**Author(s):** Hellstrand, H.; Holopainen, S.; Korhonen, J.; Räsänen, P.; Hakkarainen, A.; Laakso, J-M.; Laine, A.; Aunio, P.

**Title:** Arithmetic fluency and number processing skills in identifying students with mathematical learning disabilities

**Year:** 2024

**Version:** Published version



**Please cite the original version:**

Hellstrand, H., Holopainen, S., Korhonen, J., Räsänen, P., Hakkarainen, A., Laakso, J-M., Laine, A., & Aunio, P. (2024). Arithmetic fluency and number processing skills in identifying students with mathematical learning disabilities. Research in Developmental Disabilities, 151, Article 104795. https://doi.org/10.1016/j.ridd.2024.104795

Research in Developmental Disabilities 151 (2024) 104795

Contents lists available at ScienceDirect

Research in Developmental Disabilities

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

# Arithmetic fluency and number processing skills in identifying students with mathematical learning disabilities☆

H. Hellstrand [a,*,1], S. Holopainen [b,2], J. Korhonen [a,3], P. Räsänen [b,4], A. Hakkarainen [c,5], M.-J. Laakso [b,6], A. Laine [d,7], P. Aunio [d,8]

[a] Faculty of Education and Welfare Studies, Åbo Akademi University, Vaasa, Finland
[b] Turku Research Institute for Learning Analytics, University of Turku, Turku, Finland
[c] University of Jyväskylä, Jyväskylä, Finland
[d] Faculty of Educational Sciences, University of Helsinki, Helsinki, Finland

ARTICLE INFO

*Keywords:*
Arithmetic fluency
Dyscalculia screener
Mathematical learning disabilities
Number processing skills
Numerical skills

ABSTRACT

*Background:* Students with mathematical learning disabilities (MLD) struggle with number processing skills (e.g., enumeration and number comparison) and arithmetic fluency. Traditionally, MLD is identified based on arithmetic fluency. However, number processing skills are suggested to differentiate low achievement (LA) from MLD.

*Aims:* This study investigated the accuracy of number processing skills in identifying students with MLD and LA, based on arithmetic fluency, and whether the classification ability of number processing skills varied as a function of grade level.

*Methods and procedures:* The participants were 18,405 students (girls = 9080) from Grades 3–9 (ages 9–15). Students' basic numerical skills were assessed with an online dyscalculia screener (Functional Numeracy Assessment –Dyscalculia Battery, FUNA-DB), which included number processing and arithmetic fluency as two factors.

*Outcomes and results:* Confirmatory factor analyses supported a two-factor structure of FUNA-DB. The two-factor structure was invariant across language groups, gender, and grade levels. Receiver operating characteristics curve analyses indicated that number processing skills are a fair classifier of MLD and LA status across grade levels. The classification accuracy of number processing skills was better when predicting MLD (cut-off < 5 %) compared to LA (cut-off < 25 %).

*Conclusions and implications:* Results highlight the need to measure both number processing and arithmetic fluency when identifying students with MLD.

☆ A preprint of this manuscript has been submitted to PsyArXiv, July 17th, 2023 (preprint doi: 10.31234/osf.io/jtk8c)

* Correspondence to: Åbo Akademi University, PB 311, 65101 Vaasa, Finland.
*E-mail address:* heidi.hellstrand@abo.fi (H. Hellstrand).
[1] https://orcid.org/0000–0002-5677–1811
[2] https://orcid.org/0000–0002-6777–6247
[3] https://orcid.org/0000–0001-9666–9400
[4] https://orcid.org/0000–0002-6290–3649
[5] https://orcid.org/0000–0001-5199–3493
[6] https://orcid.org/0000–0001-9163–2676
[7] https://orcid.org/0000–0003-3881–8134
[8] https://orcid.org/0000–0002-0901–3874

https://doi.org/10.1016/j.ridd.2024.104795
Received 19 November 2023; Received in revised form 20 June 2024; Accepted 20 June 2024

**What this paper adds?**

The present study contributes to the literature by assessing the precision of number processing skills in identifying mathematical learning disabilities (MLD) and low achievement (LA) among students aged 9–15. Both number processing skills and arithmetic fluency are crucial for later mathematical learning and are recognized indicators for MLD. Notably, traditional identification methods primarily rely on arithmetic fluency in this age group. Our research enhances understanding by evaluating the classification accuracy of number processing skills in identifying MLD (cut-off below the 5th percentile) and LA (cut-off below the 25th percentile) while using arithmetic fluency as the reference standard ("gold standard"). We explore whether the classification accuracy of number processing skills varies across grade levels in a comprehensive cross-sectional analysis spanning Grades 3 to 9. A two-factor structure of the used dyscalculia screener was supported, with arithmetic fluency and number processing skills as distinguishable factors. The two-factor structure was invariant across language groups, gender, and grade levels. The classification accuracy of number processing skills was better when predicting MLD (cut-off < 5 %) compared to LA (cut-off < 25 %) status. Results highlight the importance of measuring both number processing and arithmetic fluency when identifying students with MLD.

## 1. Introduction

Numerical skills are essential in modern society. Students lacking proficient numerical skills are at risk for educational and societal dropout (Aro et al., 2019; Gross et al., 2009), emphasizing the importance of students' mathematical learning for future adult outcomes. Therefore, understanding the foundational numerical skills and identifying students with mathematical learning disabilities (MLD) is vital for preventing and ameliorating difficulties.

Several researchers have suggested that difficulties in basic mathematical skills could be divided into severe (MLD) and milder forms (persistent low achievement, LA) (Geary et al., 2012; Mazzocco & Räsänen, 2013; Zhang et al., 2020). MLD (developmental dyscalculia) is a severe deficit in basic numerical skills that cannot be explained by a general intelligence deficit or an inadequate learning environment (World Health Organization, 2019). It affects a person's mathematics performance in school and everyday life (Kucian & von Aster, 2015). LA refers to a mild but persistent underachievement in mathematics (Murphy et al., 2007; Zhang et al., 2020). The prevalence of LA varies in the literature between 15 % and 35 %, but a commonly used cut-off is at the 25th percentile, and correspondingly, the prevalence of MLD is typically around the 5th percentile (4 % to 7 %; e.g., Devine et al., 2013; Geary, 2011). However, the prevalence of MLD and LA differs based on the criteria and measures used for identifying these students (Devine et al., 2013; Kißler et al., 2021; Murphy et al., 2007).

The process of learning mathematics requires a variety of skills. Number processing skills and arithmetic fluency have been shown to be foundational for later mathematical learning (Li et al., 2018). Arithmetic fluency (e.g., speed and accuracy in one and two-digit addition and subtraction) has commonly been used as a proxy for identifying students with MLD (Mazzocco et al., 2008). Number processing skills (e.g., enumeration and symbolic number comparison) are considered critical components in mathematical learning (Cirino et al., 2015; De Smedt et al., 2013) and predict the development of arithmetic skills (Liu & Wong, 2020; Reeve et al., 2012). Therefore, it is suggested that these skills could function as identifiers of MLD (Räsänen et al., 2021; Schneider et al., 2017). However, the ability of number processing skills to differentiate students with MLD and LA in different age groups is not yet clear. This is essential because, in many countries, students with MLD are legally entitled to remedial services, but those with LA are not. Secondly, if number processing is more strongly associated with MLD than LA, it has information value for treatment. Therefore, analysing whether performance in basic number processing skills could strengthen our ability to differentiate between MLD and LA is essential. Hereby, this study investigated the accuracy of number processing skills in identifying students with MLD and LA based on arithmetic fluency. Even though we acknowledge that mathematical skills are a multi-componential construct, this current study focuses on basic numerical skills instead of applied skills such as word problems or geometry that require language and spatial skills and have not been identified as core indicators for MLD.

### *1.1. Arithmetic fluency and MLD*

Arithmetic fluency is the ability to perform arithmetic operations (i.e., addition, subtraction, multiplication, and division) accurately, effortlessly, and quickly (Vanbinst et al., 2015). It is usually assessed through short, timed arithmetic tasks (Wang et al., 2016). Deficits in understanding counting concepts, reliance on counting-based strategies, difficulty retrieving arithmetic facts and making more procedural errors than their peers are common in students with MLD (Cirino et al., 2015; Geary, Hoard et al., 2011; Psyridou et al., 2024). Arithmetic difficulties have been suggested as an essential aspect of MLD (Mazzocco et al., 2008). They are central to the diagnostic criteria for MLD in *the Diagnostic and Statistical Manual* (DSM-5; American Psychiatric Association, 2013) and *the International Statistical Classification of Diseases and Related Health Problems* (ICD-11; World Health Organization, 2019). The DSM-5 distinguishes between basic academic skills (i.e., basic mathematics: number processing, arithmetic facts, and calculations) and complex or higher-level skills (i.e., problem-solving and reasoning).

An increasing number of studies have made a distinction between MLD and LA, where the former is more connected to difficulties in foundational skills and the latter to various reasons, including poor instruction, motivational issues and general cognitive difficulties (Desoete & Grégoire, 2006; Price & Ansari, 2013). Students who struggle with gaining arithmetic fluency also struggle with foundational numerical skills, including understanding numbers, counting, comparison, and operations (Geary et al., 2012; Huijsmans et al., 2022; Zhang et al., 2020). However, the terminology, diagnostic criteria, or cut-offs concerning the MLD are not yet coherent.

### 1.2. Number processing skills and MLD

Number processing skills are more fundamental than arithmetic fluency (Jordan et al., 2010; Schneider et al., 2017). Number processing skills comprise number comparison (e.g., tasks were to choose the larger or smaller number) and mapping numerals and quantities (e.g., enumeration tasks). Number processing skills partly stem from very early developing or even innate skills. Already infants can discriminate quantities (Bremner et al., 2017; McCrink & Wynn, 2007). Non-symbolic number processing skills expand to symbolic number processing skills with increasing knowledge of number words and symbols. Proficient number processing skills enable students to process and operate with numbers and magnitudes flexibly (Lyons et al., 2014) and to develop arithmetic fluency (Halberda et al., 2008; Wang et al., 2016).

Recent research findings have shown that symbolic number processing (e.g., number comparison) is more associated with later mathematical skills than non-symbolic magnitude processing (Schneider et al., 2017). Number comparison tasks (Ansari, 2008; Dehaene et al., 1990; Schneider et al., 2017) and enumeration tasks (Major et al., 2017) are the most used measures of number processing skills. Symbolic numerical processing skills have been found to correlate strongly with individual differences in arithmetic skills (Bartelet et al., 2014; Vanbinst et al., 2016). They also predict mathematics performance in general (Schneider et al., 2017; Xenidou-Dervou et al., 2013). Number processing skills are suggested to form the core deficit in MLD (De Smedt & Gilmore, 2011; Skagerlund & Träff, 2016), but there are also contradictory findings (Mammarella et al., 2021).

### 1.3. Present study

The present study investigates how well number processing skills identify students with MLD and LA. We used arithmetic fluency as a reference standard ("gold standard") because it is the most typical criterion for identifying students with MLD in this age group (Mazzocco et al., 2008). Based on commonly used cut-offs, we set the 5th percentile as the cut-off for MLD and the 25th percentile for LA (e.g., Geary et al., 2011; Zhang et al., 2020). Furthermore, we investigated whether the classification ability of number processing skills varied as a function of grade level. The grades were combined into three grade groups (i.e., Grades 3–4, Grades 5–6, and Grades 7–9) based on a common division in the Finnish education system, where the upper secondary school comprises Grades 7–9. The following research questions were formed:

1. How well do number processing skills classify MLD (<5 %) based on arithmetic fluency?
2. How well do number processing skills classify LA (<25 %) based on arithmetic fluency?
3. How does the classification ability of number processing differ among students in Grades 3–4, Grades 5–6, and Grades 7–9?

We used a digital online dyscalculia screener to identify students with MLD (Räsänen et al., 2021) to collect a sizeable cross-sectional sample from Grades 3 to 9 (9–15 years of age). First, we verified the a priori two-dimensional structure of the test battery and the measurement invariance across language groups, gender, and grade levels. Hereafter, in the main analyses, we investigated the accuracy of number processing skills in identifying students with MLD and LA, based on arithmetic fluency, and whether the classification ability of number processing skills varied across age.

## 2. Method

### 2.1. Participants

This study is part of a larger [name of the project] project to develop assessment tools for mathematical and reading skills. The sample consisted of 18,405 third- to ninth-grade students (9080 girls and 9325 boys, aged 9 to 15 years) from Finnish (91.4 %) and Swedish-speaking (8.6 %) schools in different parts of Finland.

Students were from seven different grade levels (Grade 3: 5622 students; Grade 4: 4550 students; Grade 5: 1661 students; Grade 6: 1552 students; Grade 7: 3839 students; Grade 8: 818 students; Grade 9: 363 students). The initial sample consisted of 18,409 students, but four students were excluded from the final sample due to incomplete data. No other background characteristics other than language, gender, and grade level were collected. The gender was self-reported by the student, and the teacher reported the grade level.

Students participated anonymously and voluntarily during regular school hours. Research permission and ethical approval were applied separately from each municipality's local educational research committee. A research permit was obtained, and the participating students' parents were informed, following the instructions and policy of each municipality's school authority.

### 2.2. Procedure

The assessments were conducted in April and May at the end of the school year. The teacher administered the assessment during one lesson during regular school hours. Using randomly generated accounts, students logged in on the online educational platform ViLLE (Laakso et al., 2018) on their computers or tablets. The system offers the contents on an internet browser and collects all user interactions and timings (in milliseconds) for further analysis. Each task was introduced with instructions and practice tasks. Students could proceed at their own pace and, if needed, with the assistance of their teacher.

### 2.3. Measurement

The dyscalculia screener FUNA-DB is designed to identify students with MLD in Grades 3–9 (Räsänen et al., 2021). FUNA-DB consists of six tasks: number comparison, digit dot matching, number series, single-digit addition, single-digit subtraction, and multi-digit addition and subtraction. In the number comparison task, two single-digit Arabic numbers were presented on the screen, and the task was to choose the larger of two numbers as fast as possible. In the digit dot matching task, an Arabic number (1−9) and a randomly organized dot pattern were presented on the screen, and the task was to determine as fast as possible if the quantities were equal or different. In the single-digit addition task, calculation tasks were presented one at a time on the screen (e.g., 3 + 8 = _), and the task was to answer as many tasks as possible within two minutes. The identical single-digit subtraction task used the same stimuli but reversed into subtraction tasks (e.g., 11–3 = __). In the multi-digit addition and subtraction task, an equal number of addition and subtraction items with increasing difficulty (e.g., 30 + 40 = _ or 280 – 50 = _) were presented within a 3-minute time limit. In the number series task, four numbers were presented (e.g., 1, 3, 5, 7), and the task was to continue the series with a fifth number based on the rule that the four numbers formed. The time limit was 3 min.

Before calculating the task-specific scores for each student, the data was cleaned by flagging certain cells or cases as missing values when a specific condition was met, effectively removing them from the data. In the tasks where the item reaction time was of interest (number comparison and digit-dot matching), the data was cleaned in two steps. Firstly, very short (< 200 ms) and long reaction times (> 60000 ms) were flagged as missing values, as these values indicated unrealistically short or long response times. After that, the reaction times over three standard deviations above the mean were excluded from the score calculation. In addition, in the number comparison task, the items containing 1 or 9 as one of the numbers were removed, as these numbers are at the end of the number range, and therefore, the solution can be rule-based, not requiring number processing. Finally, because the probability of guessing the correct answer is 0.5 in dual-choice tasks, cases with less than 65 % of correct answers were removed. The probability of correctly guessing more than 65 % of the items is below 5 %. In the tasks with a time limit (number series, single-digit addition, single-digit subtraction, and multi-digit addition and subtraction), cases with less than two correct answers were removed from the data.

After cleaning the data, a task-specific score was calculated for each student. We calculated an efficiency score (the median reaction time of the correct responses divided by the percentage of correct responses) for the number comparison and digit dot matching tasks. The scores of number series, single-digit addition, single-digit subtraction, and multi-digit addition and subtraction were sum scores of the correct answers. The tasks' split-half reliabilities (Spearman-Brown and Guttman) varied between 0.75 and 0.98. Table 1 summarises the information about the FUNA-DB.

The correlation between the tasks varied from .47 to .86. In number comparison and digit dot matching, a high score indicated a low performance (i.e., efficiency score), whereas a high score in number series, single-digit addition, single-digit subtraction, and multi-digit addition and subtraction indicated a high performance (i.e., sum score). Therefore, the sum scores and efficiency scores have negative correlations. The correlations are presented in Table 2.

### 2.4. Statistical analyses

The analyses were conducted with the R (version 4.0.5) and Mplus (version 8.6) statistical software. First, confirmatory factor analyses (CFA) were performed to ratify the factor structure of FUNA-DB. In previous research (Räsänen et al., 2021), a two-factor structure of FUNA-DB was supported, with the factors of number processing skills (number comparison, digit dot matching) and arithmetic fluency (number series, single-digit addition, single-digit subtraction, multi-digit addition, and subtraction). To evaluate this two-factor model for the current sample, it was compared to a one-factor model where all tasks load on one numerical skills factor.

After evaluating the two-factor model, we tested the measurement invariance with multigroup CFA for the different student groups to determine whether the factor structure was invariant between different language groups (Finnish and Swedish), gender (girls and boys), and Grades (Grades 3–4, 5–6, and 7–9). In multigroup CFA, a series of nested models are fitted to the data where the endpoints are the least restrictive model with no invariance constraints and the most restrictive model where all parameters are forced to equality across groups (Bollen, 1989). In all the analyses, the full information maximum likelihood (FIML) was the preferred estimation method

**Table 1**
Information Regarding the FUNA-DB Tasks.

| FUNA-DB task | Description | Type of score | Number of items/ Time limit | Split-half reliabilities | |
|---|---|---|---|---|---|
| | | | | Spearman-Brown | Guttman |
| Number Comparison | Two single-digit numbers, which number is larger? | Efficiency score | 52 items | 0.769 | 0.787 |
| Digit Dot Matching | Are the quantities equal or different for a single-digit number and a group of dots? | Efficiency score | 42 items | 0.751 | 0.765 |
| Number Series | The fifth number in a number series is missing, what is the missing number? | Sum score | 3 min | 0.927 | 0.916 |
| Single-digit Addition | Addition with two single-digit numbers | Sum score | 2 min | 0.966 | 0.984 |
| Single-digit Subtraction | Subtraction with two single-digit numbers | Sum score | 2 min | 0.961 | 0.983 |
| Multi-digit Addition and Subtraction | Addition and subtraction with two multi-digit numbers | Sum score | 3 min | 0.941 | 0.953 |

**Table 2**
Descriptive Statistics and Pearson Correlations for the FUNA-DB Tasks.

| Task | *n*[a] | *M* | *SD* | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|---|---|
| 1. Number Comparison | 18,306 | 1001.10 | 293.52 | — | | | | | |
| 2. Digit Dot Matching | 18,137 | 2184.04 | 690.24 | .68*** | — | | | | |
| 3. Number Series | 18,082 | 14.35 | 4.87 | -.47*** | -.48*** | — | | | |
| 4. Single-digit Addition | 18,146 | 38.18 | 13.05 | -.59*** | -.59*** | .70*** | — | | |
| 5. Single-digit Subtraction | 18,044 | 34.83 | 12.14 | -.55*** | -.56*** | .73*** | .86*** | — | |
| 6. Multi-digit Addition and Subtraction | 17,625 | 21.60 | 8.95 | -.52*** | -.55*** | .74*** | .79*** | .84*** | — |

*** $p < .001$.
[a] After casewise deletion of missing values.

as it uses all available data in the analyses. FIML estimates seem unbiased and more efficient than other missing data methods under structural equation modelling when the data is missing at random or completely at random (Arbuckle, 1996). The Comparative Fit Index (CFI), the Tucker-Lewis Index (TLI), and the Root Mean Square Error of Approximation (RMSEA) were used as the indicators for model fit (West et al., 2012). Since these values depend on the Chi-square ($\chi^2$) test score's value, we also reported the Chi-square test results. In the CFI, values greater than 0.90 and in the TLI, values greater than 0.95 are considered to indicate an acceptable and an excellent fit to the data, whereas for RMSEA, values of less than 0.05 and 0.08 indicate a close and a reasonable fit to the data, respectively (Marsh et al., 2004). To compare the nested models in the tests of measurement invariance, we investigated the changes in CFI and RMSEA. Support for the more parsimonious model requires a change in CFI (ΔCFI) of less than 0.01 or a change in RMSEA (ΔRMSEA) of less than 0.015 (Chen, 2007).

After confirming the factor structure and measurement invariance, the classification accuracy of number processing skills to differentiate students with MLD and LA was investigated to answer the research questions. Grade-level-based standardized sum variables were formed for number processing skills and arithmetic fluency in these analyses. The classification accuracy of number processing skills was tested using the R software package 'pROC' (Robin et al., 2011). Firstly, a receiver operating characteristic (ROC) analysis was conducted using the total sample. The purpose of this was to see whether the student's performances in number processing skills tasks could be used to classify the students as MLD (True Positive, TP) and non-MLD (True Negative, TN) based on their performances in arithmetic fluency (cut-off at the 5th percentile). Second, the same procedure was conducted to investigate whether the students' performances in number processing skills tasks could be used to classify the students as LA (TP) and non-LA (TN) (cut-off at 25th percentile). Third, ROC analyses were conducted separately for Grades 3–4, 5–6, and 7–9 to see whether the classification ability of the number processing skills varied according to age.

The ROC analyses consisted of drawing an ROC curve that shows the performance of a classification model by showing the relationship between sensitivity and specificity parameters for all the different cut-off values that specify which values of the classifier separate the students as MLD/LA and non-MLD/LA. In our case, sensitivity means the proportion of students that the classification model classified as MLD/LA of all students that performed below the defined cut-off (5 % vs. 25 %) in the arithmetic fluency tasks. Sensitivity is calculated with the formula TP/(TP+FN), where TP (True Positive) is the number of students classified correctly as MLD/LA based on their performances in number processing skills. FN (False Negative) is the number of students classified as non-MLD/LA based on their performance in number processing skills but have MLD/LA. It means that TP + FN is the total number of students considered to have MLD/LA. The specificity means the proportion of students that the classification model classified as non-MLD/LA out of those who are non-MLD/LA based on their performances in the arithmetic fluency tasks. Specificity is calculated with the formula TN/(TN+FP) where TN (True Negative) is the number of students that have been classified to perform above the defined cut-off (5 % vs. 25 %) and have done so according to their performances in the arithmetic fluency tasks, in other words, these students are correctly classified as non-MLD/LA. FP (False Positive) is the number of students classified as MLD/LA but performing above the defined cut-off (5 % vs. 25 %) in the arithmetic fluency tasks. TN + FP is the total number of students who are non-MLD/LA.

Furthermore, we calculated the area under the curve (AUC) from all the ROC curves. The range of the AUC is [0, 1]. AUC summarizes the performance of a classification model as it can be understood as the average value of sensitivity for all possible specificity values (Bozikov & Lijana, 2010). A classification model where AUC = 1 classifies 100 % of the students correctly. Instead, a classification model where AUC = 0.5 is equivalent to a coin flip situation and is useless as a classification model. A value of AUC between 0.7 and 0.8 indicates that the model is a fair classifier, between 0.8 and 0.9 a good classifier, and above 0.9 an excellent classifier (Bozikov & Lijana, 2010). In addition to the AUC of the ROC curves, we were also interested in seeing how the number processing skills ability in classifying the MLD/LA students changed when using specific cut-off values for number processing skills, starting from the 25th percentile and then decreasing the rank of the percentile by five until the lowest 5th percentile was reached.

The values of the sensitivity, specificity, Positive Predictive Value (PPV), Negative Predictive Value (NPV), and accuracy parameters for different threshold percentiles for the different grade levels were investigated. The value of the PPV parameter reflects the probability that a student classified as MLD/LA truly performs lower than non-MLD/LA students (PPV = TP / (TP + FP)). The value of the NPV parameter reflects the probability that a student classified as non-MLD/LA truly performs better than a student classified as MLD/LA (NPV = TN / (FN + TN)). The value of the accuracy parameter tells the proportion of all of the correctly classified students of all of the students in a grade level (accuracy = (TP + TN) / (TP + TN + FP + FN)).

## 3. Results

### 3.1. Preliminary Analysis

First, we analyzed the factor structure of the FUNA-DB for the current sample. The one-factor model where all of the tasks load on one numerical skills factor did not fit the data very well, $\chi^2(9) = 5920.796$, $p < .001$; CFI = .929; TLI = .882; RMSEA = .189. The two-factor model with a number processing skills factor and an arithmetic fluency factor fitted the data well, $\chi^2(8) = 1242.041$, $p < .001$; CFI = .985; TLI = .972; RMSEA = .092. Therefore, the two-factor structure was supported as the optimal factor structure for FUNA-DB. Factor 1, number processing skills, was comprised of two tasks (number comparison and digit dot matching) with factor loadings.82 and.83. Factor 2, arithmetic fluency, was comprised of four tasks (number series, single-digit addition, single-digit subtraction, and multi-digit addition and subtraction) with factor loadings from.79 to.94. Fig. 1 presents the two-factor solution and the factor loadings.

After confirming the optimal factor structure, we tested measurement invariance across language groups, gender, and grade by conducting multigroup CFAs on FUNA-DB. Three models (i.e., configural, metric, and scalar invariance model) with different levels of restrictions were compared to each other in the multigroup CFAs. The configural model, which assumes that the factor structure is the same across groups, was set as the baseline model. It was then compared to the metric invariance model (equal factor loadings across groups) and the scalar invariance model (equal factor loadings and intercepts across groups). According to the changes in CFI and RMSEA, scalar invariance applied across language groups ($\Delta$CFI < 0.01; $\Delta$RMSEA < 0.015) and gender ($\Delta$CFI <= 0.011; $\Delta$RMSEA < 0.015). Support for scalar invariance across grade levels was not as obvious. In this case, the changes in CFI and RMSEA were $\Delta$CFI = 0.02 (> 0.01) and $\Delta$RMSEA = 0.019 (> 0.015) at the highest. However, the scalar model did fit the data reasonably well ($\chi^2(104) = 2464.597$, p < .001; CFI = .963; TLI = .963; RMSEA = .093). Therefore, the two-factor structure of FUNA-DB showed invariance across the whole sample.

### 3.2. Main analysis

#### 3.2.1. Classification accuracy of number processing skills in identifying MLD

We conducted ROC analyses to investigate whether the students' performances in the number processing skills tasks could be used to classify them as MLD and non-MLD based on their performances in the arithmetic fluency tasks. To draw the ROC curves, a binary variable was formed out of the standardized arithmetic fluency sum variable. The binary variable had a value of 1 when the value of the arithmetic fluency sum variable was less than or equal to the 5th percentile (MLD) and a value of 0 otherwise (non-MLD).

The value of the AUC parameter was calculated for the ROC curve. When the total sample was used for the ROC analysis, the ROC curve's AUC was 0.799, indicating a reasonably good classifier. The ROC curve for the total sample (all grades) is presented in Fig. 2.

The accuracy of number processing skills in classifying MLD and non-MLD was also investigated based on the values of the sensitivity, specificity, PPV, NPV, and accuracy parameters for different threshold percentiles. From the 25th percentile downwards,

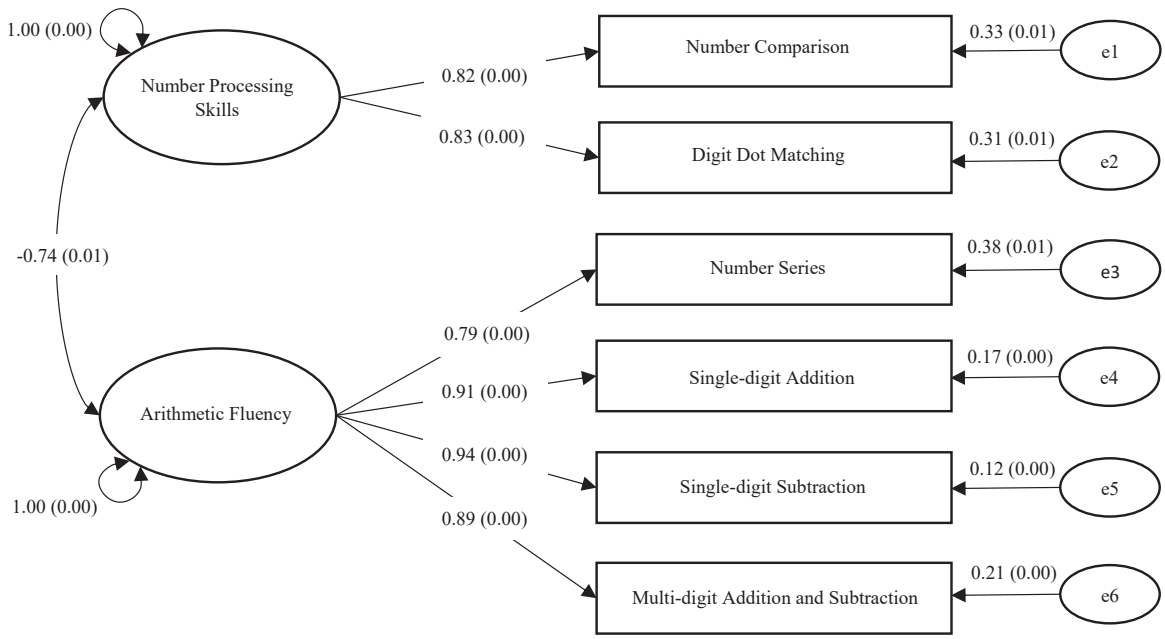


**Fig. 1.** The Two-Factor Solution with Item Loadings for FUNA-DB. Note. $N = 18{,}405$. e = error. The estimation method in CFA was the full information maximum likelihood method (FIML). The parameter estimates are standardized, and the variances of the latent factors are set to 1 by default. The standard errors are in brackets. All estimated parameters were statistically significantly different from 0 at the confidence level of 0.001.

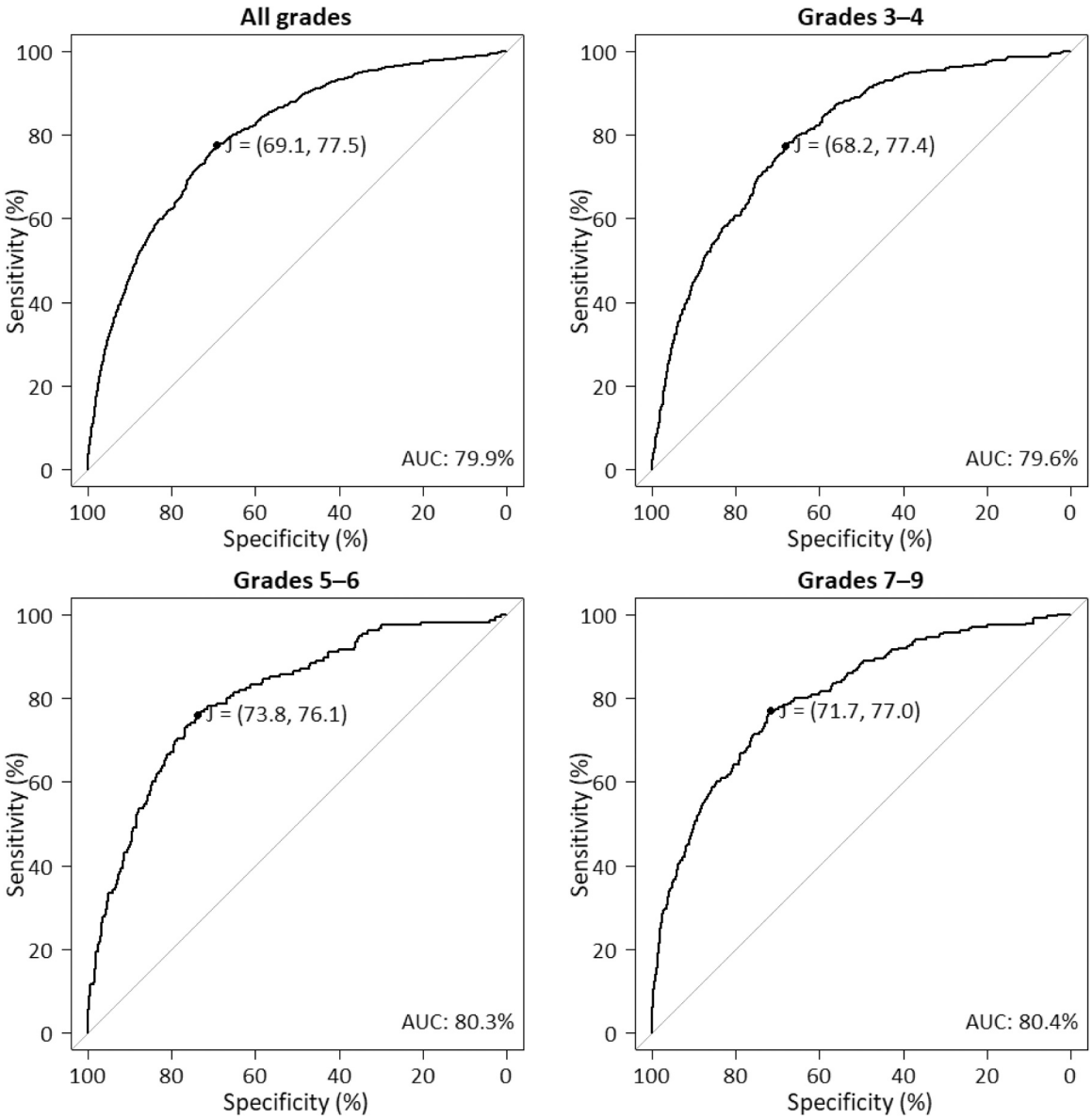


**Fig. 2.** Number Processing Skills in Identifying Mathematical Learning Disabilities based on Performance in Arithmetic Fluency: ROC curves for different grade levels and all grades combined. Note. After casewise deletion of missing values, the number of students in Grades 3–4: 9452; Grades 5–6: 3067; Grades 7–9: 4669. AUC = area under the curve. $J$ = the optimal cut-off value coordinates based on the maximal Youden's index: $J$ = max {sensitivity + specificity − 100}.

the accuracy increased when the cut-off for number processing skills was lowered. Using a cut-off at the 25th percentile, 3.4 % TPs, 73.3 % TNs, 1.7 % FNs, and 21.6 % FPs were identified in the total sample. When lowering the cut-off to the 5th percentile, the number of students identified as TP and FP decreased, and the number of students identified as TN or FN increased. When the cut-off was set at the 25th percentile, 13.4 % of the students identified as MLD truly performed low. When the cut-off was set as 5th percentile, 27.0 % of the students identified as MLD truly performed low (PPV). The values of TP, TN, FN, FP, sensitivity, specificity, PPV, NPV, and accuracy are presented in Table 3 (Appendix Table A.1 for all cut-off percentiles).

### *3.2.2. Classification accuracy of number processing skills in identifying LA*

To investigate the classification accuracy of number processing skills in identifying LA, a similar procedure was performed when identifying MLD. The binary variable had a value of 1 when the arithmetic fluency sum variable was less than or equal to the 25th percentile in the grade level in question (LA) and a value of 0 otherwise (non-LA).

The value of the AUC parameter was calculated for each ROC curve. When the total sample was used for the ROC analysis, the ROC

**Table 3**
Sensitivity and Specificity Parameters when Identifying Mathematical Learning Disabilities Based on Performance in Arithmetic Fluency.

| Percentile | Cut-off[a] | TP | TN | FN | FP | Sensitivity (%) | Specificity (%) | PPV (%) | NPV (%) | Accuracy (%) | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | | Value | Lower | Upper |
| All grades | | | | | | | | | | | | |
| 5 | -1.876 | 232 | 15,697 | 632 | 627 | 26.9 | 96.2 | 27.0 | 96.1 | 92.7 | 92.3 | 93.0 |
| 25 | -0.524 | 576 | 12,603 | 288 | 3721 | 66.7 | 77.2 | 13.4 | 97.8 | 76.7 | 76.0 | 77.3 |
| Grades 3–4 | | | | | | | | | | | | |
| 5 | -1.841 | 120 | 8625 | 354 | 353 | 25.3 | 96.1 | 25.4 | 96.1 | 92.5 | 92.1 | 93.0 |
| 25 | -0.546 | 302 | 6917 | 172 | 2061 | 63.7 | 77.0 | 12.8 | 97.6 | 76.4 | 75.6 | 77.3 |
| Grades 5–6 | | | | | | | | | | | | |
| 5 | -1.929 | 43 | 2801 | 112 | 111 | 27.7 | 96.2 | 27.9 | 96.2 | 92.7 | 91.9 | 93.5 |
| 25 | -0.537 | 109 | 2254 | 46 | 658 | 70.3 | 77.4 | 14.2 | 98.0 | 77.0 | 75.6 | 78.4 |
| Grades 7–9 | | | | | | | | | | | | |
| 5 | -1.924 | 71 | 4272 | 164 | 162 | 30.2 | 96.3 | 30.5 | 96.3 | 93.0 | 92.4 | 93.6 |
| 25 | -0.470 | 160 | 3426 | 75 | 1008 | 68.1 | 77.3 | 13.7 | 97.9 | 76.8 | 75.6 | 78.0 |

*Note*. After casewise deletion of missing values, the number of students in Grades 3–4: 9452; Grades 5–6: 3067; Grades 7–9: 4669. [a] Value of the standardized number processing skills sum variable. TP = true positive; TN = true negative; FN = false negative; FP = false positive; PPV = positive predictive value; NPV = negative predictive value.

curve's AUC was 0.763, indicating a reasonably good classifier. The ROC curve is presented in Fig. 3.

From the 25th percentile downwards, the accuracy increased when the cut-off of number processing skills was lowered. Using a cut-off at the 25th percentile, 12.8 % TPs, 62.8 % TNs, 12.2 % FNs, and 12.2 % FPs were identified in the total sample. When lowering the cut-off to the 5th percentile, the number of students identified as TP and FP decreased, and the number of students identified as TN or FN increased. When the cut-off was set at the 25th percentile, 51.4 % of the students identified as LA truly performed low. When the cut-off was set at the 5th percentile, 73.2 % of the students identified as LA truly performed low (PPV). The values of TP, TN, FN, FP, sensitivity, specificity, PPV, NPV, and accuracy are presented in Table 4 (Appendix Table A.2 for all cut-off percentiles).

### *3.2.3. Classification accuracy of number processing skills in different grade levels*

To investigate whether the classification ability of number processing skills differed between the grade levels, ROC analysis for the different grade levels was performed. When identifying MLD, the AUC parameter values were 0.796 (Grades 3–4), 0.803 (Grades 5–6), and 0.804 (Grades 7–9) (see Fig. 2). When identifying LA, the AUC parameter values were 0.765 (Grades 3–4), 0.761 (Grades 5–6), and 0.761 (Grades 7–9) (Fig. 3). The values of the sensitivity, specificity, PPV, NPV, and accuracy parameters for different threshold percentiles showed a similar trend across all age groups, both for MLD (Table 3) and LA (Table 4).

The classification of number processing skills and arithmetic fluency is illustrated in Fig. 4. The relationship between the student's performance in number processing skills and arithmetic fluency shows a greater spread in the lower end of the continuum. If only arithmetic fluency or number processing skills were used for identifying MLD or LA, some students performed above the cut-off in the other factor. When the cut-off for number processing skills was lowered, a more precise group of students could be identified, indicating that the accuracy for identifying MLD would benefit from including both number processing skills and arithmetic fluency. The results show the overlap of students identified as MLD or LA based on their performance in number processing skills and arithmetic fluency.

## 4. Discussion

The current study investigated the accuracy of number processing skills in identifying students with MLD and LA based on arithmetic fluency and whether the classification ability of number processing skills varied across age groups. We conducted a cross-sectional study with 18,405 students from third to ninth grade. A two-factor structure of the dyscalculia screener was supported, with arithmetic fluency and number processing skills as intercorrelated but distinguishable factors. The two-factor structure was invariant across language groups, gender, and grade levels. We used arithmetic fluency as the reference standard ("gold standard") to identify students with MLD and LA against which number processing skills were compared. Our results indicated that number processing skills could differentiate students with MLD and LA reasonably well. This classification ability of number processing tasks remained the same when the data were grouped by grade level. Furthermore, number processing skills were better in identifying MLD (cut-off $< 5$ %) compared to identifying LA (cut-off $> 25$ %). Our results support using both arithmetic fluency and number processing skills as criteria when identifying students with MLD.

This study confirmed the two-factor model of the dyscalculia screener with a four-times sample than used in our previous study (Räsänen et al., 2021). Of importance is the invariance of the results across language groups, gender, and grade levels. This is consistent with studies demonstrating that both arithmetic fluency and number processing skills are two separate components of basic numerical skills essential for mathematical learning (Li et al., 2018).

The measurement invariance supported the idea that tasks used in the dyscalculia screener work across a broad age span. The cognitive mechanisms for mathematical development seem to be similar for students at different skill levels and ages (Geary et al., 2012; Huijsmans et al., 2022). Arithmetic fluency is a traditional measure, whereas recent literature highlights number processing

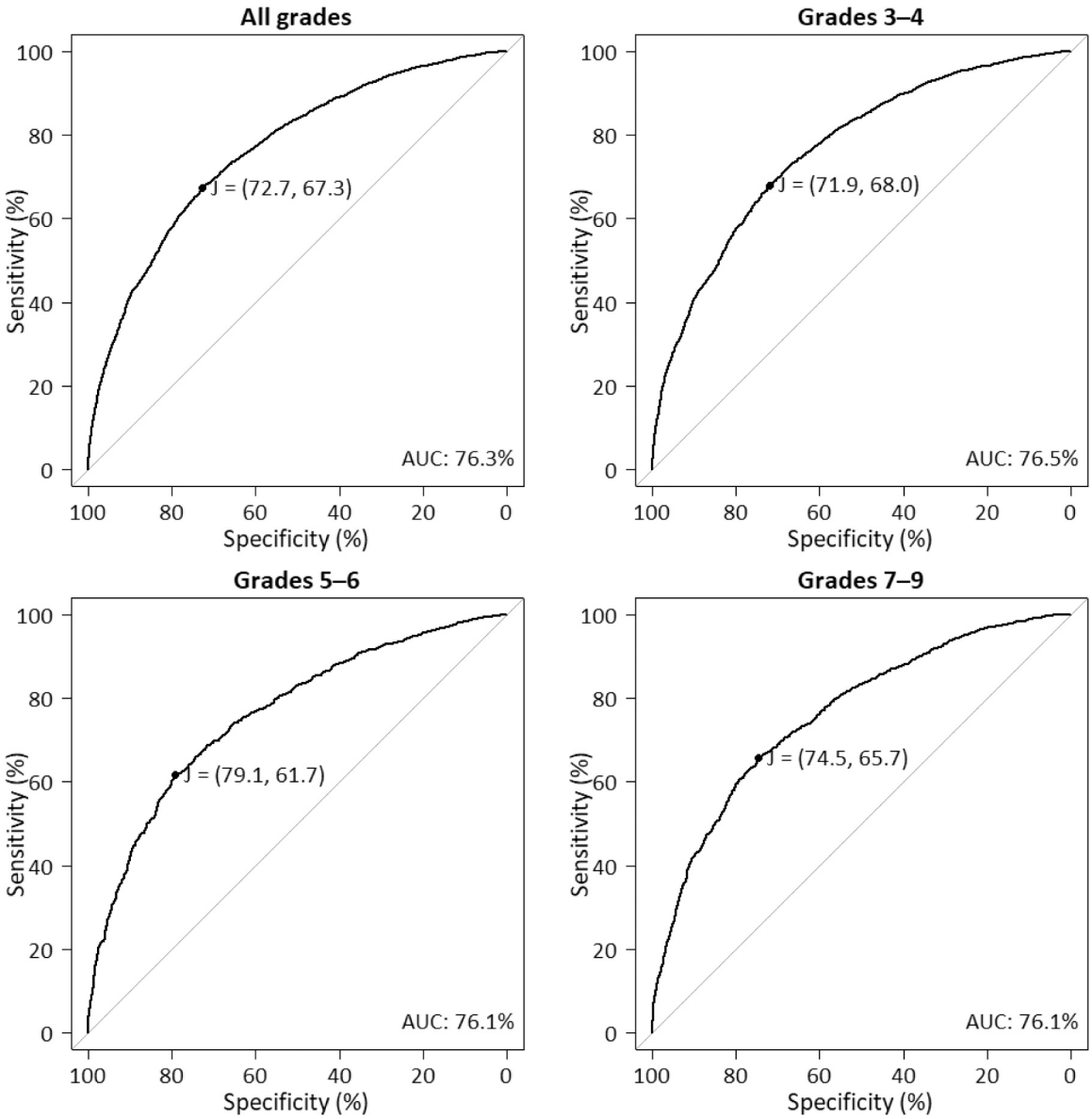


**Fig. 3.** Number Processing Skills in Identifying Low Achievement based on Performance in Arithmetic Fluency: ROC curves for different grade levels and all grades combined.

skills as a key indicator for MLD (Räsänen et al., 2021; Schneider et al., 2017).

The tasks of the screener also work in two different languages. Even though the language groups in this study come from the same educational culture, the languages (Finnish and Swedish) come from different language families. In addition, recent international assessments show that these language groups in Finland differ in their mathematical skills (Leino et al., 2019; (Räsänen et al., 2021)). However, these results indicate that FUNA-DB can identify MLD among students from diverse language backgrounds. More studies are needed from different educational cultures, especially countries with low-performance levels in international mathematical evaluations. MLD and LA differentiation have additional confounding factors in these countries, and we lack information about the relationship between and the diagnostic value of arithmetic fluency and number processing skills.

We used arithmetic fluency as the reference standard or the "gold standard", against which number processing skills were compared. The results indicated that measuring number processing skills could differentiate students with MLD across grade levels. The classification accuracy was acceptable in all grade levels and remained the same across the age groups. When setting the cut-off at the 25th percentile, the positive predictive value was lower than when tightening the cut-off to the 5th percentile, indicating that when predicting MLD (lowest 5 %) compared to LA (lowest 25 %), the number processing deficit was more pronounced in the MLD group compared to the LA group. This finding aligns with studies suggesting number processing is a key indicator for MLD (De Smedt &

**Table 4**
Sensitivity and Specificity Parameters when Identifying Low Achievement Based on Performance in Arithmetic Fluency.

| Percentile | Cut-off[a] | TP | TN | FN | FP | Sensitivity (%) | Specificity (%) | PPV (%) | NPV (%) | Accuracy (%) | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | | Value | Lower | Upper |
| All grades | | | | | | | | | | | | |
| 5 | -1.876 | 629 | 12,658 | 3671 | 230 | 14.6 | 98.2 | 73.2 | 77.5 | 77.3 | 77.0 | 77.6 |
| 25 | -0.524 | 2207 | 10,798 | 2093 | 2090 | 51.3 | 83.8 | 51.4 | 83.8 | 75.7 | 75.1 | 76.3 |
| Grades 3–4 | | | | | | | | | | | | |
| 5 | -1.841 | 350 | 6965 | 2014 | 123 | 14.8 | 98.3 | 74.0 | 77.6 | 77.4 | 77.0 | 77.8 |
| 25 | -0.546 | 1196 | 5921 | 1168 | 1167 | 50.6 | 83.5 | 50.6 | 83.5 | 75.3 | 74.5 | 76.1 |
| Grades 5–6 | | | | | | | | | | | | |
| 5 | -1.929 | 116 | 2262 | 651 | 38 | 15.1 | 98.3 | 75.3 | 77.7 | 77.5 | 76.8 | 78.3 |
| 25 | -0.537 | 399 | 1932 | 368 | 368 | 52.0 | 84.0 | 52.0 | 84.0 | 76.0 | 74.6 | 77.4 |
| Grades 7–9 | | | | | | | | | | | | |
| 5 | -1.924 | 164 | 3431 | 1005 | 69 | 14.0 | 98.0 | 70.4 | 77.3 | 77.0 | 76.4 | 77.6 |
| 25 | -0.470 | 599 | 2931 | 570 | 569 | 51.2 | 83.7 | 51.3 | 83.7 | 75.6 | 74.4 | 76.7 |

*Note*. After casewise deletion of missing values, the number of students in Grades 3–4: 9452; Grades 5–6: 3067; Grades 7–9: 4669. [a] Value of the standardized number processing skills sum variable. TP = true positive; TN = true negative; FN = false negative; FP = false positive; PPV = positive predictive value; NPV = negative predictive value.

Gilmore, 2011; Skagerlund & Träff, 2016). The students who were identified as LA based on arithmetic fluency were only partly the same as those who were identified as LA based on number processing skills. Previous studies have shown that LA is a heterogeneous group: motivational (e.g., interest), emotional (e.g., math anxiety), environmental (e.g., family SES), and other cognitive factors (e.g., working memory) (Friso-Van den Bos et al., 2013; Murayama et al., 2013; Namkung et al., 2019) are related to low performance in arithmetic.

### *4.1. Limitations and future research*

This cross-sectional study used a nationally representative and large-scale dataset with over 18,000 students to evaluate our research question. Despite the large and representative sample, this study is not without limitations. First, the number of ninth-grade students was relatively small compared to the other grade levels, and in fact, all the ninth-graders were Finnish-speaking. It would be eligible to complement the sample with Swedish-speaking students to secure cross-cultural validity for the ninth graders. However, as the results showed that the FUNA-DB works in a large age span, it would be interesting to investigate how the dyscalculia screener identifies students in higher and lower grades. Especially for the beginning learners of school mathematics, accuracy may be a more important factor than fluency when using arithmetic skills as a measure. Second, as this study was based on cross-sectional data, a longitudinal design is needed to confirm the stability of factors across development. Furthermore, a longitudinal design would enable an investigation of the persistence of MLD and LA over time.

A longitudinal design would also secure the predictive validity of the FUNA-DB, which is essential when developing screening tests. In terms of reliability and validity evidence, this study supports evidence for structural validity, cross-cultural validity (i.e., language groups), known-group validity (i.e., gender and grade levels), and concurrent validity of the FUNA-DB. In addition, even though the factor structure supported two distinct factors (i.e., arithmetic fluency and number processing skills), the number processing skills comprised two types of tasks, one comparing symbolic numbers (i.e., mapping number-to-number) and one enumeration task (i.e., mapping number symbol to quantity). As recent literature highlights fluency in operating with symbolic numbers as a key indicator in mathematical learning (Brankaer et al., 2014; Lyons et al., 2014; Vanbist et al., 2015), it is of interest in future studies to investigate whether one of these skills is more salient than the other and could more accurately identify students with MLD.

### *4.2. Implications for practice*

There is evidence that MLD has long-term consequences on well-being and employment (Aro et al., 2019). In many countries, diagnostic procedures are essential for receiving educational support. Recognising reliably those children with severe forms of learning disorders is essential for school systems to be able to target the limited resources of educational support using fair and research-based procedures, methods, and tools. A very limited number of studies have evaluated the validity and reliability evidence of the assessment tools used at schools and research. This study adds important information about the requirements for valid assessment tools for MLD. Test batteries that include both tasks measuring arithmetic and basic number processing skills are recommended as part of the diagnostic procedures for MLD.

## Funding

This research was partly supported by a grant from the Swedish Cultural Foundation in Finland (140884) and from the Academy of Finland EDUCA Flagship Programme (358947) to the research project.

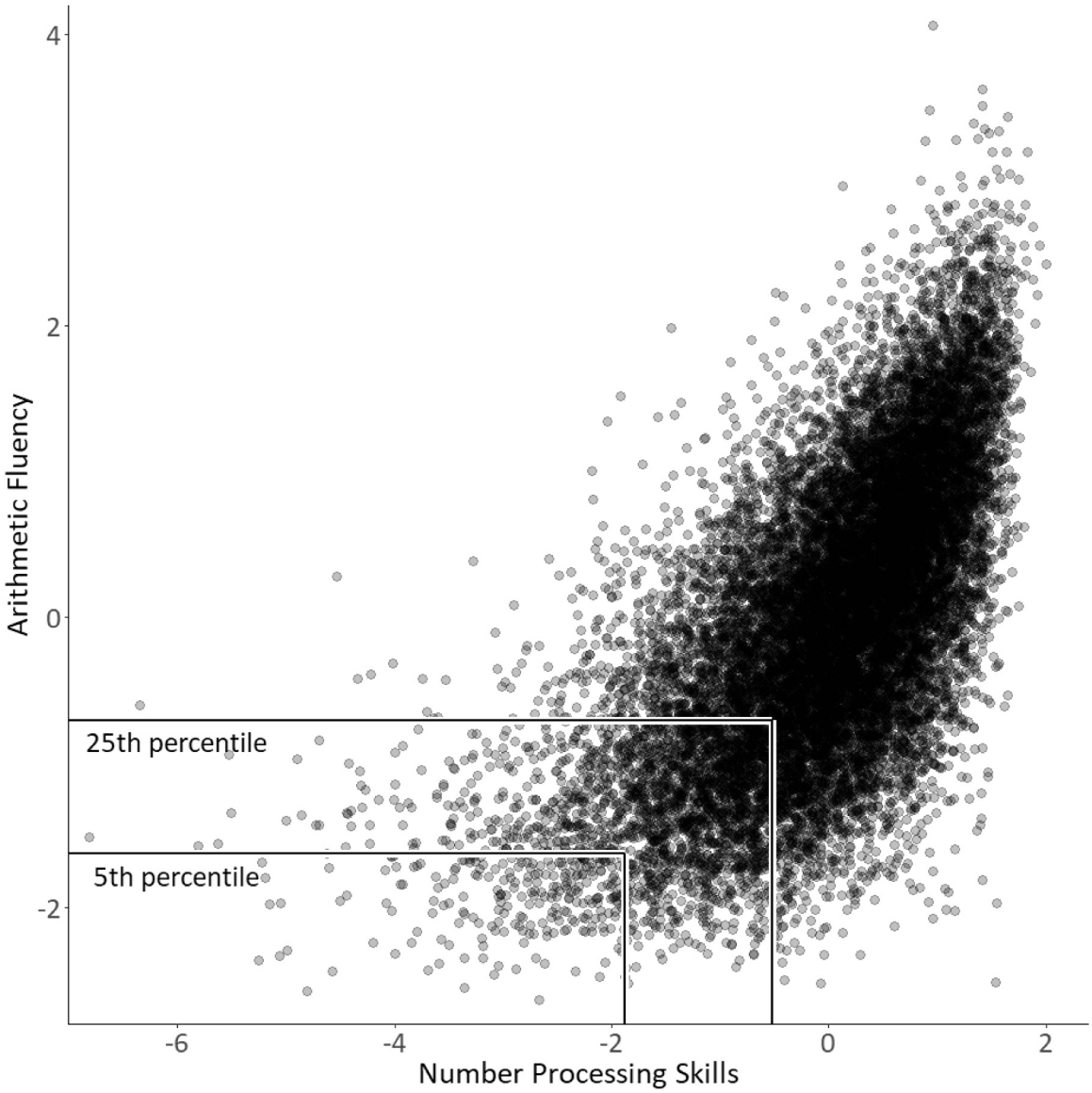


**Fig. 4.** The Classification of Number Processing Skills and Arithmetic Fluency on Different Cut-offs (5 and 25 percentile) in the Total Sample (n = 17,188).

## Declaration of Competing Interest

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

## Data availability

The authors do not have permission to share data.

## Appendices

**Table A.1**
Sensitivity and Specificity Parameters when Identifying Mathematical Learning Disabilities Based on Performance in Arithmetic Fluency (cut-off 5th percentile).

| Percentile | Cut-off[a] | TP | TN | FN | FP | Sensitivity (%) | Specificity (%) | PPV (%) | NPV (%) | Accuracy (%) | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | | Value | Lower | Upper |
| All grades | | | | | | | | | | | | |
| 5 | -1.876 | 232 | 15,697 | 632 | 627 | 26.9 | 96.2 | 27.0 | 96.1 | 92.7 | 92.3 | 93.0 |
| 10 | -1.323 | 355 | 14,961 | 509 | 1363 | 41.1 | 91.7 | 20.7 | 96.7 | 89.1 | 88.7 | 89.5 |
| 15 | -0.982 | 457 | 14,203 | 407 | 2121 | 52.9 | 87.0 | 17.7 | 97.2 | 85.3 | 84.8 | 85.8 |
| 20 | -0.731 | 519 | 13,406 | 345 | 2918 | 60.1 | 82.1 | 15.1 | 97.5 | 81.0 | 80.4 | 81.6 |
| 25 | -0.524 | 576 | 12,603 | 288 | 3721 | 66.7 | 77.2 | 13.4 | 97.8 | 76.7 | 76.0 | 77.3 |
| 33,3 | -0.230 | 670 | 11,279 | 194 | 5045 | 77.5 | 69.1 | 11.7 | 98.3 | 69.5 | 67.7 | 75.2 |
| Grades 3–4 | | | | | | | | | | | | |
| 5 | -1.841 | 120 | 8625 | 354 | 353 | 25.3 | 96.1 | 25.4 | 96.1 | 92.5 | 92.1 | 93.0 |
| 10 | -1.319 | 189 | 8222 | 285 | 756 | 39.9 | 91.6 | 20.0 | 96.6 | 89.0 | 88.4 | 89.6 |
| 15 | -0.985 | 244 | 7804 | 230 | 1174 | 51.5 | 86.9 | 17.2 | 97.1 | 85.1 | 84.4 | 85.8 |
| 20 | -0.739 | 276 | 7363 | 198 | 1615 | 58.2 | 82.0 | 14.6 | 97.4 | 80.8 | 80.0 | 81.6 |
| 25 | -0.546 | 302 | 6917 | 172 | 2061 | 63.7 | 77.0 | 12.8 | 97.6 | 76.4 | 75.6 | 77.3 |
| 34,1 | -0.230 | 367 | 6127 | 107 | 2851 | 77.4 | 68.2 | 11.4 | 98.3 | 68.7 | 58.4 | 74.9 |
| Grades 5–6 | | | | | | | | | | | | |
| 5 | -1.929 | 43 | 2801 | 112 | 111 | 27.7 | 96.2 | 27.9 | 96.2 | 92.7 | 91.9 | 93.5 |
| 10 | -1.309 | 62 | 2668 | 93 | 244 | 40.0 | 91.6 | 20.3 | 96.6 | 89.0 | 88.0 | 90.0 |
| 15 | -1.009 | 83 | 2535 | 72 | 377 | 53.5 | 87.1 | 18.0 | 97.2 | 85.4 | 84.1 | 86.6 |
| 20 | -0.738 | 98 | 2396 | 57 | 516 | 63.2 | 82.3 | 16.0 | 97.7 | 81.3 | 80.0 | 82.7 |
| 25 | -0.537 | 109 | 2254 | 46 | 658 | 70.3 | 77.4 | 14.2 | 98.0 | 77.0 | 75.6 | 78.4 |
| 28,8 | -0.378 | 118 | 2149 | 37 | 763 | 76.1 | 73.8 | 13.4 | 98.3 | 73.9 | 70.7 | 81.0 |
| Grades 7–9 | | | | | | | | | | | | |
| 5 | -1.924 | 71 | 4272 | 164 | 162 | 30.2 | 96.3 | 30.5 | 96.3 | 93.0 | 92.4 | 93.6 |
| 10 | -1.339 | 105 | 4072 | 130 | 362 | 44.7 | 91.8 | 22.5 | 96.9 | 89.5 | 88.6 | 90.3 |
| 15 | -0.958 | 131 | 3864 | 104 | 570 | 55.7 | 87.1 | 18.7 | 97.4 | 85.6 | 84.6 | 86.6 |
| 20 | -0.710 | 144 | 3644 | 91 | 790 | 61.3 | 82.2 | 15.4 | 97.6 | 81.1 | 80.0 | 82.2 |
| 25 | -0.470 | 160 | 3426 | 75 | 1008 | 68.1 | 77.3 | 13.7 | 97.9 | 76.8 | 75.6 | 78.0 |
| 30,8 | -0.256 | 181 | 3179 | 54 | 1255 | 77.0 | 71.7 | 12.6 | 98.3 | 72.0 | 70.2 | 83.9 |

*Note*. After casewise deletion of missing values, the number of students in Grades 3–4: 9452; Grades 5–6: 3067; Grades 7–9: 4669. [a] Value of the standardized number processing skills sum variable. TP = true positive; TN = true negative; FN = false negative; FP = false positive; PPV = positive predictive value; NPV = negative predictive value.

**Table A.2**
Sensitivity and Specificity Parameters when Identifying Low Achievement Based on Performance in Arithmetic Fluency (cut-off 25th percentile).

| Percentile | Cut-off[a] | TP | TN | FN | FP | Sensitivity (%) | Specificity (%) | PPV (%) | NPV (%) | Accuracy (%) | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | | Value | Lower | Upper |
| All grades | | | | | | | | | | | | |
| 5 | -1.876 | 629 | 12,658 | 3671 | 230 | 14.6 | 98.2 | 73.2 | 77.5 | 77.3 | 77.0 | 77.6 |
| 10 | -1.323 | 1120 | 12,290 | 3180 | 598 | 26.0 | 95.4 | 65.2 | 79.4 | 78.0 | 77.6 | 78.5 |
| 15 | -0.982 | 1541 | 11,851 | 2759 | 1037 | 35.8 | 92.0 | 59.8 | 81.1 | 77.9 | 77.4 | 78.4 |
| 20 | -0.731 | 1900 | 11,351 | 2400 | 1537 | 44.2 | 88.1 | 55.3 | 82.5 | 77.1 | 76.6 | 77.7 |
| 25 | -0.524 | 2207 | 10,798 | 2093 | 2090 | 51.3 | 83.8 | 51.4 | 83.8 | 75.7 | 75.1 | 76.3 |
| 37,3 | -0.122 | 2894 | 9364 | 1406 | 3524 | 67.3 | 72.7 | 45.1 | 86.9 | 71.3 | 68.8 | 72.7 |
| Grades 3–4 | | | | | | | | | | | | |
| 5 | -1.841 | 350 | 6965 | 2014 | 123 | 14.8 | 98.3 | 74.0 | 77.6 | 77.4 | 77.0 | 77.8 |
| 10 | -1.319 | 622 | 6765 | 1742 | 323 | 26.3 | 95.4 | 65.8 | 79.5 | 78.2 | 77.6 | 78.7 |
| 15 | -0.985 | 841 | 6511 | 1523 | 577 | 35.6 | 91.9 | 59.3 | 81.0 | 77.8 | 77.1 | 78.4 |
| 20 | -0.739 | 1033 | 6230 | 1331 | 858 | 43.7 | 87.9 | 54.6 | 82.4 | 76.8 | 76.1 | 77.6 |
| 25 | -0.546 | 1196 | 5921 | 1168 | 1167 | 50.6 | 83.5 | 50.6 | 83.5 | 75.3 | 74.5 | 76.1 |
| 38,1 | -0.122 | 1608 | 5097 | 756 | 1991 | 68.0 | 71.9 | 44.7 | 87.1 | 70.9 | 67.7 | 72.3 |
| Grades 5–6 | | | | | | | | | | | | |
| 5 | -1.929 | 116 | 2262 | 651 | 38 | 15.1 | 98.3 | 75.3 | 77.7 | 77.5 | 76.8 | 78.3 |
| 10 | -1.309 | 199 | 2193 | 568 | 107 | 25.9 | 95.3 | 65.0 | 79.4 | 78.0 | 77.0 | 79.0 |
| 15 | -1.009 | 275 | 2115 | 492 | 185 | 35.9 | 92.0 | 59.8 | 81.1 | 77.9 | 76.7 | 79.2 |
| 20 | -0.738 | 350 | 2036 | 417 | 264 | 45.6 | 885 | 57.0 | 83.0 | 77.8 | 76.5 | 79.1 |

(*continued on next page*)

**Table A.2** (*continued*)

| Percentile | Cut-off[a] | TP | TN | FN | FP | Sensitivity (%) | Specificity (%) | PPV (%) | NPV (%) | Accuracy (%) | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 25 | -0.537 | 399 | 1932 | 368 | 368 | 52.0 | 84.0 | 52.0 | 84.0 | 76.0 | 74.6 | 77.4 |
| 31,1 | -0.300 | 473 | 1820 | 294 | 480 | 61.7 | 79.1 | 49.6 | 86.1 | 74.8 | 67.9 | 76.5 |
| Grades 7–9 | | | | | | | | | | | | |
| 5 | -1.924 | 164 | 3431 | 1005 | 69 | 14.0 | 98.0 | 70.4 | 77.3 | 77.0 | 76.4 | 77.6 |
| 10 | -1.339 | 299 | 3332 | 870 | 168 | 25.6 | 95.2 | 64.0 | 79.3 | 77.8 | 77.0 | 78.6 |
| 15 | -0.958 | 423 | 3222 | 746 | 278 | 36.2 | 92.1 | 60.3 | 81.2 | 78.1 | 77.1 | 79.0 |
| 20 | -0.710 | 518 | 3084 | 651 | 416 | 44.3 | 88.1 | 55.5 | 82.6 | 77.1 | 76.1 | 78.2 |
| 25 | -0.470 | 599 | 2931 | 570 | 569 | 51.2 | 83.7 | 51.3 | 83.7 | 75.6 | 74.4 | 76.7 |
| 35,6 | -0.123 | 768 | 2607 | 401 | 893 | 65.7 | 74.5 | 46.2 | 86.7 | 72.3 | 70.1 | 75.5 |

*Note*. After casewise deletion of missing values, the number of students in Grades 3–4: 9452; Grades 5–6: 3067; Grades 7–9: 4669. [a] Value of the standardized number processing skills sum variable. TP = true positive; TN = true negative; FN = false negative; FP = false positive; PPV = positive predictive value; NPV = negative predictive value.

## References


American Psychiatric Association. (2013). *Diagnostic and Statistical Manual of Mental Disorders*. https://doi.org/10.1176/appi.books.9780890425596

Ansari, D. (2008). Effects of development and enculturation on number representation in the brain. *Nature Reviews Neuroscience, 9*(4), 278–291. https://doi.org/10.1038/nrn2334

Arbuckle, J. L. (1996). Full information estimation in the presence of incomplete data. In G. A. Marcoulides, & R. E. Schumacker (Eds.), *Advanced structural equation modeling* (pp. 243–277). Lawrence Erlbaum Associates, Inc.

Aro, T., Eklund, K., Eloranta, A. K., Närhi, V., Korhonen, E., & Ahonen, T. (2019). Associations between childhood learning disabilities and adult-age mental health problems, lack of education, and unemployment. *Journal of Learning Disabilities, 52*(1), 71–83. https://doi.org/10.1177/0022219418775118

Bartelet, D., Ansari, D., Vaessen, A., & Blomert, L. (2014). Cognitive subtypes of mathematics learning difficulties in primary education. *Research in Developmental Disabilities, 35*(3), 657–670. https://doi.org/10.1016/j.ridd.2013.12.010

Bollen, K. A. (1989). *Structural equations with latent variables*. Wiley.

Bozikov, J., & Lijana, Z. K. (2010). Test validity measures and receiver operating characteristic (ROC) analysis. In L. Zaletel-Kragelj, & J. Boţikov (Eds.), *Methods and tools in public health: A handbook for teachers, researchers and health professionals* (pp. 749–770). Hans Jacobs Publishing Company.

Brankaer, C., Ghesquière, P., & De Smedt, B. (2014). Children's mapping between non-symbolic and symbolic numerical magnitudes and its association with timed and untimed tests of mathematics achievement. *PLOS ONE, 9*(4), Article e93565. https://doi.org/10.1371/journal.pone.0093565

Bremner, J. G., Slater, A. M., Hayes, R. A., Mason, U. C., Murphy, C., Spring, J., Draper, L., Gaskell, D., & Johnson, S. P. (2017). Young infants' visual fixation patterns in addition and subtraction tasks support an object tracking account. *Journal of Experimental Child Psychology, 162*, 199–208. https://doi.org/10.1016/j.jecp.2017.05.007

Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal, 14*(3), 464–504. https://doi.org/10.1080/10705510701301834

Cirino, P. T., Fuchs, L. S., Elias, J. T., Powell, S. R., & Schumacher, R. F. (2015). Cognitive and mathematical profiles for different forms of learning difficulties. *Journal of Learning Disabilities, 48*(2), 156–175. https://doi.org/10.1016/j.jecp.2015.12.010

Dehaene, S., Dupoux, E., & Mehler, J. (1990). Is numerical comparison digital? Analogical and symbolic effects in two-digit number comparison. *Journal of Experimental Psychology: Human Perception and Performance, 16*(3), 626. https://doi.org/10.1037/0096-1523.16.3.626

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

De Smedt, B., Noël, P.-L., 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

Desoete, A., & Grégoire, J. (2006). Numerical competence in young children and in children with mathematics learning disabilities. *Learning and Individual Differences, 16*(4), 351–367. https://doi.org/10.1016/j.lindif.2006.12.006

Devine, A., Soltész, F., Nobes, A., Goswami, U., & Szűcs, D. (2013). Gender differences in developmental dyscalculia depend on diagnostic criteria. *Learning and Instruction, 27*, 31–39. https://doi.org/10.1016/j.learninstruc.2013.02.004

Friso-Van den Bos, I., Van der Ven, S. H., Kroesbergen, E. H., & Van Luit, J. E. (2013). Working memory and mathematics in primary school children: A meta-analysis. *Educational Research Review, 10*, 29–44. https://doi.org/10.1016/j.edurev.2013.05.003

Geary, D. C. (2011). Consequences, characteristics, and causes of poor mathematics achievement and mathematical learning disabilities. *Journal of Developmental and Behavioral Pediatrics, 32*(3), 250–263. https://doi.org/10.1097/DBP.0b013e318209edef

Geary, D. C., Hoard, M. K., & Bailey, D. H. (2011). Fact retrieval deficits in low achieving children and children with mathematical learning disability. *Journal of Learning Disabilities, 45*, 291–307. https://doi.org/10.1177/0022219410392046

Geary, D. C., Hoard, M. K., Nugent, L., & Bailey, D. H. (2012). Mathematical cognition deficits in children with learning disabilities and persistent low achievement: A five-year prospective study. *Journal of Educational Psychology, 104*, 206–223. https://doi.org/10.1037/a0025398

Gross, J., Hudson, C., & Price, D. (2009). Every child a chance Trust/KPMG. *The Longest-term costs of numeracy difficulties*.

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

Huijsmans, M., Kleemans, T., & Kroesbergen, E. H. (2022). The cognitive profiles for different samples of mathematical learning difficulties and their similarity to typical development: Evidence from a longitudinal study. *Journal of Experimental Child Psychology, 214*, Article 105288. https://doi.org/10.1016/j.jecp.2021.105288

Jordan, N. C., Glutting, J., & Ramineni, C. (2010). The importance of number sense to mathematics achievement in first and third grades. *Learning and Individual Differences, 20*(2), 82–88. https://doi.org/10.1016/j.lindif.2009.07.004

Kißler, C., Schwenk, C., & Kuhn, J. T. (2021). Two dyscalculia subtypes with similar, low comorbidity profiles: A mixture model analysis. *Frontiers in Psychology, 12*, Article 589506. https://doi.org/10.3389/fpsyg.2021.589506

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

Laakso, M.-J., Kaila, E., & Rajala, T. (2018). ViLLE – collaborative education tool: Designing and utilizing an exercise-based learning environment. *Education and Information Technologies, 23*, 1655–1676. https://doi.org/10.1007/s10639-017-9659-1

Leino, K., Ahonen, A. K., Hienonen, N., Hiltunen, J., Lintuvuori, M., Lähteinen, S., Lämsä, J., Nissinen, K., Nissinen, V., & Puhakka, E. (2019). PISA 18: ensituloksia. Suomi parhaiden joukossa [PISA 18: First results. Finland among the best]. *Opetus-ja kulttuuriministeriön julkaisuja, 40*. https://jyx.jyu.fi/handle/123456789/75294.

Li, Y., Zhang, M., Chen, Y., Deng, Z., Zhu, X., & Yan, S. (2018). Children's non-symbolic and symbolic numerical representations and their associations with mathematical ability. *Frontiers in Psychology, 9*, 1035. https://doi.org/10.3389/fpsyg.2018.01035

Liu, Y., & Wong, T. T. Y. (2020). The growth rates of dot enumeration ability predict mathematics achievements: A 5-year longitudinal study. *British Journal of Educational Psychology, 90*(3), 604–617. https://doi.org/10.1111/bjep.12318

Lyons, I. M., Price, G. R., Vaessen, A., Blomert, L., & Ansari, D. (2014). Numerical predictors of arithmetic success in grades 1–6. *Developmental Science, 17*(5), 714–726. https://doi.org/10.1111/desc.12152

Major, C. S., Paul, J. M., & Reeve, R. A. (2017). TEMA and dot enumeration profiles predict mental addition problem solving speed longitudinally. *Frontiers in Psychology, 8*, Article 313803. https://doi.org/10.3389/fpsyg.2017.02263

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

Marsh, H., Hau, K.-T., & Wen, Z. (2004). In search of golden rules: comment on hypothesis-testing approaches to setting cutoff values for fit indexes and dangers in overgeneralizing Hu and Bentler's (1999) findings. *Structural Equation Modeling: A Multidisciplinary Journal, 11*(3), 320–341. https://doi.org/10.1207/s15328007sem1103_2

Mazzocco, M. M. M., & Räsänen, P. (2013). Contributions of longitudinal studies to evolving definitions and knowledge of developmental dyscalculia. *Trends Neuroscience Education, 2*(2), 65–73. https://doi.org/10.1016/j.tine.2013.05.001

Mazzocco, M. M. M., Devlin, K. T., & McKenney, S. J. (2008). Is it a fact? Timed arithmetic performance of children with mathematical learning disabilities (MLD) varies as a function of how MLD is defined. *Developmental Neuropsychology, 33*, 318–344. https://doi.org/10.1080/87565640801982403

McCrink, K., & Wynn, K. (2007). Ratio abstraction by 6-month-old infants. *Psychological Science, 18*(8), 740–745. https://doi.org/10.1111/j.1467-9280.2007.01969.x

Murayama, K., Pekrun, R., & Lichtenfeld, S. (2013). Predicting long-term growth in students' mathematics achievement: The unique contributions of motivation and cognitive strategies. *Child Development, 84*(4), 1475–1490. https://doi.org/10.1111/cdev.12036

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*, 458–478. https://doi.org/10.1177/0022219407040005090 1

Namkung, J. M., Peng, P., & Lin, X. (2019). The relation between mathematics anxiety and mathematics performance among school-aged students: A Meta-Analysis. *Review of Educational Research*. https://doi.org/10.3102/0034654319843494

Price, G. R., & Ansari, D. (2013). Dyscalculia: Characteristics, causes, and treatments. *Numeracy, 6*(1), 1–16. https://doi.org/10.5038/1936-4660.6.1.2

Psyridou, M., Koponen, T., Tolvanen, A., Aunola, K., Lerkkanen, M.-K., Poikkeus, A.-M., & Torppa, M. (2024). Early prediction of math difficulties with the use of a neural networks model. *Journal of Educational Psychology, 116*(2), 212–232. https://doi.org/10.1037/edu0000835

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. https://doi.org/10.1037/a0027520

Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J.-C., & Müller, M. (2011). pROC: An open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics, 12*, 77. https://doi.org/10.1186/1471-2105-12-77

Räsänen, P., Aunio, P., Laine, A., Hakkarainen, A., Väisänen, E., Finell, J., Rajala, T., Laakso, M.-J., & Korhonen, J. (2021). Effects of gender on basic numerical and arithmetic skills: Pilot data from third to ninth grade for a large-scale online dyscalculia screener. *Frontiers in Education, 6*, Article 683672. https://doi.org/10.3389/feduc.2021.683672

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

Skagerlund, K., & Träff, U. (2016). Number processing and heterogeneity of developmental dyscalculia: Subtypes with different cognitive profiles and deficits. *Journal of Learning Disabilities, 49*(1), 36–50. https://doi.org/10.1177/0022219414522707

Vanbinst, K., Ceulemans, E., Ghesquière, P., & De Smedt, B. (2015). Profiles of children's arithmetic fact development: A model-based clustering approach. *Journal of Experimental Child Psychology, 133*, 29–46. https://doi.org/10.1016/j.jecp.2015.01.003

Wang, L., Sun, Y., & Zhou, X. (2016). Relation between approximate number system acuity and mathematical achievement: The influence of fluency. *Frontiers in Psychology, 7*, 1966. https://doi.org/10.3389/fpsyg.2016.01966

West, S. G., Taylor, A. B., & Wu, W. (2012). Model fit and model selection in structural equation modeling. In R. H. Hoyle (Ed.), *Handbook of structural equation modeling* (pp. 209–231). Guilford Press.

World Health Organization. (2019). *International classification of Diseases* (https://icd.who.int/).

Xenidou-Dervou, I., De Smedt, B., van der Schoot, M., & van Lieshout, E. C. (2013). Individual differences in kindergarten math achievement: The integrative roles of approximation skills and working memory. *Learning and Individual Differences, 28*, 119–129. https://doi.org/10.1016/j.lindif.2013.09.012

Zhang, X., Räsänen, P., Koponen, T., Aunola, K., Lerkkanen, M. K., & Nurmi, J. E. (2020). Early cognitive precursors of children's mathematics learning disability and persistent low achievement: A 5-year longitudinal study. *Child Development, 91*(1), 7–27. https://doi.org/10.1111/cdev.13123