Research in Developmental Disabilities 136 (2023) 104478

Contents lists available at ScienceDirect

Research in Developmental Disabilities

journal homepage: www.elsevier.com/locate/redevdis

Low discriminative power of WISC cognitive profile in developmental dyscalculia

Maristella Lunardon [a], Gisella Decarli [b,1], Francesco Sella [c], Silvia Lanfranchi [d], Silvia Gerola [e], Giuseppe Cossu [e], Marco Zorzi [b,f,*]

[a] Neuroscience Area, Scuola Internazionale Superiore di Studi Avanzati (SISSA), Trieste, Italy
[b] Department of General Psychology and Padova Neuroscience Center, University of Padova, Italy
[c] Centre for Mathematical Cognition, Loughborough University, UK
[d] Department of Developmental Psychology and Socialisation, University of Padova, Italy
[e] Centro Medico di Foniatria, Padova, Italy
[f] IRCCS San Camillo Hospital, Venice-Lido, Italy

ARTICLE INFO

Keywords:
Developmental dyscalculia
Math disability
Specific learning disabilities
Diagnosis

ABSTRACT

Background: The role of domain-general cognitive abilities in the etiology of Developmental Dyscalculia (DD) is a hotly debated issue.
Aims: In the present study, we tested whether WISC-IV cognitive profiles can be useful to single out DD.
Methods and procedures: Using a stringent 2-SD cutoff in a standardized numeracy battery, we identified children with DD ($N = 43$) within a clinical sample referred for assessment of learning disability and compared them in terms of WISC cognitive indexes to the remaining children without DD ($N = 100$) employing cross-validated logistic regression.
Outcomes and results: Both groups showed higher Verbal Comprehension and Perceptual Reasoning than Working Memory and Processing Speed, and DD scores were generally lower. Predictive accuracy of WISC indexes in identifying DD individuals was low (AUC = 0.67) and it dropped to chance level in discriminating DD from selected controls ($N = 43$) with average math performance but matched on global IQ. The inclusion of a visuospatial memory score as an additional predictor did not improve classification accuracy.
Conclusions and implications: These results demonstrate that cognitive profiles do not reliably discriminate DD from non-DD children, thereby weakening the appeal of domain-general accounts.

1. Introduction

Developmental Dyscalculia (DD) (or Mathematical Learning Disability, MLD) is a neurodevelopmental disorder characterized by significantly low mathematical skills. According to the DSM-5 (American Psychiatric Association, APA, 2013), DD belongs to the category of Specific Learning Disabilities. The difficulties in arithmetic, calculation, number sense and math reasoning must persist for at least 6 months after instruction begins, cannot be attributed to other mental, neurological, sensory or motor conditions, emerge

* Corresponding author at: Department of General Psychology, Via Venezia 8, 35131 Padova, Italy.
E-mail address: marco.zorzi@unipd.it (M. Zorzi).
[1] Present address: Department of Psychology and Cognitive Science, University of Trento, Italy.

https://doi.org/10.1016/j.ridd.2023.104478
Received 29 June 2022; Received in revised form 2 February 2023; Accepted 8 March 2023

during the school years and lead to academic skills significantly below to what expected according to chronological age. On the other hand, in the ICD-11[2] (World Health Organization, WHO, 2019/2021) the “Developmental learning disorder with impairment in mathematics” represents a distinct category from the other learning disorders, but its description resembles that of Specific learning disabilities, with the exception that ICD-11 stresses that academic skills should be well below the expectations according to age and general intellectual level.

The prevalence of DD is estimated between 5% and 7% (Butterworth et al., 2011; Morsanyi et al., 2018), though the lack of international agreement on the diagnosis and the use of widely different inclusion criteria is a recurring issue in DD/MLD research (Kucian & von Aster, 2015; Murphy et al., 2007; Peters & Ansari, 2019; Shalev & von Aster, 2018).

Several studies have highlighted that DD is a heterogeneous disorder including both domain-specific (numerical) and domain-general (non-numerical) deficits, and whether these deficits are predictive of DD is still under debate (e.g., Mammarella et al., 2021). Children with DD present difficulties in processing both non-symbolic (e.g., dots comparison; Decarli et al., 2020; Libertus et al., 2013; Mazzocco et al., 2011; Piazza et al., 2010) and symbolic numerical quantities (e.g., digit comparison and number ordering; Decarli et al., 2023; De Smedt & Gilmore, 2011; Iuculano et al., 2008; Rousselle & Noël, 2007). Conversely, other studies reported domain-general cognitive deficits in children with DD/MLD including processing speed, working memory and inhibition (Attout & Majerus, 2015; Gashaj et al., 2019; Mammarella et al., 2018; Peng et al., 2018; Rotzer et al., 2009; Szucs et al., 2013). Specifically, processing speed plays a crucial role in learning to count and fluently solve arithmetic problems (Geary, 1993), working memory allows to keep arithmetic partial results active while implementing the solving algorithm (e.g., Berg, 2008; Fung & Swanson, 2017) and inhibition processes reduce the interference of task-irrelevant features (Szucs et al., 2013; but see also Castaldi et al., 2018). Visuo-spatial skills are also considered a key component of DD as they show correlation with mathematical performance (Gilmore et al., 2013; Szucs et al., 2013).

In this vein, scales that assess multiple domain-general cognitive abilities, such as the Wechsler Intelligence Scale for Children (WISC; Wechsler, 2003), can provide useful information on the cognitive profiles of children with learning disabilities (Dickerson Mayes et al., 1998; Hale et al., 2001; Harrison & Armstrong, 2014; Styck & Watkins, 2016; Van Iterson & Kaufman, 2009; Watkins & Worrell, 2000; Watkins et al., 1997). Besides the estimation of the full-scale intelligence quotient (FSIQ), the fourth edition of the WISC (Wechsler, 2003) provides indexes for verbal comprehension, perceptual reasoning, working memory and processing speed. These measures can also be combined to obtain further composite scores, such as the General Ability Index (GAI), which includes verbal comprehension and perceptual reasoning, and the Cognitive Proficiency Index (CPI), which includes working memory and processing speed. A large discrepancy between relative high GAI and low CPI characterizes children with learning disabilities compared to typically developing children (Poletti, 2016; Giofrè et al., 2017) and to children with intellectual disabilities (Cornoldi et al., 2014).

Whether WISC indexes could be helpful in discriminating severe mathematical deficits from other learning difficulties has been overlooked. However, comparisons between children with DD and with other diagnosed learning disorders have been reported in previous studies. Compared to reading and writing disorders, DD presents lower FSIQ (Poletti, 2016; Poletti et al., 2018; Toffalini et al., 2017), suggesting a more pronounced domain-general impairment compared to other specific learning disabilities. Nevertheless, cognitive indexes present specific variations between DD and other learning disabilities. Children with DD showed lower GAI and, when considering single indices, lower perceptual reasoning and working memory scores than children with reading disorders (Poletti, 2016). Moreover, in DD, the perceptual reasoning index was lower than the verbal comprehension index, while the opposite was observed in the reading disorder group (Giofrè et al., 2022; Toffalini et al., 2017). Low math proficiency was also associated with low scores in reasoning, working memory, verbal comprehension, and processing speed indexes in children with learning disabilities (Poletti, 2016; Poletti et al., 2018).

To sum up, children with DD seem to present a distinct profile in terms of the domain-general cognitive abilities addressed by the WISC compared to the other specific learning disorders. However, it is yet to be demonstrated whether this profile allows discrimination of children with severe mathematical deficits, who receive the diagnosis of DD, from children without a mathematical impairment but who manifest other learning difficulties. Although the WISC-IV battery was not designed for the diagnosis of specific learning disabilities, detecting an index, or a combination of indexes, that reliably distinguish DD from non-DD children would provide the clinicians with an additional marker to support the diagnosis and, on the other hand, would inform the theory on the etiology of DD supporting the domain-general impairment account. Therefore, in the present study we tested whether cognitive profiles can be useful to single out children with DD from a sample of children referred to a clinical center for assessment of learning disabilities. We selected children with severe deficits in mathematical performance (i.e., below 2 *SD*s in a standardized numeracy battery) as the DD group and the remaining children were assigned to the non-DD group. Children with an average FSIQ were included in the study. A normative profile is characterized by a score of 100 in all indexes. However, as the reviewed literature showed, children with specific learning disorders present a specific pattern of weaknesses and strengths. The weaknesses are compensated by the strengths, resulting in an average FSIQ. Additionally, as it was previously demonstrated (e.g., Cornoldi et al., 2014), patterns of weaknesses and strengths in the WISC profile emerge also when the general cognitive level is in the normal range. Moreover, from the control group we also selected a subsample showing average math performance (thereby increasing the gap with DD) and matched for global IQ to the DD sample. In this way, we controlled for the difference in global cognitive functioning as a possible confound for the different profiles displayed by DD and control children. Differently from many other studies, we applied stringent criteria for the selection of children with DD (see Peters & Ansari, 2019, for discussion) and involved a control group with learning difficulties, thus improving ecological validity. We

[2] In the present study, children’s diagnosis was based on ICD-10 (World health organization, 1992) criteria because this was the classification system used at the time of our experiment. ICD-10 criteria do not significantly differ from the ICD-11 edition.

compared the two groups on the WISC-IV cognitive indexes, determined the multivariate pattern that best discriminated DD from non-DD, and tested the predictive accuracy of the classifiers trained to identify DD from the cognitive profile.

2. Material and methods

2.1. Participants

One hundred and forty-three children took part in the present study after obtaining parents' informed consent. The protocol was approved by the Psychological Science Ethics Committee of University of Padova and the study was conducted following the principles of the Declaration of Helsinki. From March 2019 to October 2020, we consecutively recruited all children referred to a specialized clinic for evaluation of learning disability, who attended a school grade from the third year of primary school to the third year of middle school. According to the ICD-10, which was the reference for diagnosis, we excluded children with WISC-IV (Wechsler, 2003; Italian version, Orsini et al., 2012) FISQ below the normal range (i.e., below 85, Flanagan & Kaufman, 2009). Other inclusion criteria were absence of ADHD, neurological, sensory or motor impairments. Children were administered a comprehensive standardized numeracy battery (BDE-2, Biancardi et al., 2016) and we assigned to the DD group ($N = 43$) those performing 2 *SD*s below the expected overall mean (considering age and grade), or performing below the 5th percentile in half or more of the battery's subtests. The remaining children were assigned to the non-DD group ($N = 100$). According to the ICD-10 coding system (World Health Organization, 1992), 14 children in the non-DD group received diagnosis of specific reading disorder (F81.0), 7 a diagnosis of specific spelling disorder (F81.1), 12 presented both in comorbidity, and 5 children were diagnosed with mixed or other disorder of scholastic skills (F81.3, F81.8). The remaining children manifested learning difficulties, which led them to the attention of the clinical service, but the assessment did not result in a diagnosis of neurodevelopmental disorder. In the DD group, 18 children presented comorbidity with reading disorder, spelling disorder or both, 5 children received a diagnosis of mixed disorder of scholastic skills (F81.3) and 9 children manifested a difficulty in another learning domain, besides the DD, not severe enough for a comorbid diagnosis.

The DD and non-DD groups (DD: 26 females, 17 males; Control: 30 females, 70 males) did not differ for chronological age ($t(141) = 1.10$, $p = .27$, BF_{10}[3] $= 0.33$, moderate evidence in favor of H0), whereas differences emerged in the total FSIQ score ($t(141) = 4.06$, $p < .001$, $BF_{10} = 260.79$, extreme evidence in favor of H1) and in mathematical achievement (Total Score of BDE-2: $t(108) = 16.06$, $p < .001$, $BF_{10} = 8.05 * 10^{26}$, extreme evidence in favor of H1; Percentage of tasks below the 5th percentile in the BDE-2: $t(31) = -9.60$, $p < .001$, $BF_{10} = 5.75 * 10^{7}$, extreme evidence in favor of H1). Descriptive statistics are reported in Table 1.

We also extracted from the non-DD group a subset of participants with a BDE-2 total score greater or equal than 85 or with less than 25% of the tasks below the 5th percentile, thereby showing average numeracy skills (Biancardi et al., 2016; see the Tasks section for the description of the battery) and paired for FISQ with the DD children. The FSIQ-matched non-DD group (hereafter FSIQ-non-DD) included 43 participants (8 females, 35 males) who did not differ from DD in FSIQ ($M = 100.30$, $SD = 6.91$, $t(84) = 1.94$, $p = .06$, $BF_{10} = 1.16$, anecdotal evidence in favor of H1) and age ($M_{\text{months}} = 133.58$, $SD = 18.77$, $t(84) = 0.96$, $p = .34$, $BF_{10} = 0.34$, moderate evidence in favor of H0), whereas they showed an average mathematical performance (Total Score of BDE-2: $M = 98.22$, $SD = 8.21$; $t(66) = 22.62$, $p < .001$, $BF_{10} = 1.66 * 10^{29}$, extreme evidence in favor of H1; Percentage of tasks below the 5th percentile: $M = 4.55$, $SD = 6.31$; $t(16) = -13.66$, $p < .001$, $BF_{10} = 1.58 * 10^{7}$, extreme evidence in favor of H1). The standardized difference between DD and FISQ-non-DD on the matching variable was $d = 0.42$ (C.I. [−0.01; 0.85]) and the variance ratio was 1.21, indicating that the groups are adequately matched according to the rules described by Kover and Atwood (2013).

2.2. Tasks

All children in the sample completed standardized cognitive and numeracy assessments, as well as ad-hoc computerized numerical tasks. The latter can provide a fine-grained evaluation of domain-specific deficits but are largely redundant with the numeracy battery. In this study, we focused on the WISC cognitive profile, overall numeracy skills and, as a control, visuo-spatial skills. The tasks used for the assessment of reading and writing skills are briefly described in the Supplementary material. For all tasks, normative scores and indexes were calculated based on the Italian standardization.

2.2.1. Numeracy

We assessed numeracy with a standardized battery for the diagnosis of DD (BDE-2, Biancardi et al., 2016). It includes several subtests for assessing three main areas: general number knowledge (counting, number reading, number writing and number repetition), calculation (multiplications, mental calculation, calculation fluency, multiplication table and written calculation) and number sense (triplets of numbers, insertions of Arabic digits on a number line, rapid calculation, number line). For fourth-to-thirteenth grade children, the raw score of each subtest was converted into normative score ($M = 10$, $SD = 3$) and the sum of all these scores composed the total quotient ($M = 100$, $SD = 15$). For third graders, the test manual does not provide conversion tables to compute the total quotient, so for each child we computed the percentage of subtests whose score fell below the 5th percentile, according to the Italian guidelines for the diagnosis of DD (Lucangeli, 2012).

[3] Bayes factor (BF_{10}) represents the probability of the data given the alternative hypothesis (H1) relative to the null hypothesis (H0). In the results we reported the evidence associated with BFs as "anecdotal" ($1/3 < BF_{10} < 3$), "moderate" ($BF < 1/3$ or $BF_{10} > 3$), "strong" ($BF < 1/10$ or $BF_{10} > 10$), "very strong" ($BF < 1/30$ or $BF_{10} > 30$), or "extreme" ($BF < 1/100$ or $BF_{10} > 100$) (Jeffreys, 1961).

Table 1
Mean, standard deviation (SD), 95% confidence interval (C.I.) and range for age (year; months), BDE scores and cognitive indexes scores for non-DD and DD groups.

	non-DD (*N* = 100)				DD (*N* = 43)			
	Mean	SD	C.I.	Range	Mean	SD	C.I.	Range
Age	11;2	1;9	129.63–137.93	104–176	10;10	1;7	123.97–135.47	97–172
BDE-2								
Total score	92.64	13.00	89.63 – 95.65	71 – 121	55.06	7.52	52.52 – 57.60	49–69
% of tasks below the 5th perc.	11.54	12.21	6.61–16.47	0 – 37.50	60.71	11.25	50.31 – 71.11	50–75
WISC indexes								
VCI	108.02	12.79	105.48–110.56	82–140	101.35	12.34	97.55–105.15	78–124
PRI	111.59	12.31	109.15–114.03	78–143	107.05	11.40	103.54–110.56	76–130
WMI	92.47	12.60	89.97–94.97	64–127	85.28	11.37	81.78–88.78	61–115
PSI	98.86	14.32	96.02–101.70	56–132	92.74	10.17	89.61–95.87	76–118
FSIQ	105.02	11.49	102.74–107.30	85 – 140	97.26	7.61	94.92–99.6	86–124
GAI	110.57	11.36	107.84–113.30	85–143	104.77	9.70	102.00–107.54	88–130
CPI	105.02	11.49	102.74–107.30	85–140	86.26	9.00	83.27–89.25	70–113

Note. VCI = Verbal comprehension index; PRI = Perceptual reasoning index; WMI = Working memory index; PSI = Processing speed index; FSIQ = Full scale IQ; GAI = General ability index; CPI = Cognitive proficiency index.

2.2.2. Intelligence

Children were assessed with the Wechsler Scale of Intelligence (WISC-IV; Wechsler, 2003; Italian version, Orsini et al., 2012). For each child, the full FSIQ and the four composite indexes were calculated: Verbal Comprehension (VCI: Similarities, Vocabulary and Comprehension subtest); Perceptual Reasoning (PRI: Block Design, Picture Concepts and Matrix Reasoning subtest); Working Memory (WMI: Digit Span and Letter-Number sequencing subtest); Processing Speed (PSI: Coding and Symbol Search subtest).

2.2.3. Visuo-spatial memory

We assessed visuo-spatial memory using the *Memory for Designs task (Immediate)* (MDI) from the NEPSY-II (Korkman et al., 2007). Children were instructed to memorize a grid (21 × 29.7 cm) with some pictures (from 6 to 10) for 10 s; then children had to choose the pictures they had seen among a set of cards (from 10 to 20) and to locate them on an empty grid in the same position they had previously memorized. For each child, we extracted the total score obtained from the number of figures and positions correctly remembered by the child. The maximum score was 150, corresponding to a performance without mistakes.

2.3. Data analysis

We ran the statistical analyses using R v. 4.1.1 (R Core Team, 2020).

First, we compared the performance of the groups on the cognitive indexes by means of a one-way MANOVA, with the four indexes as dependent variables and Group [DD vs. non-DD] as factor. We ran post-hoc two-tailed *t*-tests to investigate significant MANOVA effects. We adjusted the alpha value to 0.0125 according to Bonferroni correction for multiple comparisons (alpha = 0.05/4). Assumptions for MANOVA (i.e., normality, linearity, lack of influential outliers, homogeneity of variance and absence of multicollinearity) were met (Tabachnick & Fidell, 2007). Additionally, we conducted a mixed ANOVA with Group [DD vs. non-DD] as between factor and Index [GAI vs. CPI vs. FSIQ] as within-subject factor to assess the presence of the GAI-CPI discrepancy and the GAI-FSIQ discrepancy in the two groups.

Then, to find among the four WISC indexes those that could better predict the level of numeracy achievement, we entered them as predictors in a series of logistic regression models with Group [DD vs. non-DD] as outcome variable. Assumptions for logistic regression (lack of influential outliers, absence of missing values and of multicollinearity) were met (Tabachnick & Fidell, 2007). We used a best subset regression strategy (Hosmer & Lemeshow, 2000), whereby all possible combinations of indexes are assessed and ranked based on the Bayesian Information Criterion (BIC). The best model (i.e., with the lowest BIC) was then selected (Vrieze, 2012).

Subsequently, we tested the predictive accuracy of the selected models using leave-one-out cross-validation to train the logistic classifiers and assess their performance on the held-out children (Browne, 2000). The receiver operating characteristic (ROC) curve and its related area under the curve (AUC) were computed for each selected model. AUC is a threshold-invariant measure of classification performance, whereby values of 0.5 to 0.7 indicate low discriminative accuracy, 0.7 to 0.9 moderate accuracy, and 0.9 to 1.0 high accuracy (Swets, 1988).

We also used the ROC curves to determine the optimal classification threshold for each model using the Youden index (i.e., the best tradeoff between specificity and sensitivity) and computed the resulting confusion matrices in terms of true positives/negatives and false positives/negatives (Youden, 1950).

We repeated the above analyses to compare children with DD to a group of non-DD children matched for FSIQ and with average math performance. We used the *matchControls* function in R (Meyer et al., 2021) to select, for each DD child, the non-DD child with the nearest FSIQ value.

Additionally, we re-trained the logistic classifiers including a visuospatial memory score as an additional predictor (MDI), because previous research has suggested this domain-general component being significantly lower in children with DD.

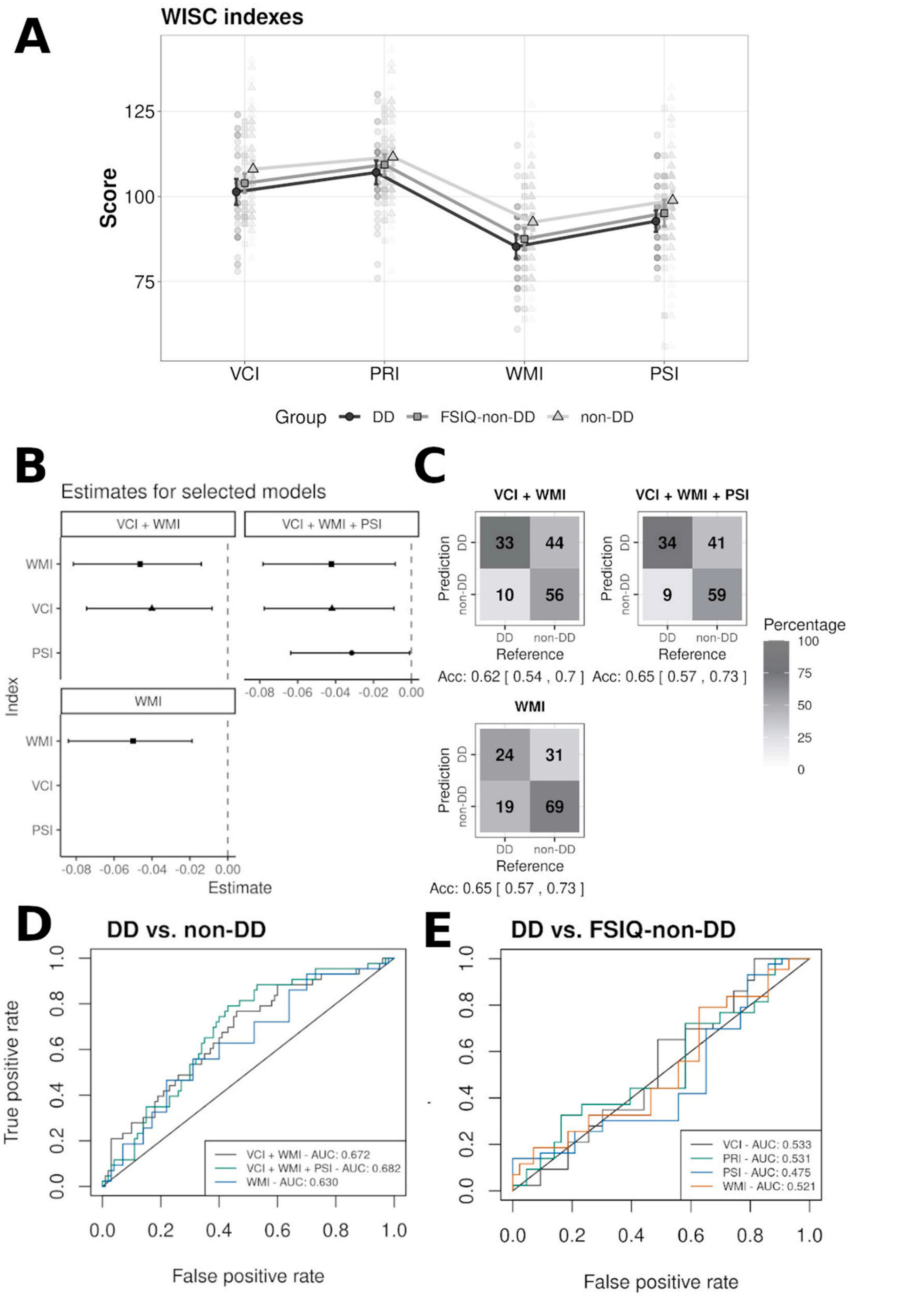

Fig. 1. The WISC cognitive profiles of DD, non-DD and FSIQ-non-DD groups and the classification performance of the selected models. VCI = Verbal comprehension index; PRI = Perceptual reasoning index; WMI = Working memory index; PSI = processing speed index. (A) Mean scores of DD, non-

DD and FSIQ-non-DD groups in each cognitive index. Error bars represent 95% confidence intervals. Transparent points represent individual scores. (B) Regression coefficients of the WISC indexes in the selected models (DD vs. non-DD classification). Error bars represent 95% confidence intervals. (C) Confusion matrices and accuracies (Acc) for the selected models (DD vs non-DD). The numbers within each cell represent how many observed cases on the y-axis are labeled as a group on the x-axis. (D) ROC curves and AUCs for the selected models (DD vs non-DD). (E) ROC curves and AUCs for the selected models (DD vs. FSIQ-non-DD).

3. Results

3.1. Indexes comparisons

Descriptive statistics for the four WISC indexes for the two groups are reported in Table 1 and displayed in Fig. 1a. The MANOVA on the four WISC indexes yielded a significant effect of Group (Pillai's trace = 0.12, $F(4,138) = 4.87$, $p = .001$, $\eta_p^2 = 0.12$). In particular, the visual inspection of Fig. 1a suggests that the DD and non-DD groups presented a similar pattern of indexes, even though the non-DD group had higher scores: the difference resulted statistically significant for VCI ($t(141) = -2.89$, $p = .004$, $BF_{10} = 7.95$, moderate evidence in favor of H1), WMI ($t(141) = -2.07$, $p = .002$, $BF_{10} = 19.08$, strong evidence in favor of H1) and PSI ($t(141) = -2.54$, $p = .012$, $BF_{10} = 3.43$, moderate evidence in favor of H1).

In the mixed ANOVA assessing GAI, CPI and FSIQ discrepancies (descriptive statistics in Table 1), a significant main effect of Group ($F(1, 423) = 37.89$, $p < .001$) and Index ($F(2, 423) = 78.82$, $p < .001$) emerged, while their interaction was not significant ($F(2, 423) = 0.45$, $p = .693$), suggesting that the indexes' discrepancies were consistent across groups. We conducted two post-hoc t-tests to assess the GAI-CPI and GAI-FSIQ discrepancies (Bonferroni corrected alpha value = 0.05/2 = 0.025) and they resulted significant, thus suggesting that GAI was higher than CPI ($t(284) = 11.57$, $p < .001$, $BF_{10} = 2.21 \times 10^{22}$, extreme evidence in favor of H1) and FSIQ ($t(284) = 4.67$, $p < .001$, $BF_{10} = 3217$, extreme evidence in favor of H1).

Comparisons for single subtests (see Supplementary material) showed a group difference in Digit span (WMI).

We then compared the WISC indexes between DD children and the group of non-DD children matched for FSIQ and with average mathematical performance. Fig. 1a shows that the pattern of cognitive indexes did not change and the DD and FSIQ-non-DD profiles were largely overlapping (index values of FSIQ-non-DD: VCI: $M = 103.91$, $SD = 9.14$, C.I. = [101.10, 106.72], range = [86,122]; PRI: $M = 109.33$, $SD = 9.87$, C.I. = [106.29, 112.37], range = [87,128]; WMI: $M = 87.51$, $SD = 10.64$, C.I. = [84.24, 90.78], range = [64,106]; PSI: $M = 95.09$, $SD = 12.80$, C.I. = [91.15, 99.03], range = [56,126]; MANOVA: Pillai's trace = 0.05, $F(4,81) = 1.03$, $p = .398$, $\eta_p^2 = 0.05$). As previously found, GAI resulted higher than CPI and FSIQ regardless of Group (values for FSIQ-non-DD: GAI: $M = 107.14$, $SD = 7.74$, C.I. = [103.72, 110.56], range = [86,126]; CPI: $M = 86.56$, $SD = 11.12$, C.I. = [87.18, 91.94], range = [70,111]; mixed ANOVA: Group: $F(1,252) = 7.04$, $p = .008$; Index: $F(2,252) = 91.74$, $p < .001$; interaction: $F(2,252) = 0.06$, $p = .934$; post-hoc t-tests: GAI vs. CPI: $t(170) = 12.43$, $p < .001$, $BF_{10} = 2.12 \times 10^{22}$, extreme evidence in favor of H1; GAI vs. FSIQ: $t(170) = 5.79$, $p < .001$, $BF_{10} = 3.30 \times 10^{5}$, extreme evidence in favor of H1). No group difference emerged at subtest level (see Supplementary material).

3.2. Classifier analyses

We selected three logistic regression models because their associated BIC values were comparable (i.e., a difference in BIC lower than 2; Raftery, 1995): the model including VCI and WMI (BIC: 173.25), the model with VCI, WMI and PSI (BIC: 174.12), and the model with just WMI (BIC: 174.46). Estimates of the index effect in each model are displayed in Fig. 1b and their ROC curves and AUCs are illustrated in Fig. 1d. The AUC of the model with VCI+WMI was 0.672 and it indicated low accuracy in discriminating between children with and without DD. The other AUCs values were 0.682 and 0.630 for the models VCI+WMI+PSI and WMI respectively, again indicating low classification accuracy.

Fig. 1c shows the confusion matrices computed with the optimal classification threshold obtained from the ROC curves. The model with VCI and WMI as predictors displayed an accuracy of 0.62 and it correctly detected 77% of the DD children, but the false positive rate was very high as the classifier misdiagnosed 44% of the non-DD children. The accuracy of the model with VCI, WMI and PSI was 0.65 and it identified 79% of the children with DD but misclassified 41% of non-DD children. Finally, the model with WMI also displayed an accuracy of 0.65 and it accurately categorized only 56% of the children with DD and 69% of children in the non-DD group.

Though the four WISC indexes did not differ between DD and FSIQ-non-DD, we entered them as predictors in a series of logistic regression analyses on Group [DD vs. FSIQ-non-DD]. Four models, each including a single index, yielded comparably low BICs: VCI (BIC: 126.91), PRI (BIC: 127.13), PSI (BIC: 127.22) and WMI (BIC: 127.23). The cross-validated logistic classifiers failed to discriminate between children with severe DD and non-DD children with average mathematical skills. The highest AUC was 0.525 for the model with PSI, which had an accuracy of 0.57. The other models showed even lower accuracies, with AUCs ranging from 0.467 to 0.479 (see ROC curves in Fig. 1e).

Finally, we retrained the classifiers including MDI. The mean MDI scores were 116.72 ($SD = 22.54$, C.I. = [109.78, 123.66], range = [74,147]) for DD, 123.81 ($SD = 23.18$, C.I. = [119.21, 128.41], range = [66,150]) for non-DD, and 120.65 ($SD = 23.46$, C.I. = [113.43, 127.87], range = [66,150]) for FSIQ-non-DD groups. The DD score did not differ from the scores of the other groups (for DD vs non-DD, $p = .093$, $BF_{10} = 0.70$, anecdotal evidence; for DD vs FSIQ-non-DD, $p = .430$, $BF_{10} = 0.30$ moderate evidence in favor of H0). The best models discriminating DD from non-DD children, according to their BIC value, were the same as those previously selected (i.e., MDI was never included); therefore, cross-validation and ROC analysis led to the same results. For the DD vs FSIQ-non-

DD discrimination, the model with just the MDI index was equivalent to the four previous models (BIC: 127.49) and was therefore retained. The cross-validated accuracy of the model with MDI as predictor was 0.51 and the AUC was 0.588, indicating the inability to discriminate between children with and without DD.

3.3. Control analyses using weighted logistic classifiers

The use of threshold-invariant AUC (as well as choosing an optimal classification threshold using the Youden index) is appropriate for reliably measuring classification accuracy even in the case of non-severe numerical imbalance between classes, as in the present case. Nevertheless, to further exclude class imbalance between DD and non-DD groups as cause for low classification accuracy, we also trained and assessed cross-validated weighted logistic classifiers (King & Zeng, 2001), which apply higher cost to wrong predictions of the minority class (here DD). The weighted classifiers (see Supplementary Figure 2) could correctly identify 72% of DD children and 57% of non-DD children when the predictors were VCI and WMI, 74% of DD and 60% of non-DD when adding PSI to the previous model, and 63% of DD and 60% of non-DD when WMI was the only predictor. The model with the highest AUC was the VCI + WMI + PSI (AUC: 0.724), indicating a limited discriminative power, as in the other two models (VCI + WMI: 0.709; WMI: 0.665).

3.4. Additional group comparisons

We compared the cognitive indexes of specific subgroups of our sample. Results are reported in the Supplementary material. When DD scores were contrasted to those of the non-DD children without a diagnosis of learning disorder, they differed in the FSIQ and the WMI (Supplementary table 2). Moreover, we compared the cognitive scores of children with a deficit in mathematics only and of children with a combined disorder in mathematics and in other learning domains, since previous research showed that the latter are likely to display lower cognitive levels (e.g., Toffalini et al., 2017). Contrary to the expectations, no significant difference emerged (Supplementary table 3).

Finally, we tested the accuracy of WISC indexes in discriminating children with a diagnosis of specific learning disorder from those who did not receive the diagnosis. The model where the group [learning disorder vs. no learning disorder] was predicted by WMI yielded the lowest BIC, but its predictive accuracy was low (AUC: 0.586). This classifier could correctly detect 94% of the children with a diagnosis of learning disorder, but misclassified 94% of the children without a diagnosis. These results may be due to the fact that all children in our sample, including those without a learning disorder, manifested learning difficulties that motivated referral for the clinical assessment.

4. Discussion

The finding that children with DD show weaknesses in domain-general cognitive abilities has led to theoretical accounts that point to deficits in general cognitive skills (rather than in domain-specific numerical processing) as etiology of mathematical disability (Kaufmann et al., 2013, for a review). Here we assessed whether cognitive indexes derived from the WISC scale can effectively identify children with DD from children without mathematical impairments but with difficulties in other learning domains. All children were referred to a clinical service for the diagnosis of specific learning disorders. In line with previous findings, children with DD displayed a lower general cognitive functioning compared to non-dyscalculic children (Poletti, 2016; Poletti et al., 2018; Toffalini et al., 2017). Both children with DD and non-DD displayed a similar pattern of cognitive indexes whereby the VCI and the PRI were higher than the WMI and PSI and presented the GAI-CPI discrepancy found in previous studies (Cornoldi et al., 2014; Giofrè et al., 2017; Poletti, 2016; Toffalini et al., 2017). Especially the PRI in both groups and the VCI in the non-DD group were higher than the normative mean. On one hand, this may be due to the fact that in the non-DD group there were children with a learning disorder. Since GAI is more strongly related to general cognitive functioning (the *g*-factor) than CPI, it has been suggested that FSIQ may underestimate the intellectual level of children with specific learning disabilities (Giofrè & Cornoldi, 2015). On the other hand, these above average values may be the consequence of our sample selection procedure, which included children with an FSIQ above 85, where the low WMI and PSI may be compensated by above average VCI and PRI.

To identify the combinations of cognitive indexes that better predicted the presence of DD, we ran a series of logistic regressions. We found that the best model was the one including VCI and WMI. However, the classifier's (cross-validated) predictive accuracy in discriminating DD from non-DD was low (AUC: 0.672), with a very high false positive rate in DD detection. The runner-up models (i.e., VCI+WMI+PSI, WMI) showed an equivalent (low) classification performance (AUCs: 0.682 and 0.630, respectively). When we tested discrimination of DD against a group of children with average mathematical performance but matched for FSIQ with DD children, all classifiers (also those including a measure of visuospatial working memory as additional predictor) exhibited very low to chance-level predictive accuracy (AUCs: 0.479 - 0.588).

The WISC is widely applied in the clinical setting to estimate general cognitive functioning and it has been proposed that the distribution of the cognitive indexes can contribute to discriminating neurodevelopmental disorders, especially specific learning disorders (Cornoldi et al., 2014; Giofrè et al., 2017; Poletti, 2016). Our study focused on DD and showed that children with severe mathematical underachievement present a global lower cognitive functioning compared to children referred for clinical evaluation of learning disabilities without math impairment, but they cannot be identified on the basis of the multivariate pattern of WISC indexes. Our results, therefore, suggest that the WISC cognitive profile does not provide a reliable cue to detecting DD children.

This is the first time that predictive modeling is applied to separate DD from children with other learning difficulties and with a data-driven approach to cognitive index selection. There were some previous attempts to single out learning disabilities from typical development based on WISC cognitive indexes. For instance, Giofrè et al. (2017) yielded an AUC of 0.78 for classification based on the four WISC-IV indexes and an AUC of 0.76 with the GAI and CPI. Even lower discriminative power resulted when indices extracted from the WISC-III were used as predictors, with AUCs ranging from 0.50 to 0.68 (Watkins et al., 1997; Watkins & Worrell, 2000). Our results are in line with these findings. WISC indexes provide information on cognitive processes that underlie academic achievement. However, they cannot provide reliable evidence for the diagnosis of specific learning disorders, and DD in particular. Nevertheless, this does not exclude their utility for a better definition of the clinical profile (also see Giofrè et al., 2017).

This study presents some limitations. First, the non-DD group is highly heterogeneous. Only 38 of these children received a diagnosis for specific learning disorders. The remaining children displayed learning difficulties not severe enough for a clinical diagnosis. It is possible then that WISC profile discriminative power may be limited to the comparison of DD to other specific learning disorders, such as reading disorder or writing disorder, as described by the literature reviewed in the introduction. Given the limited sample size, our data did not allow these categorical comparisons, which was however beyond our interests: the goal of this study was to detect WISC cognitive indexes that contribute to single out children with DD from children without DD, regardless of the type of learning difficulty in the latter group. The low scholastic performance of non-DD children who did not get a diagnosis could also be the consequence of other forms of difficulties, for example at socio-emotional level. Another limitation of our study is that we did not have information about participants' emotional states and family background, which could have contributed to clarify these cases.

Secondly, the sample size is relatively small compared to other studies addressing cognitive profiles in specific learning disabilities (e.g., Toffalini et al., 2017). However, it is considerably large for DD given the very stringent selection criteria, the same used by professionals for clinical diagnosis.

Thirdly, 23 out of 43 children in the DD group presented a comorbid learning disorder, and 38 out of 100 non-DD children were diagnosed with a specific learning disorder, thus possibly explaining their similar cognitive profiles. Therefore, our results should be interpreted with caution, since we did not consider in the logistic regressions "pure" DD separately from the cases of comorbidity between DD and other learning disorders (e.g., Mammarella et al., 2018), though the former are infrequent in the clinical setting and the contrasts run as supplementary analysis did not show differences between the two groups in any cognitive index. Nevertheless, the proportion of comorbidity in our DD group resembles the rates reported in previous studies (e.g., Morsanyi et al., 2018) and children with pure DD and those with comorbid dyslexia typically show the same profile in terms of numerical skills (e.g., Decarli et al., 2023). Moreover, the use of a standardized battery to assess numeracy skills ensured a reliable identification of mathematical under-achievement and the recruitment of participants among the population of children referred to a clinical service for the diagnosis of learning disorders guarantee the ecological validity of the findings.

An additional limitation is the exclusion of children with FSIQ below the normative mean. In the current version of the DSM-5, the achievement-IQ discrepancy is no longer a criterion for diagnosis of specific learning disorders. Moreover, excluding children with below average FSIQ reduced the variability in the distributions of the four cognitive indexes, which may have affected their predictive accuracy. However, in our view, the contrast between an average IQ and extremely low mathematical performance holds theoretical value for our research question regarding DD, as it allows to study the actual cognitive correlates of DD without the confound of borderline intellectual functioning (which is characterized by its own specific cognitive profile; e.g., Pulina et al., 2019). In addition, an average FSIQ could hide multivariate patterns of component scores, which were our target in this study. It was previously found that in children with a learning disorder, regardless of the fact that their IQ was higher or lower than 85, the GAI was higher than the CPI, while in children with intellectual disability (FSIQ < 70) the profile was flat, meaning that they scored approximately the same in all indexes (Cornoldi et al., 2014). Therefore, we could expect to find specific patterns of strengths and weaknesses in the WISC-IV components in DD children also when the general FSIQ is in the normal range.

In conclusion, our results provide further evidence that children with learning disorders tend to show cognitive profiles characterized by lower working memory and processing speed, compared to verbal comprehension and perceptual reasoning. However, these profiles do not discriminate DD from matched non-DD children, suggesting that it may be a common marker of specific learning disorders. Evaluating general cognitive skills is a fundamental step in the assessment of learning disabilities, which however does not contribute to characterizing the specific disorder components. Therefore, to reliably diagnose DD and provide adequate intervention, the focus should be mainly on domain-specific abilities, namely, numerical and arithmetic skills.

CRediT authorship contribution statement

Conceptualization: Maristella Lunardon, Gisella Decarli, Marco Zorzi. Methodology: All authors. Formal analysis: Maristella Lunardon, Gisella Decarli. Investigation: Maristella Lunardon, Gisella Decarli, Silvia Gerola. Writing - Original Draft: Maristella Lunardon, Gisella Decarli, Francesco Sella, Marco Zorzi. Writing - Review and Editing: All authors. Supervision: Silvia Lanfranchi, Giuseppe Cossu, Marco Zorzi. Project Administration: Marco Zorzi. Funding Acquisition: Marco Zorzi.

Conflict of Interest

None.

Data availability

The authors do not have permission to share data.

Acknowledgements

The study was supported by Cariparo Foundation (Italy) through Excellence Grant "NUMSENSE" to M.Z. We thank the staff of the Centro Medico di Foniatria for support in data collection, all children who took part in this study and their families.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.ridd.2023.104478.

References

American Psychiatric Association (2013). Diagnostic and statistical manual of mental disorders (5th ed.). Arlington, VA: American Psychiatric Publishing.

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

Berg, D. H. (2008). Working memory and arithmetic calculation in children: The contributory roles of processing speed, short-term memory, and reading. *Journal of Experimental Child Psychology, 99*(4), 288–308. https://doi.org/10.1016/j.jecp.2007.12.002

Biancardi, A., Bachmann, C., Nicoletti, 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. Trento, Italy: Edizioni Centro Studi Erickson.

Browne, M. W. (2000). Cross-validation methods. *Journal of Mathematical Psychology, 44*, 108–132. https://doi.org/10.1006/jmps.1999.1279

Butterworth, B., Varma, S., & Laurillard, D. (2011). Dyscalculia: From brain to education. *Science, 332*(6033), 1049–1053. https://doi.org/10.1126/science.1201536

Castaldi, E., Mirassou, A., Dehaene, S., Piazza, M., & Eger, E. (2018). Asymmetrical interference between number and item size perception provides evidence for a domain specific impairment in dyscalculia. *PloS One, 13*(12), Article e0209256.

Cornoldi, C., Giofrè, D., Orsini, A., & Pezzuti, L. (2014). Differences in the intellectual profile of children with intellectual vs. Learning disability. *Research in Developmental Disabilities, 35*(9), 2224–2230. https://doi.org/10.1016/j.ridd.2014.05.013

Decarli, G., Paris, E., Tencati, C., Nardelli, C., Vescovi, M., Surian, L., & Piazza, M. (2020). Impaired large numerosity estimation and intact subitizing in developmental dyscalculia. *PLOS ONE, 15*(12), Article e0244578. https://doi.org/10.1371/journal.pone.0244578

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*(2), 278–292. https://doi.org/10.1016/j.jecp.2010.09.003

Decarli, G., Sella, F., Lanfranchi, G., Gerotto, G., Gerola, S., Cossu, G., & Zorzi, M. (2023). Severe developmental dyscalculia is characterized by core deficits in both symbolic and nonsymbolic number sense. *Psychological Science, 34*(1), 8–21. https://doi.org/10.1177/09567976221097947

Dickerson Mayes, S., Calhoun, S. L., & Crowell, E. W. (1998). WISC-III profiles for children with and without learning disabilities. *Psychology in the Schools, 35*(4). https://doi.org/10.1002/(SICI)1520-6807(199810)35:4<309::AID-PITS1>3.0.CO;2-V

Flanagan, D.P., Kaufman, A.S. (2009). Essentials of WISC-IV assessment (2nd ed). Wiley.

Fung, W., & Swanson, H. L. (2017). Working memory components that predict word problem solving: Is it merely a function of reading, calculation, and fluid intelligence. *Memory & Cognition, 45*(5), 804–823. https://doi.org/10.3758/s13421-017-0697-0

Gashaj, V., Oberer, N., Mast, F. W., & Roebers, C. M. (2019). Individual differences in basic numerical skills: The role of executive functions and motor skills. *Journal of Experimental Child Psychology, 182*, 187–195. https://doi.org/10.1016/j.jecp.2019.01.021

Geary, D. C. (1993). Mathematical disabilities: Cognitive, neuropsychological, and genetic components. *Psychological Bulletin, 114*, 345–362. https://doi.org/10.1037/0033-2909.114.2.345

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

Giofrè, D., Allen, K., Toffalini, E., Mammarella, I. C., & Caviola, S. (2022). Decoding gender differences: Intellectual profiles of children with specific learning disabilities. *Intelligence, 90*, Article 101615. https://doi.org/10.1016/j.intell.2021.101615

Giofrè, D., & Cornoldi, C. (2015). The structure of intelligence in children with specific learning disabilities is different as compared to typically development children. *Intelligence, 52*, 36–43. https://doi.org/10.1016/j.intell.2015.07.002

Giofrè, D., Toffalini, E., Altoè, G., & Cornoldi, C. (2017). Intelligence measures as diagnostic tools for children with specific learning disabilities. *Intelligence, 61*, 140–145. https://doi.org/10.1016/j.intell.2017.01.014

Hale, J. B., Fiorello, C. A., Kavanagh, J. A., Hoeppner, J.-A. B., & Gaither, R. A. (2001). WISC-III predictors of academic achievement for children with learning disabilities: Are global and factor scores comparable. *School Psychology Quarterly, 16*(1), 31–55. https://doi.org/10.1521/scpq.16.1.31.19158

Harrison, A. G., & Armstrong, I. (2014). WISC-IV Unusual Digit Span Performance in a Sample of Adolescents With Learning Disabilities. *Applied Neuropsychology: Child, 3*(2), 152–160. https://doi.org/10.1080/21622965.2012.753570

Hosmer, D.W., Lemeshow, S. (2000). Applied logistic regression, Wiley.

Iuculano, T., Tang, J., Hall, C. W., & Butterworth, B. (2008). Core information processing deficits in developmental dyscalculia and low numeracy. *Developmental Science, 11*(5), 669–680. https://doi.org/10.1111/j.1467-7687.2008.00716.x

Jeffreys, H. (1961). Theory of probability (3rd ed.). Oxford: Oxford University Press, Clarendon Press.

Kaufmann, L., Mazzocco, M. M., Dowker, A., von Aster, M., Goebel, S., Grabner, R., & Nuerk, H. C. (2013). Dyscalculia from a developmental and differential perspective. *Frontiers in Psychology, 4*, 516. https://doi.org/10.3389%2Ffpsyg.2013.00516.

King, G., & Zeng, L. (2001). Logistic Regression in Rare Events Data. *Political Analysis, 9*(2), 137–163. https://doi.org/10.1093/oxfordjournals.pan.a004868

Korkman, M., Kirk, U. & Kemp, S. (2007). NEPSY Second Edition (C. Urgesi, F. Campanella & F. Fabbro, tran.). Firenze: Giunti O.S.

Kover, S. T., & Atwood, A. K. (2013). Establishing Equivalence: Methodological Progress in Group-matching Design and Analysis. *American Journal on Intellectual and Developmental Disabilities, 118*(1), 3–15. https://doi.org/10.1352/1944-7558-118.1.3

Kucian, K., & von Aster, M. (2015). Developmental dyscalculia. *European Journal of Pediatrics, 174*(1), 1–13. https://doi.org/10.1007/s00431-014-2455-7

Libertus, M. E., Feigenson, L., & Halberda, J. (2013). Is approximate number precision a stable predictor of math ability. *Learning and Individual differences, 25*, 126–133. https://doi.org/10.1016/j.lindif.2013.02.001

Lucangeli, D. (2012). La diagnosi di discalculia: il documento d'accordo AIRIPA-AID. *QI-Questioni e idee in psicologia, 3*.

Mammarella, I. C., Caviola, S., Giofrè, D., & Szűcs, D. (2018). The underlying structure of visuospatial working memory in children with mathematical learning disability. *British Journal of Developmental Psychology, 36*(2), 220–235. https://doi.org/10.1111/bjdp.12202

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/jc

Mazzocco, M. M., Feigenson, L., & Halberda, J. (2011). Impaired acuity of the approximate number system underlies mathematical learning disability (dyscalculia). *Child Development, 82*(4), 1224–1237. https://doi.org/10.1111/j.1467-8624.2011.01608.x

Meyer, D., Dimitriadou, E., Hornik, K., Weingessel, A., and Leisch, F. (2021). e1071: Misc Functions of the Department of Statistics, Probability Theory Group (Formerly: E1071), TU Wien. R package version 1.7–8. ⟨https://CRAN.R-project.org/package=e1071⟩.

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

Murphy, M. M., Mazzocco, M. M. M., Hanich, L. B., & Early, M. C. (2007). Cognitive Characteristics of Children With Mathematics Learning Disability (MLD) Vary as a Function of the Cutoff Criterion Used to Define MLD. *Journal of Learning Disabilities, 40*(5), 458–478. https://doi.org/10.1177/00222194070400050901

Orsini, A., Pezzuti, L., Picone, L. (2012). WISC-IV: Contributo alla taratura Italiana. [WISC-IV Italian edition]. Florence, Italy: Giunti O. S.

Peng, P., Wang, C., & Namkung, J. (2018). Understanding the cognition related to mathematics difficulties: A meta-analysis on the cognitive deficit profiles and the bottleneck theory. *Review of Educational Research, 88*(3), 434–476. https://doi.org/10.3102/0034654317753350

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., 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

Poletti, M. (2016). WISC-IV intellectual profiles in Italian children with specific learning disorder and related impairments in reading, written expression, and mathematics. *Journal of Learning Disabilities, 49*(3), 320–335. https://doi.org/10.1177/0022219414555416

Poletti, M., Carretta, E., Bonvicini, L., & Giorgi-Rossi, P. (2018). Cognitive clusters in specific learning disorder. *Journal of Learning Disabilities, 51*(1), 32–42. https://doi.org/10.1177/0022219416678407

Pulina, F., Lanfranchi, S., Henry, L., & Vianello, R. (2019). Intellectual profile in school-aged children with borderline intellectual functioning. *Research in Developmental Disabilities, 95*, Article 103498. https://doi.org/10.1016/j.ridd.2019.103498

R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL ⟨https://www.R-project.org/⟩.

Raftery, A. E. (1995). Bayesian model selection in social research. *Sociological Methodology*, 111–163. https://doi.org/271063.

Rotzer, S., Loenneker, T., Kucian, K., Martin, E., Klaver, P., & von Aster, M. (2009). Dysfunctional neural network of spatial working memory contributes to developmental dyscalculia. *Neuropsychologia, 47*(13), 2859–2865. https://doi.org/10.1016/j.neuropsychologia.2009.06.009

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

Shalev, R. S., & von Aster, M. G. (2008). Identification, classification, and prevalence of development dyscalculia. *Encyclopedia of Language and Literacy Development*, 1–9. https://doi.org/10.5167/uzh-12874

Swets, J. A. (1988). Measuring the accuracy of diagnostic systems. *Science, 240*(4857), 1285–1293. https://doi.org/10.1126/science.3287615

Styck, K. M., & Watkins, M. W. (2016). Structural validity of the WISC-IV for students with learning disabilities. *Journal of Learning Disabilities, 49*(2), 216–224. https://doi.org/10.1177/0022219414539565

Szucs, 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

B.G. Tabachnick L.S. Fidell *Using multivariate statistics* (5th ed.) 2007 Pearsons.

Toffalini, E., Giofrè, D., & Cornoldi, C. (2017). Strengths and weaknesses in the intellectual profile of different subtypes of specific learning disorder: A study on 1,049 diagnosed children. *Clinical Psychological Science, 5*(2), 402–409. https://doi.org/10.1177/2167702616672038

Van Iterson, L., & Kaufman, A. S. (2009). Intra-individual subtest variability on the dutch wechsler intelligence scales for children—revised (WISC–RNL) for children with learning disabilities, psychiatric disorders, and epilepsy. *Psychological Reports, 105*(3), 995–1008. https://doi.org/10.2466/PR0.105.F.995-1008

Vrieze, S. I. (2012). Model selection and psychological theory: A discussion of the differences between the Akaike information criterion (AIC) and the Bayesian information criterion (BIC). *Psychological Methods, 17*(2), 228. https://doi.org/10.1037/a0027127

Watkins, M. W., & Worrell, F. C. (2000). Diagnostic utility of the number of WISC-III subtests deviating from mean performance among students with learning disabilities. *Psychology in the Schools, 37*(4). https://doi.org/10.1002/1520-6807(20002)37:4<311::AID-PITS1>3.0.CO;2-X

Watkins, Marley W., Kush, J. C., & Glutting, J. J. (1997). Discriminant and predictive validity of the WISC-III ACID profile among children with learning disabilities. *Psychology in the Schools, 34*(4), 10.1002/1097-0142(1950)3:1<32::aid-cncr2820030106>3.0.co;2-3.https://doi.org/10.1002/(SICI)1520-6807(199710)34:4<309::AID-PITS2>3.0.CO;2-G.

Wechsler, D. (2003). WISC-IV technical and interpretive manual. San Antonio, TX: The Psychological Association.

World Health Organization (1992). The ICD-10 classification of mental and behavioral disorders: Clinical descriptions and diagnostic guidelines. Geneva, Switzerland.

World Health Organization (2019/2021). International Classification of Diseases, Eleventh Revision (ICD-11). Retrieved from ⟨https://icd.who.int/browse11⟩.

Youden, W. J. (1950). Index for rating diagnostic tests. *Cancer, 3*(1). https://doi.org/10.1002/1097-0142(1950)3:1<32::aid-cncr2820030106>3.0.co;2-3