscalculia showed deficits in both symbolic and nonsymbolic number sense assessed with simple computerized tasks. Performance in the digit-comparison task and the numerosity match-to-sample task reliably separated children with developmental dyscalculia from controls in cross-validated logistic regression (area under the curve = .84). These results support a number-sense-deficit theory and highlight basic numerical abilities that could be targeted for early identification of at-risk children as well as for intervention.

Developmental dyscalculia was originally described as a disorder in mathematical abilities without a deficit in general mental abilities (Kosc, 1974). More recently, developmental dyscalculia or mathematical learning disability (MLD) has been referred to as a specific learning disability characterized by a deficit in numerical and mathematical skills (for a review, see Butterworth et al., 2011) that cannot be attributed to a lack of learning opportunities, inadequate education or environmental disadvantage, intellectual disabilities, global development delay, hearing/vision/motor disorders, or neurological deficits, according to the fifth edition of the *Diagnostic and Statistical Manual of Mental Disorders* (*DSM-5*; American Psychiatric Association, 2013).

Although low mathematical achievement can be identified using standardized tests, mathematical skills are heterogeneous and rely on multiple domain-specific numerical (both verbal and nonverbal) as well as domain-general cognitive processes (e.g., working memory). Multiple "core-deficit" hypotheses (for a critical discussion, see Astle & Fletcher-Watson, 2020) have attempted to identify a single and distinct neurocognitive mechanism as the distal cause of mathematical

Corresponding Author:
Marco Zorzi, University of Padova, Department of General Psychology and Padova Neuroscience Center
Email: marco.zorzi@unipd.it

disability. Accordingly, developmental dyscalculia has been related to a variety of deficits, both domain specific and domain general. Recent domain-specific accounts have focused on a core number-sense deficit that might be traced back to impaired numerosity perception (Piazza et al., 2010). Nevertheless, a long-standing debate has concerned whether the core deficit concerns nonsymbolic numerosities, symbolic numbers, or both (De Smedt et al., 2013; Landerl et al., 2004; Rousselle & Noël, 2007). Moreover, dissection of symbolic number processing has also suggested a deficit in the processing of ordinality and possibly of general ordinal information (Attout & Majerus, 2015). Although single-deficit theories are unlikely to explain the heterogeneity of individual profiles in learning disabilities (for dyslexia, see Perry et al., 2019), domain-specific accounts of developmental dyscalculia have important implications for both diagnosis and intervention. However, the picture becomes even more complex when we consider that individuals with developmental dyscalculia show weaknesses in domain-general cognitive abilities, which has led to theoretical accounts that point to deficits in general cognitive skills rather than in numerical processing as etiology of mathematical disability (for a review, see Kaufmann et al., 2013). Accordingly, poor visuospatial working memory has been highlighted as a key contributor to developmental dyscalculia at both the behavioral and neural levels (Ashkenazi et al., 2013), and the domain-general account has been further supported by failures to observe domain-specific deficits in children with MLD (Mammarella et al., 2021; Szűcs et al., 2013).

A crucial issue that plagues the ongoing debate and prevents drawing solid inferences about the etiology of developmental dyscalculia is the use of lenient and highly variable inclusion criteria to identify children with developmental dyscalculia. For instance, children were assigned to the developmental dyscalculia or MLD group when their math achievement was below the 15th percentile (e.g., Rousselle & Noël, 2007), the 16th percentile (Szűcs et al., 2013), or even the 35th percentile (Geary et al., 2008; for a discussion, see Peters & Ansari, 2019). A lenient threshold implies that the putative developmental-dyscalculia/MLD group can include many children who are indeed low achievers but do not qualify for a diagnosis of specific learning disability (note that prevalence of developmental dyscalculia according to *DSM-5* criteria is estimated to be around 5%; e.g., Morsanyi et al., 2018). Focusing on children with a severe disorder is therefore important to characterize the impairment of basic number skills that potentially cause mathematical learning difficulties. Moreover, the extremely low mathematical scores would be in sharp contrast to average cognitive skills. Although

Statement of Relevance

One of the causes for struggling with math is a learning disability known as developmental dyscalculia. This study addressed the heated debate on whether dyscalculia originates from a deficit in perceiving the numerosity of object sets, or a difficulty in processing the meaning of number symbols, or even from more general cognitive weaknesses such as poor working memory. We selected severely dyscalculic and non-dyscalculic children closely matched for age, IQ, and visuospatial working memory from a sample referred for learning difficulties and compared their basic number skills with computerized tasks. Performance in choosing the larger among two Arabic digits and in matching two consecutive sets of dots was sufficient to reliably classify a child as dyscalculic. These findings support the hypothesis that dyscalculia stems from a number sense deficit and suggest that these simple numerical tasks could help identify at-risk children and should be a target for intervention.

some manuals (e.g., *DSM-5*) have dropped the discrepancy between mathematical skills and IQ as a diagnostic criterion for developmental dyscalculia, a sharp contrast between extremely low mathematical scores and average cognitive abilities maintains theoretical value because it guarantees the specificity of the observed mathematical difficulties and drastically reduces the possibility to ascribe the deficit to domain-general factors. Another challenge is recruiting a control group that perfectly matches the developmental-dyscalculia group but displays average mathematical performance. Ideally, the two groups should come from the same sample, have undergone the same assessment procedure, and display comparable nonnumerical skills. Such level of control would return an accurate quantification of numerical processing deficits in children with developmental dyscalculia when compared with matched controls. In summary, stringent selection criteria for developmental dyscalculia, an evident discrepancy between numerical and general cognitive abilities, and a fine-grained matching with the control group would lead to an optimal scenario for contrasting competing theoretical perspectives on the cognitive bases of developmental dyscalculia.

Here, we selected children with severe developmental dyscalculia and controls from a large sample ($N >$ 200) of school-age children referred to a specialized center for a formal assessment of cognitive and learning

disabilities. In the developmental-dyscalculia group, we included children who displayed general intelligence within normal limits (IQ > 85) and performance in a standardized numeracy battery that was 2 standard deviations below the expected mean considering age and scholastic grade (i.e., 2nd–3rd percentile). We selected a control group that matched the developmental-dyscalculia group on age, IQ, visuospatial memory, and the presence of dyslexia but displayed average numeracy skills.

The two groups completed several tasks to assess symbolic and nonsymbolic number sense. Symbolic number processing is typically investigated using tasks that require judging the magnitude or the order of digits. Several studies have shown that children with developmental dyscalculia display more errors and slower responses compared with controls when choosing the larger between two digits (Landerl et al., 2004; Rousselle & Noël, 2007). The lower performance in digit comparison has been interpreted as a deficit in accessing numerical magnitude from the symbols. Children with developmental dyscalculia are also slower in judging whether three digits are in order (e.g., 1-2-3) or not (e.g., 2-1-3; Attout & Majerus, 2015), leading to the proposal that developmental dyscalculia is characterized by a deficit in the processing of ordinality rather than cardinality. Children with developmental dyscalculia also show lower performance in estimating the position of target numbers on a visual line, as assessed in the number-line task (Geary et al., 2008).

Nonsymbolic number sense usually refers to the ability to perceive and manipulate the numerosity of object sets. Small numerosities (up to three to four elements) are processed as distinctive objects via a limited-capacity object-tracking system (OTS), whereas larger numerosities (greater than four) are represented as (noisy) summary statistics in the approximate number system (ANS; Feigenson et al., 2004). There is substantial evidence that ANS precision is related to mathematical achievement (e.g., Halberda et al., 2008; for a meta-analysis, see M. Schneider et al., 2017), whereas the role of the OTS remains unclear (e.g., Anobile et al., 2019).

Some studies have suggested that children with developmental dyscalculia display difficulties in processing small numerosities as indexed by a reduced subitizing limit in enumeration tasks (e.g., Landerl et al., 2004). This deficit may stem from a reduced OTS capacity. However, other studies reported that children with developmental dyscalculia display performance comparable with that of typically developing children. Children with developmental dyscalculia have also been shown to display lower accuracy than typically developing children in comparing large numerosities presented as sets of dots (e.g., Piazza et al., 2010), even

though not all studies have reported such impairment (Rousselle & Noël, 2007). Impaired numerosity comparison is thought to index a lower acuity of the ANS (i.e., noisier representations), but it might also reflect a deficit in filtering out irrelevant visual cues (such as the convex hull or total surface area of the sets) that covary with numerosity (Bugden & Ansari, 2016; Piazza et al., 2018). Accordingly, the discrepancy between children with developmental dyscalculia and typically developing children is more marked in (if not limited to) the incongruent trials (Bugden & Ansari, 2016). In this light, visuospatial memory and inhibition may explain the ability to focus on numerical information while filtering out the effect of nonnumerical visual cues (Gilmore et al., 2013), thereby emphasizing a deficit in these domain-general processes as an alternative account of developmental dyscalculia (Szűcs et al., 2013).

Method

Participants

Two hundred forty-seven children completed a full assessment at the neuropsychiatric unit to evaluate the presence of learning and cognitive disability. Data collection lasted for 24 months because of project constraints and negotiation with the clinical facility. At the time of testing, all children attended school levels between Grade 4 of primary school and Grade 3 of middle school (in Italy, primary school consists of five grades and includes children ages 6 to 10 years, whereas middle school consists of three grades and includes children ages 11 to 13 years). Note that, according to Italian regulations, a formal diagnosis of developmental dyscalculia cannot be made before the end of Grade 3 of primary school. We classified children as having developmental dyscalculia if they displayed a performance 2 standard deviations (i.e., scores ≤ 70; $M = 100$, $SD = 15$) below the expected mean considering age and school grade in a standardized numeracy battery (Biancardi et al., 2016; see the Cognitive Assessment section) while having an IQ within the normal range (> 85; Wechsler Intelligence Scale for Children IV [WISC IV]; Wechsler, 2003), no history of neuropsychological disorders, no symptoms of attention-deficit/hyperactivity disorder, and no motor disorders. Thus, children with developmental dyscalculia ($n = 58$) presented average general cognitive abilities but a severe deficit in mathematical skills. We also identified 53 children with both IQ and numeracy scores within the normal range (> 85 for both tests), no history of neuropsychological disorders, no symptoms of attention-deficit/hyperactivity disorder, and no motor disorders. From these children, we selected a control group matched for IQ, age, and

Table 1. Descriptive Statistics and Bayesian *t* Test Comparisons for the Developmental-Dyscalculia and Control Groups

Variable	Developmental dyscalculia				Control				BF_{10}
	M	*SD*	Minimum	Maximum	*M*	*SD*	Minimum	Maximum	
Age in months	136.19	18.89	110	174	140.71	20.29	108	178	0.4
IQ	100.07	9.93	86	120	103.79	9.97	86	124	0.9
Visuospatial memory (Memory for Designs of NEPSY-II)	100.9	26.87	41	147	110.26	24.56	36	144	0.9
Numeracy score (Batteria Discalculia Evolutiva)	56.59	7.08	49	69	99.52	8.58	87	128	8.4×10^{43}

Note: BF_{10} = Bayes factor of the *t* tests between the two groups.

visuospatial memory (i.e., Memory for Designs subtest of the NEPSY-II; Korkman et al., 2007) to the developmental-dyscalculia group. Note that visuospatial memory is one of the domain-general processes that most often relates to mathematical achievement (e.g., Szűcs et al., 2013). The initial sample of control children displayed higher scores in visuospatial working memory but not in IQ and age (Bayesian *t* tests). We then ranked the control sample according to the visuospatial memory score and iteratively eliminated the best-performing child from the control sample until the two groups had similar scores (i.e., Bayes factor [BF] < 1 across the three scores). Therefore, the final control group was composed of 42 children matched for age, IQ, and visuospatial memory to the developmental-dyscalculia group (see Table 1). Thirty-two children in the developmental-dyscalculia group and 13 in the control group also received a diagnosis of dyslexia, $\chi^2(1) = 2.46$, $p = .12$.

Cognitive assessment and numerical tasks

Intelligence. Children completed the WISC-IV (Wechsler, 2003), from which we extracted a measure of full IQ ($M = 100$, $SD = 15$) based on chronological age.

Numeracy. Children completed a standardized numeracy battery used in Italy for the clinical assessment of developmental dyscalculia in children from the third year of primary school to the third year of middle school (Batteria Discalculia Evolutiva [BDE-2]; Biancardi et al., 2016). The battery is composed of multiple subtests assessing a variety of numerical and mathematical skills: counting, number reading, writing and repetition, mental calculation, arithmetic facts, insertions, numerical triplets, approximate calculations, number-line estimation, written operations, and

arithmetic problems. The raw scores of each subtest are converted into normative scores ($M = 10$, $SD = 3$) and the sum composes the total score ($M = 100$, $SD = 15$), which we used for selecting the two groups of children. The battery presents good internal consistency (average Cronbach's α across subtests = 0.7).

Visuospatial memory. We assessed visuospatial memory using the Memory for Designs subtest of the NEPSY-II (Korkman et al., 2007). The experimenter presented some abstract figures (six to 10) located in different positions on the grid (21 cm × 29.7 cm) for 10 s with the instructions for the child to memorize the figures and their locations. Then, the child chose the abstract figures previously memorized from a set of cards (from 10 to 20) and placed them on a blank grid. We used the total score, which combines the ability to remember both figures and location, as a measure of visuospatial memory.

Match to sample. The match-to-sample task (Sella et al., 2013) assessed the ability to sequentially compare the numerosity of two sets of dots. Participants were shown a fixation cross for 400 ms and then a blank screen for 150 ms (see Fig. 1a). Thereafter, a sample set appeared for 300 ms, followed by a blank screen for 1,000 ms. Then, a target set appeared and remained on screen until a response was given. Participants decided whether the two sets contained the same or a different number of dots by pressing the left or the right key of the keypad, respectively. The numerosity of the target set matched the numerosity of the sample set (match condition) in half of the trials, whereas in the other half, the numerosity of the target set was −1 or +1 with respect to the sample set (nonmatch condition). When one dot or eight dots composed the sample set, the numerosity of the

Nonsymbolic Tasks

a

Match-to-Sample Task

Fixation
400 ms

Blank
150 ms

Sample
300 ms

Blank
1000 ms

Match   Nonmatch   Nonmatch

Time

b

Panamath

Fixation
Until a key
is pressed
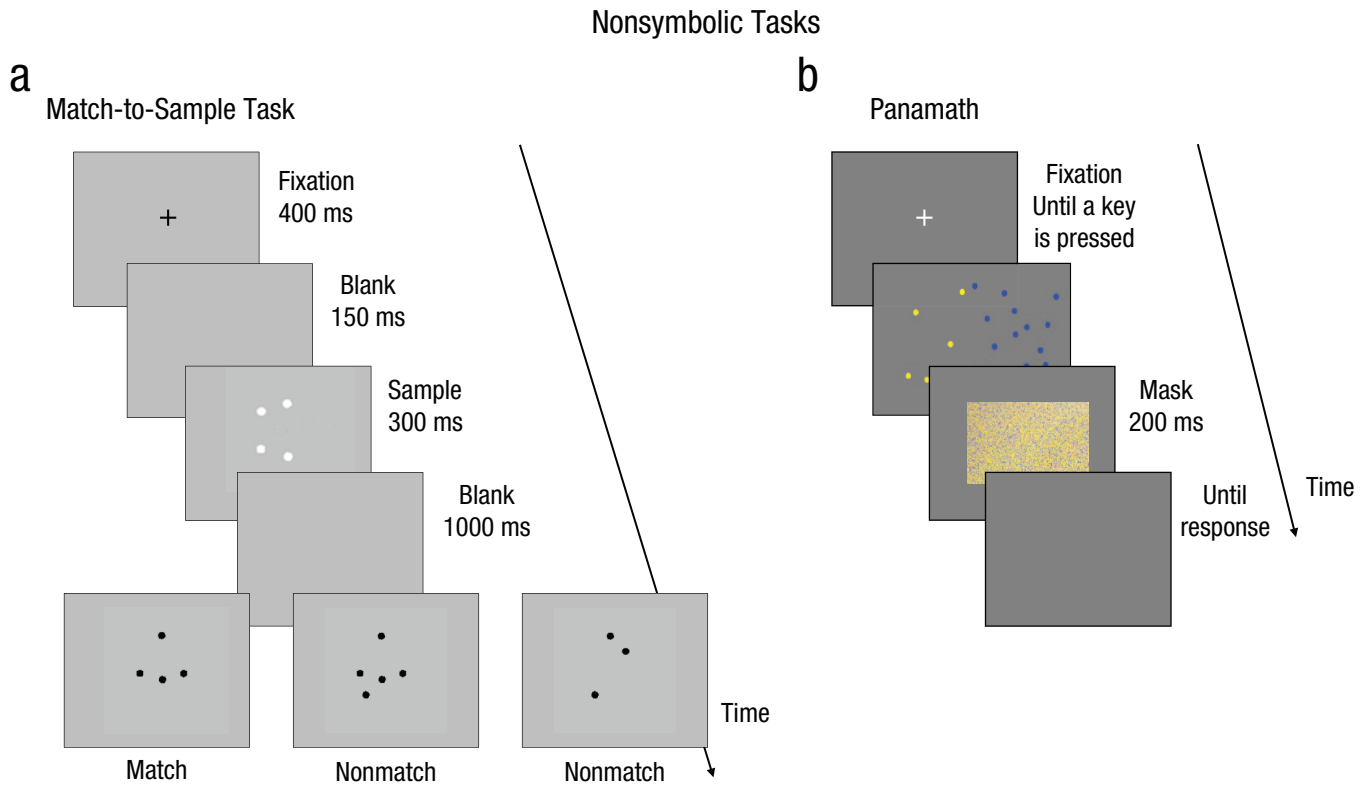
Mask
200 ms

Time

Until
response

Fig. 1. (a) Match-to-sample task: Children decided whether the two sequentially presented sets contained the same or a different number of dots. (b) Panamath: Children indicated which of two visually presented arrays contained more dots.

target set in the nonmatch condition was two dots or seven dots, respectively. The size of the dots and their spatial arrangement varied across trials, and the sample and the target sets had opposite polarity; that is, the sample set was composed of white dots, whereas the target set was composed of black dots. These manipulations ensured that participants extracted numerical information rather than basing their judgment on nonnumerical visual cues. After 10 practice trials, there were 12 trials for each numerosity from one to seven, and the numerosity eight was presented only in six trials, yielding a total of 90 test trials. For each target numerosity, we calculated both response accuracy and median response time (RT). In each trial, children briefly saw a sample array followed by the presentation of a target array, which remained on screen until a response was given. Therefore, accuracy reflects precision in perceiving the number of dots in the sample array, whereas RT reflects speed in enumerating the target array, and both measures can highlight the transition between small and large numerosities (e.g., Fu et al., 2022).

The match-to-sample task assesses both OTS capacity (e.g., one vs. two, two vs. three, three vs. four) and ANS acuity (comparison of large numerosities from four vs. five to eight vs. nine) by presenting numerosities

from one to nine. The ±1 deviation or matching between the sample and the target arrays entails incremental difficulty as the numerical ratio between the arrays gets closer to 1 (e.g., from one vs. two to eight vs. nine). The small numerical deviation also forces participants to extract numerical information because the two arrays are similar in terms of nonnumerical visual cues (e.g., convex hull). Moreover, the sequential presentation reduces the possibility of visually comparing the two arrays, which more likely happens when the two arrays are presented simultaneously. Finally, the task does not require any verbal response.

Panamath. In the Panamath (Halberda et al., 2008) numerosity comparison task, children had to choose, as quickly as possible, which of two visually presented arrays contained more dots (see Fig. 1b). The numerical sets varied between five and 21 dots and were shown in blue and yellow. We used the default setting of the software that adjusted the presented numerical ratios and the timing of stimuli presentation depending on age. Notably, in half of the trials, the size of the dots correlated with the numerosity (i.e., congruent condition), whereas in the other half, the size of the dots was equated between the two sets (i.e., incongruent condition). We extracted
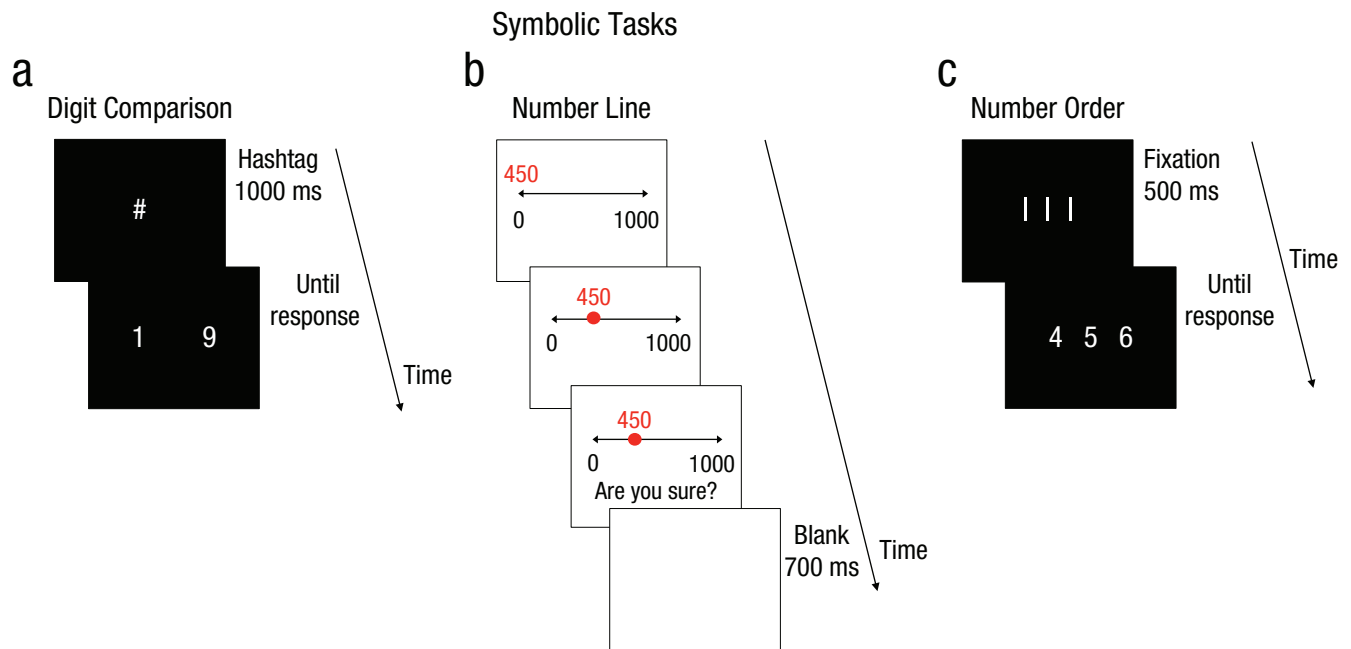
Symbolic Tasks

a
Digit Comparison

Hashtag
1000 ms

1 9

Until response

Time

b
Number Line

450

0 1000

450

0 1000

450

0 1000

Are you sure?

Blank 700 ms

Time

c
Number Order

Fixation 500 ms

| | |

4 5 6

Until response

Time

Fig. 2. (a) Digit-comparison task: Children indicated the larger of two digits. (b) Number-line task: Children moved the target number to the selected position on the line and clicked one of the mouse buttons to place the number. (c) Number-order task: Children judged whether the triplets presented on screen were in ascending order or not.

the estimated Weber fraction (*wf*) separately for the congruent and incongruent conditions.

Digit comparison. Participants were presented with two Arabic digits from 1 to 9 and were asked to choose the larger as quickly and as accurately as possible by clicking the keypad response button on the side of the chosen number. The digits remained on screen until the participant responded. Between each trial, a central fixation hashtag appeared for 1,000 ms (see Fig. 2a). There were two practice trials followed by 72 test trials entailing all the possible comparisons of digits between 1 and 9 repeated twice. The larger number appeared on the left side of the screen in half of the trials. For each participant, we calculated the median RTs for correct responses, accuracy, and an efficiency score. Following Lyons et al. (2014), we computed efficiency as RT(1 + 2 × error rate). A linear combination of speed and accuracy provides a useful summary index, provided that the speed and accuracy data are also inspected (Vandierendonck, 2017). In this regard, we ensured that accuracy was high (i.e., more than 90% correct responses) and that there was a high positive correlation between RTs and accuracy (*r* = .95).

Number line. In the number-line task (Siegler & Opfer, 2003), a horizontal bounded line was presented in the middle of the screen. The left and right ends of the line were labeled 0 and 1,000, respectively. A red target number appeared just above the line either at the beginning or end of the line alternatively, and participants moved the target number using the mouse and clicked one of the mouse buttons to place it on the line (see Fig. 2b). Children were instructed to place the target number in the correct position on the line. After the target number was positioned, a red dot appeared on the selected location, and children pressed the "Y" key to confirm their response or the "N" key to repeat the trial. There were three training trials, in which children had to position 0, 1,000, and 500. The experimenter showed the correct position of the training trials 0 and 1,000 in case of inaccurate positioning, thereby ensuring that children were aware of the numerical interval. Thereafter, 24 target numbers (i.e., 10, 130, 140, 230, 260, 270, 320, 360, 390, 410, 450, 490, 530, 540, 580, 620, 660, 670, 720, 750, 790, 850, 880, 980) were randomly presented. For each participant, we calculated the mean RTs and the mean absolute deviation (i.e., |estimate-target number|).

Number order. In the number-order task, three vertical lines appeared in the middle of the screen for 500 ms and were replaced by a triplet of Arabic digits (e.g., 1-2-3; see Fig. 2c), which remained on screen until a response was given. Participants judged as fast as they could whether the triplet was in ascending order or not. After four training trials, there were 28 trials: 14 in ascending order (1-3-5, 2-5-8, 3-5-7, 3-6-9, 4-5-6, and 6-7-8, repeated twice; 2-3-4 and 5-7-9, repeated once), seven in descending order (4-3-2, 5-3-1, 6-5-4, 7-4-1, 8-5-2, 8-7-6, 9-7-5), and

seven not ordered (4-2-3, 5-3-7, 5-8-2, 6-3-9, 6-4-5, 7-1-4, and 7-5-9). For each participant, we calculated the median RTs for correct responses and the efficiency score. For the latter, we ensured that accuracy was high and that there was a high positive correlation between RTs and accuracy ($r = .82$).

Procedure

Children were tested by research assistants in a quiet room of the child neuropsychiatric unit. Children were assessed using a PC with a screen resolution of 1,366 × 768, and stimuli were presented using E-Prime software (Version 2.0; W. Schneider et al., 2012). The administration of the tasks was conducted in two sessions on different days. During the first session, children completed the cognitive assessment (WISC-IV, BDE-2); during the second session, they completed the computerized tasks and the visuospatial memory test. Parents gave written informed consent for their child to participate in the study, whereas children gave verbal consent. The protocol was approved by the Psychological Science Ethics Committee of the University of Padova.

Results

We reported both frequentist and Bayesian analyses using the R programming environment. We used the *BayesFactor* package with default priors (Morey et al., 2015) for computing Bayesian t tests and analyses of variance (ANOVAs). We reported BFs (BF_{10}) expressing the probability of the data given Hypothesis 1 relative to Hypothesis 0: Values larger than 1 are in favor of Hypothesis 1, and values smaller than 1 are in favor of Hypothesis 0. We described the evidence associated with BFs as "anecdotal" ($1/3 < BF < 3$), "moderate" ($BF < 1/3$ or $BF > 3$), "strong" ($BF < 1/10$ or $BF > 10$), "very strong" ($BF < 1/30$ or $BF > 30$), or "extreme" ($BF < 1/100$ or $BF > 100$). The number of participants varied across tasks because of testing constraints or computer failures. We reported the number of participants in the developmental-dyscalculia and control groups for each task. Note that the two groups remained matched for IQ, age, and visuospatial memory (see Table S1 in the Supplemental Material available online). The sample size (58 with developmental dyscalculia and 42 controls) provides 80% power to detect an effect size of 0.57 (i.e., Cohen's d) with .05 (two-tailed) significance level (G*Power, Version 3.1; Faul et al., 2007). We also carried out separate analyses for the group of children with pure developmental dyscalculia and the group with comorbidity between developmental dyscalculia and dyslexia. The results, reported in the Supplemental Material, did not show any difference between the two groups in all the numerical tasks, whereas both groups showed worse performance compared with the control group.

Match to sample

In line with previous studies, we performed separate analyses for large and small target numerosities (the analyses on each target numerosity can be found in the Supplemental Material). We compared the neighboring numerosities (one vs. two, two vs. three, three vs. four,...), and we found the first significant difference in RTs and accuracy when comparing target numerosities three and four—accuracy: $t(91) = 4.8$, $p < .001$, $d = 0.55$, $BF_{10} = 2535$, extreme evidence; RTs: $t(91) = -7.6$, $p < .001$, $d = 0.8$, $BF_{10} = 2.9 \times 10^{11}$, extreme evidence. Therefore, we included in the small numerical range the target numerosities one, two, and three, whereas target numerosities four, five, six, and seven were included in the large numerical range. We excluded target numerosity eight because participants were at the chance level.

We performed a Bayesian mixed ANOVA on accuracy with numerical range (small, large) as the within-subjects factor and group (developmental dyscalculia, control) as the between-subjects factor. The model with the main effect of numerical range and group yielded the highest evidence ($BF_{10} = 5.9 \times 10^{27}$) and was superior to the two models including only the single main effect of numerical range ($BF_{10} = 2.6 \times 10^{26}$) and group ($BF_{10} = 2$). The two groups displayed similar accuracy in comparing small numerosities (developmental dyscalculia: $n = 54$, $M = .88$, $SD = .09$; control: $n = 38$, $M = .91$, $SD = .09$), $t(90) = -1.28$, $p = .2$, $d = 0.27$, 95% confidence interval (CI) = $[-0.15, 0.69]$, $BF_{10} = 0.4$, anecdotal evidence (see Fig. 3a), whereas children with developmental dyscalculia were less accurate ($M = .63$, $SD = .12$) than controls ($M = .72$, $SD = .13$) in comparing large numerosities, $t(90) = -3.09$, $p = .003$, $d = 0.65$, 95% CI = $[0.21, 1.08]$, $BF_{10} = 13.11$ (strong evidence; see Fig. 3a).

A Bayesian mixed ANOVA was also performed on median RTs with numerical range (small, large) as the within-subjects factor and group (developmental dyscalculia, control) as the between-subjects factor. The model with the main effect of numerical range yielded the highest evidence ($BF_{10} = 5.6 \times 10^{11}$) compared with the model with the main effects of numerical range and group ($BF_{10} = 4.1 \times 10^{11}$) as well as with the model including only group ($BF_{10} = 0.5$, anecdotal evidence). Children with developmental dyscalculia were slower than controls in comparing small numerosities (developmental dyscalculia: $M = 1,164$ ms, $SD = 291$ ms; control: $M = 956$ ms, $SD = 204$ ms), $t(90) = 3.81$, $p <$
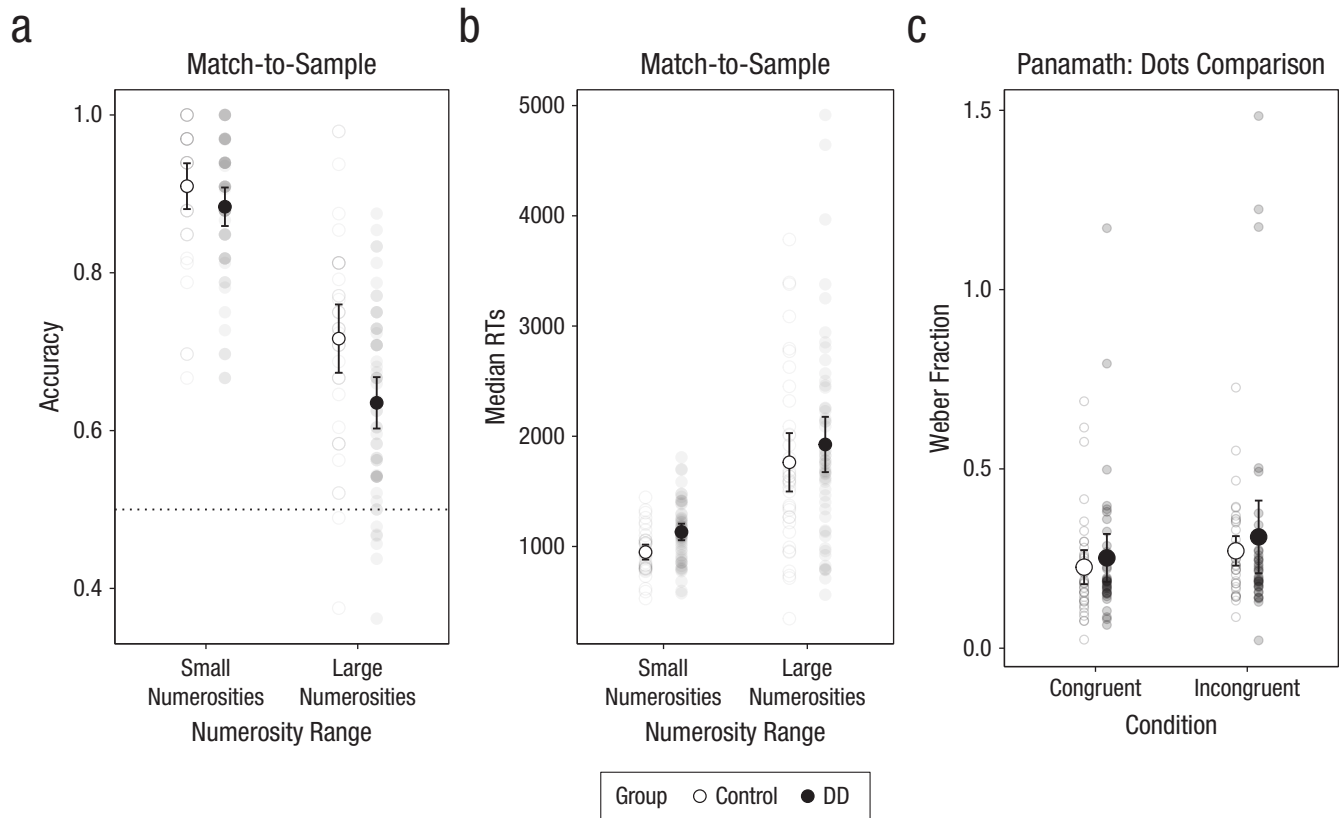
Fig. 3. Performance in the nonsymbolic tasks for the developmental dyscalculia (DD) group (black dots) and the control group (white dots). (a) Accuracy (proportion of correct responses) in the match-to-sample task as a function of target numerosity (small range vs. large range). The dashed line represents the chance level. (b) Median response times (RTs) in the match-to-sample task as a function of the target numerosity (small range vs. large range). (c) Weber fraction in the Panamath (numerosity comparison) task in congruent (i.e., numerosity and total surface area correlated) and incongruent (i.e., total surface area is equated between the comparison sets) trials. Error bars represent 95% confidence intervals, and transparent dots represent individual scores.

.001, $d = 0.81$, 95% CI = [0.36, 1.25], $BF_{10} = 97.46$ (very strong evidence), whereas both groups did not show different RTs in comparing large numerosities (developmental dyscalculia: $M = 2,019$ ms, $SD = 964$ ms; control: $M = 1,790$ ms, $SD = 789$ ms), $t(90) = 1.21$, $p = .23$, $d = 0.26$, 95% CI = [−0.16, 0.68], $BF_{10} = 0.42$, anecdotal evidence (see Fig. 3b).

Panamath

We analyzed the Weber fraction in a Bayesian mixed ANOVA with condition (congruent, incongruent) as the within-subjects factor and group (developmental dyscalculia, control) as the between-subjects factor. The model with the main effect of condition provided the highest evidence ($BF_{10} = 5.84$, moderate evidence, $p = .006$, $\eta_G^2 = .02$). We found anecdotal evidence for a null main effect of the group ($BF_{10} = 0.43$, $p = .46$), and the two groups displayed similar performance (developmental dyscalculia: $n = 38$, $M = .27$, $SD = .23$; control: $n = 38$, $M = .24$, $SD = .11$; see Fig. 3c). Moreover, we

performed a Bayesian mixed ANOVA for the RTs with condition as the within-subjects factor and group as the between-subjects factor. The model with the main effect of group provided the highest evidence ($BF_{10} = 0.53$, $p = .52$, $\eta_G^2 = .005$, anecdotal evidence for the null). The two groups showed similar RTs (developmental dyscalculia: $M = 1,792.89$, $SD = 650.18$; control: $M = 1,569.62$, $SD = 520.75$).

Digit comparison

Children with developmental dyscalculia were slower ($n = 42$, $M = 986$ ms, $SD = 265$) than controls ($n = 35$, $M = 747$ ms, $SD = 171$) in choosing the larger of two digits, $t(75) = 4.6$, $p < .001$, $d = 1.05$, 95% CI = [0.54, 1.55], $BF_{10} = 1,075$, extreme evidence (see Fig. 4a), whereas the proportion of correct responses was high in both groups (developmental dyscalculia: $M = 0.96$, $SD = 0.03$; control: $M = 0.97$, $SD = 0.02$), $t(75) = −1.9$, $p = .06$, $d = 0.44$, 95% CI = [−0.02, 0.9], $BF_{10} = 1.13$, anecdotal evidence. In line with the RT analysis, the

a

Digit Comparison

b

Number Line
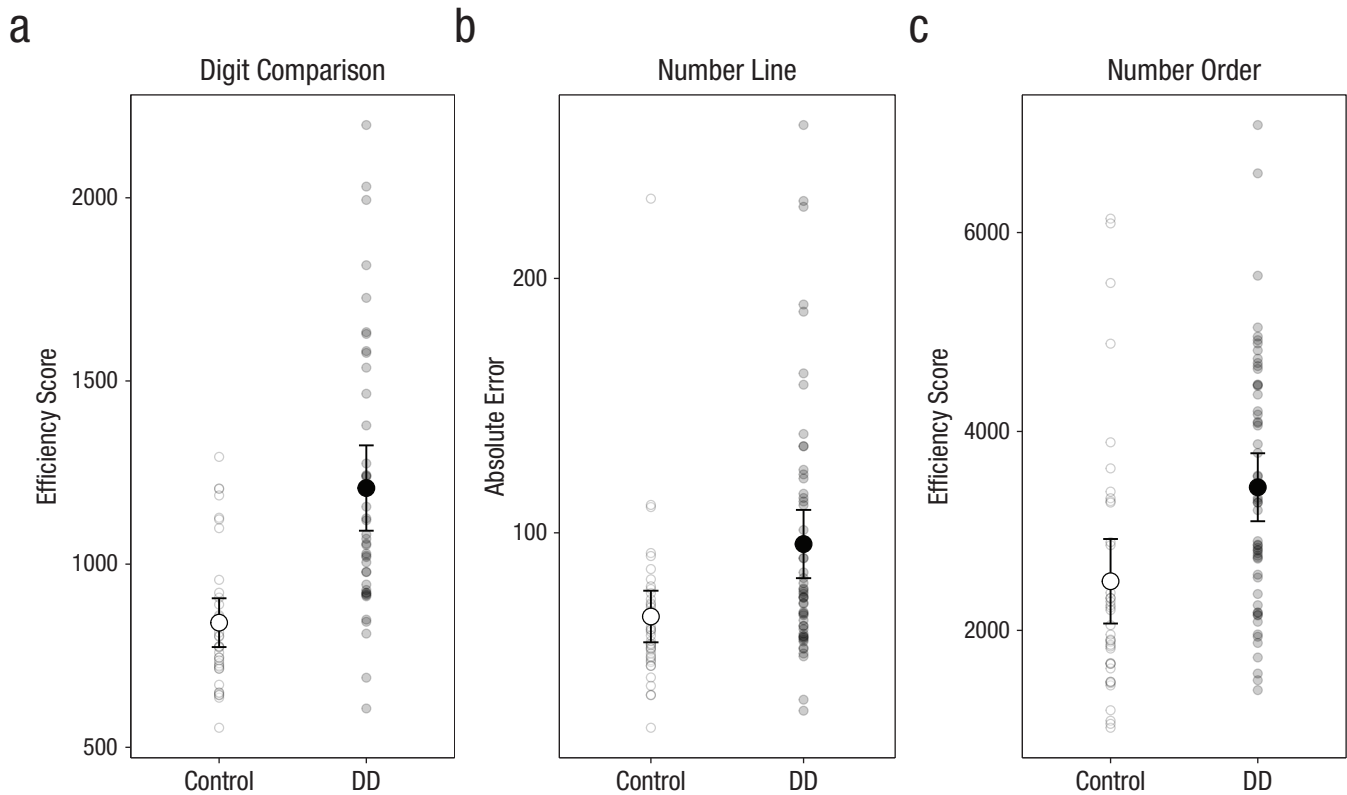

c

Number Order

Fig. 4. Performance in the symbolic tasks for the developmental dyscalculia (DD) group (black dots) and the control group (white dots). (a) Efficiency score in the digit-comparison task. (b) Mean absolute error (i.e., |estimate – target|) in the number-line task. (c) Efficiency score in the number-order task. Error bars represent 95% confidence intervals, and transparent dots represent individual scores.

developmental-dyscalculia group displayed less efficiency when comparing Arabic digits (developmental dyscalculia: $M = 1,208$, $SD = 374$; control: $M = 840$, $SD = 194$), $t(75) = 5.25$, $p < .001$, $d = 1.2$, 95% CI = [0.68, 1.71], $BF_{10} = 10,465$, extreme evidence.

Number line

Children with developmental dyscalculia displayed larger absolute error ($n = 55$, $M = 95.59$, $SD = 49.63$) when placing target numbers on the line compared with the control group ($n = 41$, $M = 67.1$, $SD = 32.12$), $t(94) = 3.21$, $p = .002$, $d = 0.66$, 95% CI = [0.23, 1.08], $BF_{10} = 17.7$ (strong evidence; see Fig. 4b). Conversely, there was anecdotal evidence for no difference between the two groups in terms of RTs (developmental dyscalculia: $M = 5,657$ ms, $SD = 1,932$; control: $M = 5,051$ ms, $SD = 2,019$), $t(94) = 1.49$, $p = .14$, $d = 0.31$, 95% CI = [−0.1, 0.72], $BF_{10} = 0.6$.

Number order

Children with developmental dyscalculia were slower ($M = 2,233$ ms, $SD = 773$) in judging whether the triplets of digits were in ascending order or not compared with controls ($M = 1,709$ ms, $SD = 711$), $t(93) = 3.36$,

$p = .001$, $d = 0.7$, 95% CI = [0.27, 1.13], $BF_{10} = 26.7$ (strong evidence), whereas the two groups displayed similar accuracy (developmental dyscalculia: $n = 56$, $M = 0.9$, $SD = 0.13$; control: $n = 39$, $M = .9$, $SD = .11$), $t(93) = −1.22$, $p = .23$, $d = 0.25$, 95% CI = [−0.16, 0.66], $BF_{10} = 0.4$, anecdotal evidence. We found the same pattern of results when we analyzed efficiency scores (developmental dyscalculia: $M = 3,439.01$, $SD = 1,276.17$; control: $M = 2,493.89$, $SD = 1,310.91$), $t(93) = 3.51$, $p < .001$, $d = 0.73$, 95% CI = [0.3, 1.16], $BF_{10} = 41$ (very strong evidence; see Fig. 4c).

Logistic regression analyses

We performed a series of logistic regression analyses with group (developmental dyscalculia = 1, control = 0) as the outcome variable and age, full IQ, visuospatial memory, RTs, and accuracy for large and small target numerosities in the match-to-sample task, the efficiency score in the digit-comparison task, the absolute error in the number-line task, and the efficiency score in the number-order task as predictor variables. We aimed to identify the combination of measures that best separate children with developmental dyscalculia from controls (all the correlations between these measures are reported in Table S2 in the Supplemental Material).
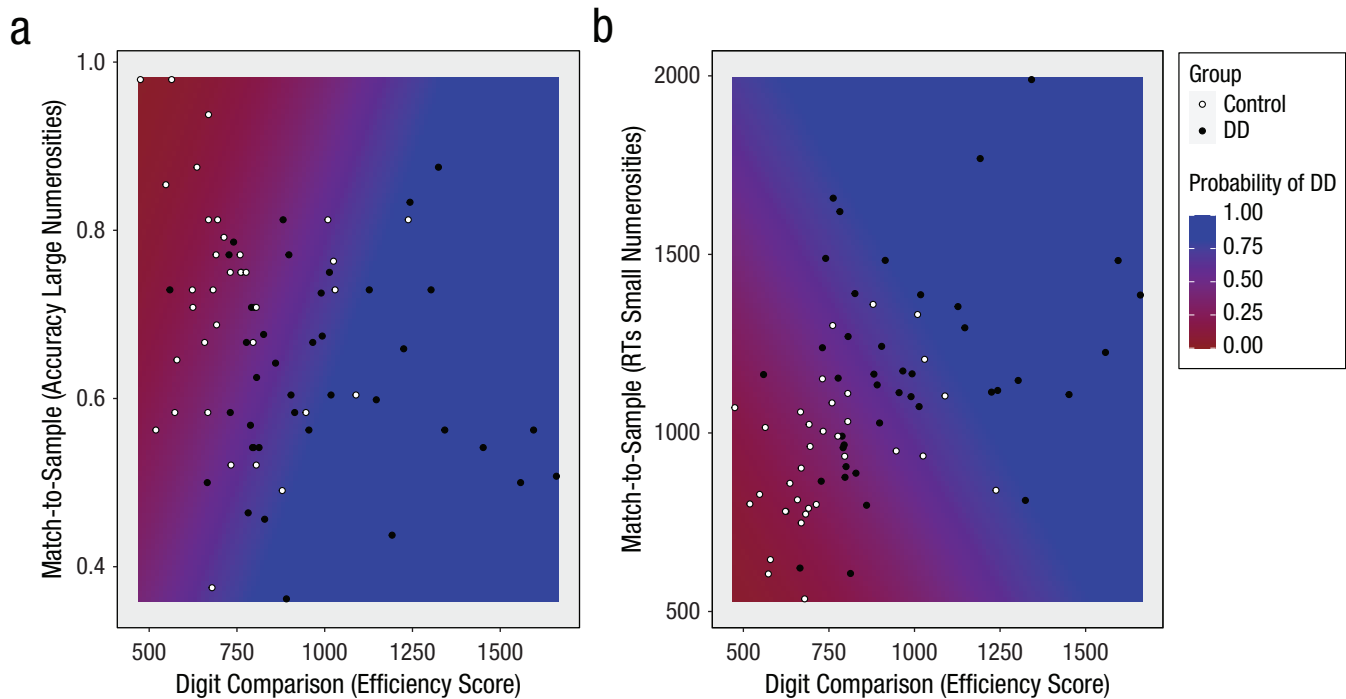
Fig. 5. (a) Accuracy for large numerical quantities in the match-to-sample task (*y*-axis) as a function of the efficiency score in the digit-comparison task (*x*-axis) in children with developmental dyscalculia (DD; black dots) and controls (white dots). The background hue represents the probability of the logistic regression model to classify a child as having DD (0 = purple, 1 = blue). (b) Median response times (RTs) for small numerical quantities in the match-to-sample task (*y*-axis) as a function of the efficiency score in the digit-comparison task (*x*-axis) in children with DD (black dots) and controls (white dots). The background hue represents the probability of the logistic regression model to classify a child as having DD (0 = purple, 1 = blue).

We assessed the models including all the possible combinations of predictors and then ranked them on the basis of the Bayesian information criterion (BIC). The model including the accuracy for large numerosities in the match-to-sample task and the efficiency score in the digit-comparison task yielded the lowest BIC (i.e., best model; BIC: 85.76; accuracy for large numerosities: $\beta = -4.52$, $p = .06$; efficiency score: $\beta = 0.005$, $p < .001$; see Fig. 5a). However, there were at least two other models including two or fewer (one) predictors that could be considered comparable with the best model (i.e., < 2 from the BIC of the best model; Raftery, 1995): (a) the model including the RTs for small numerosities in the match-to-sample task and the efficiency score in the digit-comparison task (BIC: 86.19; RTs for small numerosities: $\beta = 0.002$, $p = .08$; efficiency score: $\beta = 0.004$, $p = .002$; see Fig. 5b) and (b) the model including only the efficiency score in the digit-comparison task (BIC: 86.81; efficiency score: $\beta = 0.005$, $p < .001$). The number of children for each model changed because not all of them had done all of the tasks. However, when we performed the same analysis on the reduced sample of children who were assessed in all of the tasks, the best models remained the same.

For the best model as well as for the two runners up, we calculated sensitivity and specificity using leave-one-out cross-validation (see Table S3 in the Supplemental Material). The model with accuracy for large numerosities and efficiency score yielded a mean proportion of correct classification of 0.72 with a sensitivity of 0.68 and a specificity of 0.76; the model with RTs for small numerosities and efficiency score led to a mean classification accuracy of 0.76 with a sensitivity of 0.76 and a specificity of 0.76. The model with just the efficiency score had a classification accuracy of 0.73 with a sensitivity of 0.76 and a specificity of 0.71. Finally, for each model, we performed a receiver operating characteristic (ROC) analysis and computed the area under the ROC curve (AUC), which is a scale-invariant and threshold-invariant measure of the quality of the model's predictions. The model with accuracy for large numerosities and efficiency score yielded an AUC of .841; the model with RTs for small numerosities and efficiency score yielded an AUC of .836; and the model with just the efficiency score yielded an AUC of .823.

General Discussion

We selected a developmental-dyscalculia group and a control group from a larger sample of children who were referred to a specialized center for formal assessment of cognitive and learning disabilities. Children in

the developmental-dyscalculia group displayed average IQ and visuospatial memory, whereas their numeracy score was 2 standard deviations below the expected mean considering age and schooling. Children in the control group matched the developmental-dyscalculia group on IQ, visuospatial memory, and age but had an average numeracy score. Both groups displayed similar reading skills, as indexed by a similar rate in the diagnosis of dyslexia. This fine-grained matching provided an optimal condition to assess nonsymbolic and symbolic number sense in children with severe dyscalculia while controlling for domain-general processes (e.g., intelligence, visuospatial memory) and other confounding factors (e.g., reading skills, being referred to the neuropsychiatric unit).

When compared with the control group, children with developmental dyscalculia displayed lower performance in all the symbolic tasks. In line with previous studies, children with developmental dyscalculia were slower when judging the magnitude and the order of Arabic digits (e.g., Attout & Majerus, 2015; Landerl et al., 2004; Rousselle & Noël, 2007) as well as less accurate when placing target numbers on the visual line (e.g., Geary et al., 2008; Landerl et al., 2009). These three symbolic numerical skills have been related to mathematical attainment via specific mechanisms. The slow access to numerical magnitude undermines children's implementation of efficient arithmetic strategies, such as choosing the larger addend when solving additions (Vanbinst et al., 2012). The ordinal judgment requires the combination of magnitude comparison and retrieval from memory. Accordingly, in ordered nonconsecutive triplets (e.g., 1-3-5) and in nonordered triplets (e.g., 1-5-3), individuals more likely adopt sequential magnitude comparison strategies, whereby the magnitude of the first digit is compared with the second and the second with the third. Conversely, in consecutive ordered sequences, individuals are more likely to identify the triplet as a portion of the counting sequence (e.g., 1-2-3) because the first digit activates the second and the second activates the third in a chain stored in long-term memory (Sella et al., 2020). The retrieval of declarative knowledge from memory explains why the number-order judgment task becomes one of the stronger correlates of arithmetic fluency from the second grade when children start using memory-based retrieval strategies to solve arithmetic problems (Sasanguie & Vos, 2018). Finally, the lower performance in the number-line task might reflect poor mapping of numbers into space (Landerl et al., 2009) or children's inefficient use of arithmetic strategies (e.g., using half and quarters of the line as anchoring points) when solving the task.

Children with developmental dyscalculia also displayed lower performance than control children in nonsymbolic numerical processing when assessed with the match-to-sample task. The structure of the task implies that children perceive the sample numerosity by either subitizing or estimation, hold the numerosity representation in memory during the delay period, and finally match it with the target numerosity. Both groups displayed high accuracy for small numerosities (up to three), whereas children with developmental dyscalculia were less accurate with larger numerosities than controls. This suggests that numerosity perception is impaired in the estimation range. However, in contrast to previous studies (Piazza et al., 2010), the estimation deficit was not detected in the Panamath numerosity comparison task, even when the analysis was restricted to the incongruent trials that require extraction of numerical information against nonnumerical visual cues (Gilmore et al., 2013) and appear to drive the poor comparison performance in dyscalculia (Bugden & Ansari, 2016). Potential differences between the two groups might have been diluted by matching them on visuospatial memory skills. Conversely, we speculate that the sequential (rather than simultaneous) presentation of numerosities in the match-to-sample task turned out to be more sensitive because it requires a more cognitively demanding three-step process of extracting and maintaining the numerical magnitude representation of the sample and target sets before their numerosities are compared.

Children with developmental dyscalculia were also slower than controls in the match-to-sample task, particularly in the subitizing range despite showing the same (high) level of accuracy. RTs to the target tap dot enumeration, which has been shown to characterize children with developmental dyscalculia in a large prevalence study (Reigosa-Crespo et al., 2012) and is longitudinally related to children's arithmetic development (Reeve et al., 2012). Our results suggest that enumeration in the subitizing range is particularly inefficient in developmental dyscalculia, in line with previous evidence of slow enumeration of small numerical quantities (Landerl et al., 2004).

The efficiency score in the digit-comparison task was a strong predictor for separating children with developmental dyscalculia from matched controls. Nevertheless, performance in the match-to-sample task (either the RTs for small numerosities or accuracy for larger numerosities) improved sensitivity. The combination of symbolic and nonsymbolic number-sense measures supported a reliable identification of children with developmental dyscalculia in cross-validated logistic regression, yielding an AUC of .84. These results are in stark contrast with those of Mammarella et al. (2021), who reported no evidence for core deficits in their MLD sample as well as low discriminative power of basic number processing measures (both symbolic and nonsymbolic) with AUCs lower than .70. This suggests that

one potential limit to the generalizability of the findings across samples stems from the selection criteria for DD/MLD groups. In particular, the use of clinical criteria for inclusion as DD implies not only a more stringent threshold but also a broader assessment of numeracy skills compared to the case of selecting MLD children from a large school-based sample. Indeed, the latter is influenced by potential variability in the choice of cut-off percentile (Peters & Ansari, 2019, for discussion) as well as by diversity in math performance assessment, which might be brief (for practical reasons) and possibly biased towards specific sub-skills (e.g., calculation).

In summary, the combination of stringent inclusion criteria, a large sample size, fine-grained matching, and thorough assessment of symbolic and nonsymbolic numerical skills makes the findings of our study relevant for both the field of mathematical cognition and developmental psychology. Although convincing evidence supports the presence of multiple core deficits in other (learning) disabilities, there is still debate on the presence and type of deficits in developmental dyscalculia. Our results suggest that severe developmental dyscalculia is characterized by core deficits in both symbolic and nonsymbolic number sense. Children with comorbid developmental dyscalculia and dyslexia did not differ from those with only developmental dyscalculia across all numerical tasks (see also Landerl et al., 2009). These domain-specific deficits reliably discriminate children with developmental dyscalculia from controls while controlling for domain-general cognitive abilities such as IQ and visuospatial memory. Our results are consistent with the view that nonsymbolic and symbolic comparison skills mutually influence each other during development, with the latter becoming the stronger predictor of arithmetic skills (Lyons et al., 2014) and overall mathematical competence (for a meta-analysis, see M. Schneider et al., 2017). A relevant implication of our findings is that performance in these basic number-sense tasks provides key information for the differential diagnosis of developmental dyscalculia and could become the cornerstone of early identification because they do not directly reflect the outcome of school-based learning and can be easily completed by young children (e.g., Sella et al., 2013).

Transparency

ORCID iDs

Gisella Decarli 🔟 https://orcid.org/0000-0002-6784-4457
Marco Zorzi 🔟 https://orcid.org/0000-0002-4651-6390

Acknowledgments

Supplemental Material

Additional supporting information can be found at http://journals.sagepub.com/doi/suppl/10.1177/09567976221097947

References

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). American Psychiatric Association.

Anobile, G., Arrighi, R., & Burr, D. C. (2019). Simultaneous and sequential subitizing are separate systems, and neither predicts math abilities. *Journal of Experimental Child Psychology*, *178*, 86–103. https://doi.org/10.1016%2Fj.jecp.2018.09.017

Ashkenazi, S., Rosenberg-Lee, M., Metcalfe, A. W., Swigart, A. G., & Menon, V. (2013). Visuo-spatial working memory is an important source of domain-general vulnerability in the development of arithmetic cognition. *Neuropsychologia*, *51*(11), 2305–2317. https://doi.org/10.1016/j.neuropsychologia.2013.06.031

Astle, D. E., & Fletcher-Watson, S. (2020). Beyond the core-deficit hypothesis in developmental disorders. *Current Directions in Psychological Science*, *29*(5), 431–437. https://doi.org/10.1177/0963721420925518

Attout, L., & Majerus, S. (2015). Working memory deficits in developmental dyscalculia: The importance of serial order. *Child Neuropsychology*, *21*(4), 432–450. https://doi.org/10.1080/09297049.2014.922170

Biancardi, A., Nicoletti, C., & Bachmann, C. (2016). *BDE 2-Batteria Discalculia Evolutiva: Test per la diagnosi dei*

disturbi dell'elaborazione numerica e del calcolo in età evolutiva–8-13 anni. [Developmental Dyscalculia Battery: Test for diagnosing numerical processing and calculation disorders in childhood - 8-13 y.o.]. Edizioni Centro Studi Erickson.

Bugden, S., & Ansari, D. (2016). Probing the nature of deficits in the 'approximate number system' in children with persistent developmental dyscalculia. *Developmental Science*, *19*(5), 817–833. https://doi.org/10.1111/desc.12324

Butterworth, B., Varma, S., & Laurillard, D. (2011). Dyscalculia: From brain to education. *Science*, *332*(6033), 1049–1053. https://doi.org/10.4324%2F9781315538112

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*(2), 48–55. https://doi.org/10.1016/j.tine.2013.06.001

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*, 175–191.

Feigenson, L., Dehaene, S., & Spelke, E. (2004). Core systems of number. *Trends in Cognitive Sciences*, *8*(7), 307–314. https://doi.org/10.1016/j.tics.2004.05.002

Fu, W., Dolfi, S., Decarli, G., Spironelli, C., & Zorzi, M. (2022). Electrophysiological signatures of numerosity encoding in a delayed match-to-sample task. *Frontiers in Human Neuroscience*, *15*, Article 750582. http://doi.org/10.3389/fnhum.2021.750582

Geary, D. C., Hoard, M. K., Nugent, L., & Byrd-Craven, J. (2008). Development of number line representations in children with mathematical learning disability. *Developmental Neuropsychology*, *33*(3), 277–299. https://doi.org/10.1080/87565640801982361

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*(6), Article e67374. https://doi.org/10.1371/journal.pone.0067374

Halberda, J., Mazzocco, M. M. M., & Feigenson, L. (2008). Individual differences in non-verbal number acuity correlate with maths achievement. *Nature*, *455*(7213), 665–668. https://doi.org/10.1038/nature07246

Kaufmann, L., Mazzocco, M. M., Dowker, A., von Aster, M., Göbel, S., Grabner, R. H., Henik, A., Jordan, N. C., Karmiloff-Smith, A. D., Kucian, K., Rubinsten, O., Szücs, D., Shalev, R., & Nuerk, H. C. (2013). Dyscalculia from a developmental and differential perspective. *Frontiers in Psychology*, *4*, Article 516. https://doi.org/10.3389%2Ffpsyg.2013.00516

Korkman, M., Kirk, U., & Kemp, S. (2007). *NEPSY-II: A developmental neuropsychological assessment.* The Psychological Corporation.

Kosc, L. (1974). Developmental dyscalculia. *Journal of Learning Disabilities*, *7*(3), 164–177. https://doi.org/10.1177%2F002221947400700309

Landerl, K., Bevan, A., & Butterworth, B. (2004). Developmental dyscalculia and basic numerical capacities: A study of 8-9-year-old students. *Cognition*, *93*(2), 99–125. https://doi.org/10.1016/j.cognition.2003.11.004

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*(3), 309–324. https://doi.org/10.1016/j.jecp.2009.03.006

Lyons, I. M., Price, G. R., Vaessen, A., Blomert, L., & Ansari, D. (2014). Numerical predictors of arithmetic success in grades 1-6. *Developmental Science*, *5*, 714–726. https://doi.org/10.1111/desc.12152

Mammarella, I. C., Toffalini, E., Caviola, S., Colling, L., & Szűcs, D. (2021). No evidence for a core deficit in developmental dyscalculia or mathematical learning disabilities. *Journal of Child Psychology and Psychiatry*, *62*(6), 704–714. https://doi.org/10.1111%2Fjcpp.13397

Morey, R. D., Rouder, J. N., & Jamil, T. (2015). *BayesFactor: Computation of Bayes factors for common designs* (R package Version 0.9.12-2). https://rdrr.io/cran/BayesFactor/

Morsanyi, K., van Bers, B. M., McCormack, T., & McGourty, J. (2018). The prevalence of specific learning disorder in mathematics and comorbidity with other developmental disorders in primary school-age children. *British Journal of Psychology*, *109*, 917–940. https://doi.org/10.1111/bjop.12322

Perry, C., Zorzi, M., & Ziegler, J. C. (2019). Understanding dyslexia through personalized large-scale computational models. *Psychological Science*, *30*(3), 386–395. https://doi.org/10.1177%2F0956797618823540

Peters, L., & Ansari, D. (2019). Are specific learning disorders truly specific, and are they disorders? *Trends in Neuroscience and Education*, *17*, Article 100115. https://doi.org/10.1016/j.tine.2019.100115

Piazza, M., De Feo, V., Panzeri, S., & Dehaene, S. (2018). Learning to focus on number. *Cognition*, *181*, 35–45. https://doi.org/10.1016/j.cognition.2018.07.011

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*(1), 33–41. https://doi.org/10.1016/j.cognition.2010.03.012

Raftery, A. E. (1995). Bayesian model selection in social research. *Sociological Methodology*, *25*, 111–163. https://doi.org/10.2307%2F271063

Reeve, R., Reynolds, F., Humberstone, J., & Butterworth, B. (2012). Stability and change in markers of core numerical competencies. *Journal of Experimental Psychology: General*, *141*(4), 649–666. https://doi.org/10.1037/a0027520

Reigosa-Crespo, V., Valdés-Sosa, M., Butterworth, B., Estévez, N., Rodríguez, M., Santos, E., Torres, P., Suárez, R., & Lage, A. (2012). Basic numerical capacities and prevalence of developmental dyscalculia: The Havana Survey. *Developmental Psychology*, *48*(1), 123–135. https://doi.org/10.1037/a0025356

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*(3), 361–395. https://doi.org/10.1016/j.cognition.2006.01.005

Sasanguie, D., & Vos, H. (2018). About why there is a shift from cardinal to ordinal processing in the association with arithmetic between first and second grade. *Developmental Science*, *21*(5), Article e12653. https://doi.org/10.1111/desc.12653

Schneider, M., Beeres, K., Coban, L., Merz, S., Schmidt, S. S., Stricker, J., & De Smedt, B. (2017). Associations of non-symbolic and symbolic numerical magnitude processing with mathematical competence: A meta-analysis. *Developmental Science*, *20*, Article e12372. https://doi.org/10.1111/desc.12372

Schneider, W., Eschman, A., & Zuccolotto, A. (2012). *E-Prime user's guide*. Psychology Software Tools.

Sella, F., Lanfranchi, S., & Zorzi, M. (2013). Enumeration skills in Down syndrome. *Research in Developmental Disabilities*, *34*(11), 3798–3806. https://doi.org/10.1016%2Fj.ridd.2013.07.038

Sella, F., Sasanguie, D., & Reynvoet, B. (2020). Judging the order of numbers relies on familiarity rather than activating the mental number line. *Acta Psychologica*, *204*, Article 103014. https://doi.org/10.1016/j.actpsy.2020.103014

Siegler, R. S., & Opfer, J. E. (2003). The development of numerical estimation: Evidence for multiple representations of numerical quantity. *Psychological Science*, *14*(3), 237–243. https://doi.org/10.1111/1467-9280.02438

Szűcs, D., Devine, A., Soltesz, F., Nobes, A., & Gabriel, F. (2013). Developmental dyscalculia is related to visuo-spatial memory and inhibition impairment. *Cortex*, *49*(10), 2674–2688. https://doi.org/10.1016/j.cortex.2013.06.007

Vanbinst, K., Ghesquière, P., & Smedt, B. D. (2012). Numerical magnitude representations and individual differences in children's arithmetic strategy use. *Mind, Brain, and Education*, *6*(3), 129–136. https://doi.org/10.1111/j.1751-228X.2012.01148.x

Vandierendonck, A. (2017). A comparison of methods to combine speed and accuracy measures of performance: A rejoinder on the binning procedure. *Behavior Research Methods*, *49*(2), 653–673. https://doi.org/10.3758%2Fs13428-016-0721-5

Wechsler, D. (2003). *Wechsler Intelligence Scale for Children* (4th ed.). The Psychological Corporation.