Author(s): Psyridou, Maria; Koponen, Tuire; Tolvanen, Asko; Aunola, Kaisa; Lerkkanen, Marja-Kristiina; Poikkeus, Anna-Maija; Torppa, Minna

Title: Early prediction of math difficulties with the use of a neural networks model

Year: 2024

Version: Accepted version (Final draft)

Please cite the original version:

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

Early Prediction of Math Difficulties with the Use of a Neural Networks Model

Maria Psyridou[1], Tuire Koponen[2], Asko Tolvanen[3], Kaisa Aunola[1], Marja-Kristiina Lerkkanen[4], Anna-Maija Poikkeus[4], Minna Torppa[4]

[1]Department of Psychology, University of Jyväskylä

[2]Department of Education, University of Jyväskylä

[3]Methodology Centre for Human Sciences, University of Jyväskylä

[4]Department of Teacher Education, University of Jyväskylä

Citation:

Psyridou, M., Torppa, M., Tolvanen, A., Poikkeus, A.-M., Lerkkanen, M.-K., & Koponen, T. (2023). Early prediction of math difficulties with the use of a neural networks model. *Journal of Educational Psychology*. https://doi.org/10.1037/edu0000835

Author Note

Maria Psyridou https://orcid.org/0000-0003-1425-2279

Tuire Koponen https://orcid.org/0000-0003-0039-1016

Asko Tolvanen https://orcid.org/0000-0001-6430-8897

Kaisa Aunola https://orcid.org/0000-0001-8866-6736

Marja-Kristiina Lerkkanen https://orcid.org/0000-0002-5709-5800

Anna-Maija Poikkeus https://orcid.org/0000-0001-7913-8691

Minna Torppa https://orcid.org/0000-0003-3834-9892

The data that support the findings of this study, materials, and analysis code are available on request from the authors. This study was not preregistered. We have no conflicts of interest to disclose.

Correspondence should be addressed to Maria Psyridou, P.O.Box 35, University of Jyväskylä, Kärki, Mattilanniemi 6, 40014-FIN. Phone: +(358) 408053016; E-mail: maria.m.psyridou@jyu.fi

Funding

The First Steps study was supported by the Academy of Finland (Grants 213486, 263891, 268586, and 292466). In addition, Maria Psyridou was supported by the Academy of Finland (Grant 339418). Minna Torppa was supported by the Academy of Finland (Grants 276239, 284439, and 313768) and by the European Research Council (ERC) under the European Union's Horizon 2020 Research and Innovation Program (Grant Agreement 101002966).

Abstract

The early prediction of math difficulties (MD) is important as it facilitates timely support. MD are multifaceted, and several factors are involved in their manifestation. This makes the accurate early prediction of MD particularly challenging. In the present study, we aim to predict MD in Grade 6 with kindergarten-age (age 6) measures by applying a neural networks model. We use a set of 49 variables assessed during kindergarten from the domains of early arithmetic skills, cognitive skills, the home learning environment, parental measures, motivation, behavioral problems, and gender, which have been shown to have associations with mathematical development and/or MD. A two-step approach was used: first, we examined whether the neural networks approach can provide a solution for the effective early identification of MD based on all 49 variables and, then, by using the most important predictors as identified by the initial model. The initial model achieved an area under the curve (AUC) of .818, demonstrating excellent performance. The most important predictors of Grade 6 MD came from the domains of arithmetic and cognitive skills (arithmetic skills, rapid automatized naming (RAN), number concepts, spatial skills, counting) and behavioral problems (attention-orientation). The model with only the most important predictors achieved an AUC of .776, indicating good performance. Our results provided proof of concept for using neural networks in MD prediction in Grade 6 using information already available in kindergarten. In schools, these results could be used to identify children at potential risk of developing MD and to provide access to early support.

Keywords: arithmetic, math difficulties, prediction, neural networks model, kindergarten-age

Educational Impact and Implications Statement

Approximately 4%–15% of children suffer from math difficulties (MD), and many more struggle with them without a formal diagnosis. MD have been shown to put children at increased risk of lower academic achievement, lower motivation, anxiety, depression, and even higher unemployment. Predicting MD accurately and early facilitates timely support. The current study demonstrates the potential of neural networks models to facilitate the early identification of those at risk of developing MD. The performance of our model provided proof of concept for using neural networks for the prediction of MD in Grade 6 using information already available in kindergarten. In a school setting, such prediction knowledge could be used to identify children at potential risk of developing MD and to provide access to early support.

Authors' note

The study was approved by the Ethical Committee of the University of Jyväskylä. At the beginning of the study, the children's parents and teachers provided informed written consent to participate.

The data used in this study and the analysis code are available upon request to the corresponding author. The study was not preregistered.

Early Prediction of Math Difficulties with the Use of a Neural Networks Model

Math difficulties (MD) refer to deficits in understanding and representing numerical magnitude, difficulties retrieving basic arithmetic facts from one's long-term memory, and delays in learning mathematical procedures despite having at least average intelligence (Geary, 2011a). The present study focuses on problems in arithmetic fluency, which are the most typical feature of MD (e.g., Geary et al., 1993). Approximately 4%–15% of children suffer from MD (Mazzocco & Myers, 2003; Shalev et al., 2005), and many more struggle with them without a formal diagnosis. Individual differences in math are already evident when students enter primary school (e.g., Aunola et al., 2004; Barnes et al., 2020; Garon-Carrier et al., 2018). Therefore, to facilitate early identification, the present study aims to predict MD in Grade 6 using information already available in kindergarten. This is important as predicting MD accurately and early facilitates timely support. MD have been shown to put children at increased risk of lower academic achievement, lower motivation, anxiety, depression, and even higher unemployment (e.g., Aro et al., 2019; Lundetræ et al., 2010; Magnuson et al., 2016; Parhiala et al., 2018). The absence of effective prediction mechanisms can result in delayed MD identification, which could increase the risk of such negative consequences.

However, predicting MD accurately is particularly challenging because, as in other neurodevelopmental disorders, there is no single core deficit causing MD, but rather a combination of various deficits that can vary from individual to individual (e.g., Pennington, 2006; Rubinsten & Henik, 2009). That is, predictive statistical approaches that can handle a larger number of predictors and identify non-linear associations are required. Neural networks models offer such an approach for prediction but are still new in the field of learning difficulties (e.g., Psyridou et al., 2023a). Therefore, the present study examines whether a neural networks approach can provide a solution for the effective early identification of MD with the use of kindergarten-age measures. In addition, the majority of what we know about the risk factors of

MD comes from models that examine linear associations (e.g., Kuhl et al., 2022). Neural networks, however, can identify linear and non-linear associations, interaction effects (e.g., several combinations of the independent variables that build cumulative risk), or a combination of these. In other words, there might not be a linear effect between a variable and MD, but the neural networks model might identify this variable as an important predictor of MD because of a non-linear effect or an interaction effect. As such, it is also possible that new knowledge with regard to reading difficulties and MD comorbidity can be gained with such a model.

Prediction Models

Two major goals related to the study of learning difficulties are inference and prediction. Inference creates a model from the data to formalize understanding or test a hypothesis. Prediction aims to forecast unobserved outcomes or future behaviors (Bzdok et al., 2018). Prediction allows, for example, the identification of whether an individual will experience MD without requiring an understanding of the underlying mechanisms that cause such difficulties. Both are necessary as inference aids the understanding of the underlying mechanisms that cause MD and enables the development of effective interventions and support systems, and prediction allows the early identification of those at risk of developing MD and can result in early access to support. Inference has been the main focus in the field of learning difficulties and, more generally, in science concerning human development. Recently, however, models that concentrate on prediction have emerged in other fields and have been shown to possess good accuracy (Choi et al., 2017; Lu et al., 2018; Mamoshina et al., 2016; Olsen et al., 2020). Such models need to be transferred to and tested in the field of learning difficulties in order to examine whether individuals with learning difficulties can be identified early on and, thereby, given access to timely support.

In the field of learning difficulties, statistical methods (including regression-type analyses and null hypothesis testing using, for example, t-tests and ANOVA) have a longstanding status, and they focus on inference. Knowledge generation has been dominated by classical statistics with the estimation of linear regression models and the statistical significance testing of whether an effect exists in a sample. These models were designed for data with few independent variables. However, in the case of neurodevelopmental disorders, the manifestation of a disorder depends on the presence and interactions of various factors (e.g., Pennington, 2006; Rubinsten & Henik, 2009), and the use of traditional linear models is not always helpful. As the number of independent variables increases, the possible associations between them also increase, causing the model to be more complex, the inferences to be less precise, and can also result in collinearity problems.

By contrast, modeling based on machine learning concentrates on prediction and aims to identify patterns in often rich and unwieldy data (Bzdok et al., 2017). The model learns from the data, which makes it well suited to addressing phenomena that are influenced by many factors, with possible complicated associations among them (Bzdok et al., 2017; Urban & Gates, 2021). The focus is on prediction, and a two-step approach is usually followed. First, a learning algorithm is fitted on a typically bigger amount of data (the training sample). Then, the ensuing model is evaluated using a typically smaller amount of data (the testing sample). This contrasts with classical statistical methods, where the aim is to reject the null hypothesis by considering the entire sample (Wasserstein & Lazar, 2016). In the present study, we used a neural networks model to predict individuals with MD using a broad set of kindergarten-age measures. In addition, we examined whether the predictive ability of the model changes when using only the most important predictors of MD. Deep artificial neural networks, a category of deep learning and a subcategory of machine learning (Goodfellow et al., 2016; Urban & Gates, 2021), are a promising methodology that is gaining attention.

Deep learning methods are a type of representation learning method with multiple levels of representation (LeCun et al., 2015). That is, they can automatically find the optimal representation from the raw data without requiring prior feature selection. This is obtained with the use of a hierarchical structure with different levels of complexity. Each level entails the application of non-linear transformations, which result in representation at a higher, more abstract level. For example, for classification tasks, higher layers of representation amplify features that are more significant for differentiation and suppress unrelated differences (LeCun et al., 2015). Deep learning has had a great impact on natural language processing (Graves et al., 2013) and game playing (Mnih et al., 2015; Silver et al., 2016) research, among others.

Deep neural networks are computational models inspired by how the human brain processes information. They are composed of many units that work in parallel and are arranged in interconnected layers: an input layer, one or more hidden layers that contain unobservable network units, and the output layer (Hinton et al., 2007; LeCun et al., 2015). The input layer includes the features, the data inputted into the model (e.g., cognitive skills, factors related to the home learning environment). The hidden layer(s) learn and save increasingly more abstract features of the data. These features travel to the output layer, that includes the target variable(s), which classifies the observations into categories (e.g., MD vs. no MD). The number of hidden layers represents the depth of the network. Each layer comprises a set of artificial neurons or "nodes" within which each neuron is connected to all the neurons in the previous layer. Each connection is associated with a weight value, which reflects the strength and direction of each neuron input, like a synapse between two biological neurons. Deep learning requires very little manual engineering; instead, the model learns the connections from the data (LeCun et al., 2015). A deep neural network learns to perform a specific task (e.g., prediction, classification) through training, during which the model learns the strength of the connections between the units in the layer(s) (Cichy & Kaiser, 2019; Urban & Gates, 2021). Once trained, the deep neural networks model can be used to perform the same task using new inputs (Cichy & Kaiser, 2019).

The use of neural networks for the prediction of learning difficulties has just started to gain attention. Psyridou et al. (2023a) used a neural networks model for the prediction of reading difficulties in the same sample used in this study, and they compared its results with those from linear and mixture models. Their results suggested that the neural networks model provided high accuracy in the prediction of both reading fluency and reading comprehension difficulties. In addition, it was shown that the neural networks model was the most accurate method, as compared to the linear and mixture models or a combination of them, for the early prediction of adolescent reading fluency and reading comprehension difficulties. These results conform to previous studies that have successfully used deep learning models in psychiatry (Calhoun & Sui, 2016; Vieira et al., 2017), in medicine and biomedicine (Mamoshina et al., 2016; Olsen et al., 2020), and in the prediction of various disorders, such as Alzheimer's disease (Lu et al., 2018), attention deficit hyperactivity disorder (ADHD) (Kuang & He, 2014), autism (Heinsfeld et al., 2018), and Parkinson's disease (Choi et al., 2017).

Despite the popularity that neural networks models have gained, there are some limitations and potential challenges in their application in the field of learning difficulties. One of limitations is the lack of interpretability. Neural networks models are often referred to as "black boxes" due to their lack of transparency in the decision-making process (Cichy & Kaiser, 2019). It can be difficult to understand how and why a neural networks model arrives at a particular conclusion, which can make it challenging to interpret the results and identify potential errors or biases. In the context of learning difficulties, the lack of interpretability of neural networks models can be a significant challenge. Understanding how a neural networks model arrived at a particular conclusion can be essential for diagnosing and addressing learning difficulties. However, despite these limitations, their ability to learn complex patterns in data and achieve state-of-the-art performance in various tasks can open the door for the early prediction of learning difficulties and thus the early access to support.

Risk Factors for MD

MD are most often linked to difficulties in basic math (i.e., simple arithmetic problems and memorizing basic facts; Geary, 2011a; Huijsmans et al., 2020). Complex math includes more complicated procedures, wherein stepwise problem solving is needed, as well as word problems. Overall, those with MD have been found to have comparable difficulties with basic and complex math skills (Kroesbergen et al., 2022). This could be because struggles with basic math will influence their performance on more complex math tasks (Kleemans et al., 2018), leading to difficulties with regard to the whole spectrum of math skills. In this study, we focused on the prediction of MD in arithmetic fluency. Taking advantage of the ability of the neural networks model to handle a large number of variables, we used a broad set of kindergarten-age measures as predictors. Consequently, in addition to well-known factors that are frequently shown to be associated with increased probabilities of developing MD, we used also less studied and less well-known elements; previous studies, though, have guided the selection of the measures (e.g., Bernabini et al., 2021; Cirino et al., 2018; Geary et al., 2018; Kroesbergen et al., 2022; Nelson & Powell, 2017; Psyridou et al., 2023b). In addition, the majority of what we know about the risk factors of MD comes from models that examine linear associations. However, neural networks can identify linear and non-linear associations, interaction effects, or a combination of these. In other words, there might not be a linear effect between a variable and MD, but the neural networks model might identify this variable as an important predictor of MD because there might be a non-linear effect or an interaction. By adding a broader set of predictors, it is possible that, with such a model, predictors that have not been studied in depth emerge as important features for the prediction of MD and new knowledge can be gained with regard to reading difficulties and MD comorbidity.

Due to the limited number of studies focusing on MD in arithmetic fluency per se, and the association between basic and complex math skills, we incorporated measures from various domains associated with arithmetic/mathematical development and/or MD. These domains included early arithmetic skills, cognitive skills, home learning environment, parental factors, motivation, behavioral problems, and gender. First, we included arithmetic fluency in kindergarten. Children's math skills in kindergarten have been shown to be associated with their later math skills, and children with lower skills in math in kindergarten tend to also have lower skills during the later grades (Aunola et al., 2004; Morgan et al., 2009). In addition to early math skills, the strategies children use to solve arithmetic problems have been shown to be related to their later achievements in math (Chu et al., 2018; Geary, 2011b; Geary et al., 2017). For this reason, we also included the kindergarten measures of the arithmetic strategies children use to solve problems as predictors of MD.

Various early numerical and non-numerical cognitive skills have been shown to be important indicators of later MD and arithmetic skills development (e.g., Bernabini et al., 2021; Cirino et al., 2018; Geary et al., 2007, 2018; Koponen et al., 2019; Moll et al., 2014a; Peng et al., 2018; Psyridou et a., 2023b). Of the cognitive skills that were included in the current study, some are closely related to later MD and arithmetic skills development, such as spatial relations (e.g., LeFevre et al., 2010; Psyridou et al., 2023b; Zhang & Lin, 2015), counting (e.g., Bernabini et al., 2021; Cirino et al., 2018; Desoete & Grégoire, 2006; Geary et al., 2009; Koponen et al., 2019; Nelson & Powell, 2017; Psyridou et al., 2023b), and number concepts (Geary et al., 2009, 2018; Kroesbergen et al., 2022; Psyridou et al., 2023b). Some others are more closely related to reading skills, such as oral language (including vocabulary and listening comprehension), letter knowledge, and word reading (e.g., Caravolas et al., 2019; Clayton et al., 2020; Psyridou et al., 2021). The rationale for including reading-related cognitive predictors in our prediction model was that MD and reading difficulties often co-occur (e.g., Joyner & Wagner, 2020; Koponen et al., 2018; Willcutt et al., 2013) and have shared predictors (e.g., Peng et al., 2020). Moreover, RAN has been associated with both literacy and arithmetic skills and MD (e.g., Donker et al., 2016; Georgiou et al., 2013; Koponen et al., 2017; Kroesbergen et al., 2022; Landerl et al., 2009; Pulkkinen et al., 2022; Psyridou et al., 2023b), while studies have shown contradictory results for the association between phonological awareness and MD (e.g., Amland et al., 2021; De Smedt & Boets, 2010; Liu et al., 2022; Vanbinst et al., 2014; see also Yang et al., 2022 for a meta-analysis).

The home learning environment refers to literacy and numeracy activities outside of traditional classroom settings, which can be grouped into more formal (i.e., explicit teaching) and more informal activities (i.e., integrated in play). This can be an important factor in math development (e.g., Lehrl et al., 2020; Napoli & Purpura, 2018). However, previous studies on the association between the home learning environment and children's math skills reveal some inconsistencies. Some have shown that the home learning environment is either associated with children's math skills (e.g., Kleemans et al., 2012; Lehrl et al., 2020; Manolitsis et al., 2013; Niklas & Schneider, 2014; Psyridou et al., 2023b) or with early skills that form math skills prerequisites (Dunst et al., 2017; Skwarchuk et al., 2014; Susperreguay et al., 2020). Other studies, though, have failed to identify such associations between the home learning environment and math development (e.g., Missall et al., 2015; Zippert & Rittle-Johnson, 2020), including a study based on the present sample (Khanolainen et al., 2020). However, as all the previous studies were based on regression-type analyses and null hypothesis testing methods, we decided to also include home learning environment measures. This is because neural networks models may reveal features that are important predictors due to interactions and/or non-linear associations. In addition to teaching numbers and arithmetic at home, we also included shared reading as, even without numerical content, it has been shown to support children's math skills (Lehrl et al., 2020; Napoli & Purpura, 2018). Similar to cognitive skills, we also included home learning environment measures that are related to literacy skills (i.e., teaching letters and reading) due to the relations that have been identified between the two domains.

The possible association between the home learning environment and later math skills could be due to masked genetic effects (e.g., Knafo & Jaffee, 2013; Taylor et al., 2010). Parental MD could be reflected in the home learning environment (e.g., via language that is used to avoid numerical concepts or that communicates negative beliefs and attitudes concerning math). They could also be reflected through math practices at home (e.g., fewer math activities or activities of a lower quality; e.g., Maloney et al., 2015; Missall et al., 2015; Susperreguy et al., 2020). Moreover, difficulties in reading or math may lead to lower levels of education, for example, due to dropouts or track selections (e.g., Hakkarainen, 2015; Magnuson, 2016). Previous studies have reported that parents' education, as well as their own difficulties in math and reading, is associated with children's math development (e.g., Silver & Libertus, 2022; Soares et al., 2018); this is also observed in the current sample (Khanolainen et al., 2020; Psyridou et al., 2023b). Therefore, in the present study, in addition to the home learning environment measures, we included parental math and reading difficulties, as well as parental education level, as predictors of children's MD.

In addition to cognitive skills, the home learning environment, and parental measures, the child's learning motivation and behavioral problems may also affect later math skills. Theories of motivation as well as related empirical research have shown that motivation plays an important role in students' learning and academic achievement in school (Eccles & Wigfield, 2002; Wigfield & Cambria, 2010). In the literature, learning motivation has been approached from different viewpoints, and there is evidence that motivation, when defined in terms of task-focused behavior, interests, and task values, is related to academic achievement. More persistent and task-focused behavior has been related to better learning outcomes (Elliot et al., 1999; Wigfield & Cambria, 2010). For example, higher task-avoidant behavior has been associated with less improvement in math skills (Hirvonen et al., 2012; Psyridou et al., 2023b). Interest provides positive learning opportunities, such as enhancing attention and goal setting (Hidi & Renninger, 2006). In their study, Fisher et al. (2012) reported a positive association between children's math interest and math skills as early as kindergarten (see also Viljaranta et al., 2009). In addition, higher task values are associated with higher levels of academic performance and achievement (Eccles et al., 1998; Murphy & Alexander, 2000). Finally, the self-concept of ability in math has been found to be associated with math performance (Cai et al., 2018; Psyridou et al., 2023b). In the present study, we included task-avoidant behavior, interest in reading and math, task values, and self-concept for reading and math as predictors of MD.

Behavioral problems may also predict later math performance. Overall, behavioral factors are associated with poor academic and learning outcomes (e.g., Darney et al., 2013), and poor academic outcomes are associated with behavioral problems (Becker & Luthar, 2002). Such associations have also been documented for math performance; students with behavioral problems have been found to perform more poorly in math compared to students without such problems (Mulcahy et al., 2014, 2016; Trout et al., 2003). For example, inattentive behavior has been found to be related to MD (Cirino et al., 2007; Gold et al., 2013). In the present study, we have included a variety of behavioral measures, namely attention, hyperactivity, impulsiveness, attention-orientation, planning, and disruptive behavior, as predictors of MD. Finally, gender was included in the study, although previous research has shown inconsistent findings to date (e.g., Hutchison et al., 2019; Moll et al., 2014b; Psyridou et al., 2023b; Wei et al., 2015).

The Present Study

In the present study, we aim to apply a neural networks model to examine whether we can already predict during kindergarten (age 6) who will develop MD (defined as scoring within the lowest 10% of the arithmetic fluency distribution) in Grade 6. In addition, we examine whether by using a smaller set of predictors (only the top-ranked features) the ability of the model to predict MD remains as high as when using the broader set of predictors. The neural networks model was selected because it combines multiple benefits for the research question at hand: it allows linear and non-linear associations, and interactions or combinations of these, between the predictors and the target variable to be identified, and it can handle many more variables than the more traditional regression-based approaches. It is a novel method which has been broadly used in other fields with very promising results. To the best of our knowledge, in the field of learning difficulties, it has so far been used only for the prediction of reading fluency and reading comprehension difficulties. Psyridou et al. (2023a) suggested that the neural networks model provided high accuracy in the prediction of both reading fluency and reading comprehension difficulties and that the neural networks model was the most accurate method, as compared to the linear regression and mixture models or a combination of them. Given these promising results it was of interest to examine how the neural networks model perform for the prediction of MD.

We use a broad set of measures assessed during kindergarten, including early arithmetic abilities (arithmetic skills and strategies), cognitive skills (number concepts, counting, spatial relations, RAN, phonological awareness, letter knowledge, word reading, vocabulary, listening comprehension), the home learning environment (teaching numeracy and literacy skills at home, shared reading), parental math and reading difficulties, parental education, motivational (interest in math and reading, task values, self-concept in numbers and counting and reading, task avoidance) and behavioral (attention, hyperactivity, impulsiveness, attention-orientation, planning, disruptive behavior) measures, and gender. The identification of a model with a small set of predictors and with good performance can help in establishing a research direction for future studies, for example, by suggesting which factors warrant being included in future models, and it is a step further toward the transfer of such models to everyday practice. Assessing all the possible predictors for every child is quite challenging in everyday life. If we manage to construct a model with a minimal group of predictors (e.g., by including only the most important predictors) without compromising the performance of the model, it will be easier to use it, for example, in schools, for the early identification of those at risk of developing MD.

Method

Participants and Procedure

The present study was part of the First Steps Study, a Finnish longitudinal study that includes data of approximately 2,000 children from kindergarten to Grade 6 (Lerkkanen et al., 2006-2016). At the beginning of the follow-up (kindergarten), 1,880 children were included, but when they entered school, all their classmates were also invited to participate. Over the years, the sample size varied. In Grade 6, we have data on the math skills of 1,817 participants. The sample was drawn from four municipalities: two in central, one in western, and one in eastern Finland. One municipality was mainly urban, one was mainly rural, and two included both urban and semi-rural environments. In three of the municipalities, the participants represented the entire age cohort of children, and in the fourth, the participating children comprised about half the age cohort. Of the parents who were contacted, 78%–89% agreed to participate in the study – depending on the town or municipality. Ethnically and culturally, the sample was very homogeneous and representative of the Finnish population, and the parental education levels were very close to the national distribution recorded in Finland (Statistics Finland, 2007). The university’s Ethical Committee approved the study, and all the participants provided informed written consent.

Measures

The children were assessed longitudinally: in kindergarten (fall 2006 and/or spring 2007) and in Grade 6 (spring 2013). The children's cognitive and arithmetic skills, home learning environment and parental, motivational, and behavioral measures were assessed in the fall and/or spring of kindergarten (i.e., aged 6), and their math skills were tested in Grade 6. The measures are described in Table 1.

Statistical Analysis

A math sum score was calculated using the participants' standardized scores on the two arithmetic and multiplication tasks in Grade 6. The math sum score variable was used in the analysis for the prediction of MD. A multilayer perceptron network (MLP) was used to produce the model for the prediction of MD in Grade 6 based on the kindergarten-age measures. The MLP analysis was conducted using the SPSS (Version 26) neural networks add-on module. We utilized the default functions provided by SPSS (https://www.ibm.com/docs/en/SSLVMB_26.0.0/pdf/en/IBM_SPSS_Neural_Network.pdf). The Identity activation function was employed for the output layer while the hyperbolic tangent (or tanh) function was used for the hidden layers. For the loss function, we opted for binary cross-entropy, as it is the default in SPSS for binary classification problems. Additionally, the optimization function used was scaled conjugate gradient descent. MD were defined as scoring in the lowest 10% of the math skills distribution (see also Appendix B in the online supplemental materials for the plots with the histograms and the distributions of math scores for the individuals belonging to the lowest 10%, the individuals belonging to the remaining 90%, and the whole sample). The selection of the cut-off matters, and it is always somewhat arbitrary. Different research groups may use somewhat different cut-offs for the definition of MD (e.g., lowest 10th, lowest 15th, or 25th percentile). The large sample of the present study allowed the selection of a rather strict cut-off for the identification of MD, which was preferred in order to be closer to the prevalence estimates of MD in population.

First, we examined whether data were missing completely at random (MCAR). Little's MCAR test (Little, 1988) suggested that the data were not MCAR, $\chi^2(6{,}103) = 7{,}867.40$, $p < .001$. Therefore, the participants with and without math data in Grade 6 were compared with regard to their kindergarten-age measures. The effect sizes (Cohen's d and Hedge's g) suggested either negligible or small differences (they ranged from .00 to .34). One of the limitations of the MLP is that it cannot be performed when there are missing cases. For this reason, we used an expectation-maximization algorithm in SPSS to impute missing data. For the imputation, we only used the kindergarten-age measures so that the Grade 6 math variable remained unchanged and independent from the kindergarten-age predictors. Individuals with missing data for the math variable in Grade 6 were thus excluded from the analysis. Imputation was performed for the kindergarten-age variables using only the kindergarten-age variables (i.e., without using information from the math scores in Grade 6). We chose to impute the missing cases instead of excluding them, because most of the 49 variables used as features had some missing cases. Excluding all cases with missing data would have reduced the sample size considerably.

When using a neural networks model, the sample is divided into two samples: the training and the testing samples. The training sample, usually the bigger part of the data, is used to train the model (i.e., the fit of the algorithm, that is, the estimation of the weights). To evaluate whether the model can be generalized, the testing sample is used with the ensuing model (that is, with the weights that were generated with the training sample). In the present study, the MLP was set to randomly choose 70% of the data for the training sample and the remaining 30% for the testing sample. We used an average model approach to evaluate the performance of multiple model architectures. We allowed the model to choose the number of hidden layers and units automatically. The model was trained 20 times across different training sets to ensure that the model is not overfitting to a specific training set. Each time the same training sample was used and as such there was no overlapping between the training and the testing samples.

Due to the imputation, extra caution was taken in order to randomly balance the imputed cases between the training (70% of the data, *n* = 1,272) and the testing (30% of the data; *n* = 545) samples. Balancing the imputed cases in the training and testing samples is important to prevent bias in the model performance evaluation. When imputing missing data using an expectation-maximization algorithm, the imputed values are based on the available data in the dataset. Therefore, imputed cases are not actual observations but are estimated values. This means that there is some level of uncertainty associated with these imputed values. To ensure that the model is exposed to a representative number of imputed values we tried to have similar number of imputed cases in the training and the testing samples. This approach could help to prevent bias in the model evaluation, leading to more accurate predictions and better generalization to new data.

The first step of our analysis was to balance the imputed cases between the training and testing samples. The split of the individuals between the training and the testing samples was random. We balanced the imputed cases by estimating the percentage of the missing cases (and essentially the percentage of imputed cases) in the training and the testing sample in the MD and no MD groups. We used 20 different seed values to randomly select participants and allocate them to the two groups that would be used for the training and the testing of the neural networks model later. For each seed, we calculated how many missing cases (and essentially how many imputed cases) were in the MD and no MD groups in the training and the testing samples. The seed with the smallest difference in the MD and no MD groups in the training and the testing samples was used further for the prediction of MD (Appendix C in the online supplemental materials). For the best identified seed, the testing and training sample difference (absolute value) of the percentage of imputed cases in the MD and no MD groups was .53%.

Once the best seed was identified, it was used for the prediction of MD. We followed a supervised approach; that is, the input data were “labelled” or associated with the true outcome (Kotsiantis et al., 2007). Such approaches are common in predictive analysis, for example, when making a diagnosis (Singh et al., 2016). In our study, the math variable was dichotomized, using as a cut-off the lowest 10% of the math distribution for identification of MD. To evaluate the performance of multiple model architectures, we used an average model approach where the model is allowed to choose the number of hidden layers and units automatically. The model was trained 20 times across different training sets to ensure that the model is not overfitting to a specific training set. Each time the same training sample was used which secures that there was no overlapping between the training and the testing samples. Consequently, 20 MLP models were trained (using the best identified seed and the same training-testing-split) to predict the dichotomized math variable. The predicted math scores were saved, and their mean was calculated.

Using the mean of the 20 predicted scores, we estimated the receiver operating characteristic (ROC) curve in order to test the ability of the model to predict MD. The ROC curve is plotted with the true positive rate (i.e., sensitivity) on the y-axis and the false positive rate (i.e., 1-specificity) on the x-axis. The sensitivity of a test is the proportion of people who test positive among all those who actually have the condition. The specificity of a test is the proportion of people who test negative among all those who do not have the condition. In the case of the present study, sensitivity represents the proportion of people that the model predicted as having MD among all those who actually had MD (True Positives / [True Positives + False Negatives]), while specificity represents the proportion of people that the model predicted as not having MD among all those who did not have MD (True Negatives / [True

Negatives + False Positives]). Overall, a test with high sensitivity is useful for ruling out a condition if an individual has tested negative. A test with high specificity is useful for ruling in a condition if an individual has tested positive. Sensitivity and specificity are characteristics of a test and, therefore, independent of the population. ROC curves compare sensitivity versus 1-specificity across a range of values for the ability to predict a dichotomous outcome. Each point on the ROC curve represents a sensitivity/1-specificity pair corresponding to a particular cut-off. The larger the area under the curve (AUC), the better the identification of those with and without MD.

Based on the ROC curve, we next calculated specificity, balanced accuracy, positive predictive values (also known as precision), and negative predictive values when the sensitivity was as close to .80 as possible. The value of .80 was selected as it was considered a good sensitivity value for our study. Balanced accuracy is a measure used to assess the performance of a classification model. We calculated the balanced accuracy instead of the accuracy because the former is more useful when the two groups are imbalanced (i.e., when one group appears much more than the other; Brodersen et al., 2010). The closer the balanced accuracy is to 1, the better the model can correctly classify observations. Finally, we calculated positive and negative predictive values as they describe an individual's probability of having the condition once the results are known. The positive predictive value is the probability that following a positive test result, that individual will really have that specific condition. The negative predictive value is the probability that following a negative test result, that individual will not really have that specific condition. In the case of our study, the positive predictive value represents the probability that after an individual has been predicted as having MD, that individual will really have MD (True Positives / [True Positives + False Positives] *or* [Sensitivity × Prevalence] / [(Sensitivity × Prevalence) + (1 - Specificity) × (1 - Prevalence)]), while the negative predictive value represents the probability that after an individual has been predicted as not having MD, that individual will not really have MD (False Positives / [True Positives + False Positives] *or* [Specificity × (1 - Prevalence)] / [(1 - Sensitivity) × Prevalence + Specificity × (1 - Prevalence)]). In contrast to sensitivity and specificity, the positive and negative predictive values are influenced by the prevalence of the condition in the population being tested (Akobeng, 2007).

Next, we examined which kindergarten-age measures were the most important for the prediction of MD. An independent variable importance analysis, which computes each kindergarten-age measure's importance in determining the neural network based on the combined training and testing samples, was conducted. The independent variable importance analysis performs a sensitivity analysis, which computes the importance of each predictor in determining the neural network. The analysis is based on the combined training and testing samples. The MLP analysis in SPSS uses the importance measure based on the percentage increase in mean squared error (%IncreaseMSE). This measure evaluates the contribution of each input variable to the prediction accuracy of the model by comparing the reduction in the mean squared error of the model when a particular variable is included versus when it is excluded. The higher the %IncreaseMSE, the more important the variable is to the model's predictive accuracy. The analysis was conducted 20 times with the best identified seed, and the mean of the normalized importance for each kindergarten-age factor was estimated. We used the normalized importance instead of the raw importance because they sum up to 1, allowing for a direct comparison of the relative importance of different variables. This was useful in our study as we wanted also to identify the most important predictors in the model and extract information for the relative contribution of different sets of predictors. The independent variable importance analysis was provided by the MLP.

Finally, a follow-up analysis using a procedure identical to the one described above (first, identification of the seed with the best balance of imputed cases between the training and the testing samples and training of the model 20 times using the best identified seed), but only including the top-ranked features (i.e., measures with average normalized importance higher than 50% in the initial model) for the prediction of MD, was conducted. This was carried out to examine whether a model based on only the top-ranked features, instead of the all the features used in the initial model, continued to achieve similar performance. This allowed the identification of a minimal group of predictors that achieved good performance in terms of the identification of MD.

Transparency and Openness

This manuscript has been prepared according to the standards described in the Journal Article Reporting Standards (JARS; Kazak, 2018) for the American Psychological Association. We have reported all necessary study information, data exclusions and manipulations, and all measures used in the study. The data used in this study and the analysis code are available upon request to the corresponding author. Data were analyzed using SPSS (Version 26). The design and the analysis of this study were not preregistered.

Results

Descriptive Statistics and Correlations

The descriptive statistics of the kindergarten-age measures, the arithmetic and multiplication measures, and the math sum score variable in Grade 6 are presented in Table 2. The correlation coefficients (Pearson and Spearman) between all the assessed measures can be found in Appendix A in the online supplemental materials. The strongest correlations between the math variable and the kindergarten-age measures were for letter knowledge (fall) (r = .30), counting (fall) (r = .41), counting (spring) (r = .39), spatial relations (r = .27), arithmetic (r = .45), and the use of the strategy of immediately providing the answer from memory (Strategy 1) (rho = .32).

Prediction Model for MD: Architecture and Classification Performance

We used an average model approach to evaluate the performance of multiple model architectures. For the best identified seed, based on the models produced, all models had 1 hidden layer (which is the default selection for the automatic architecture selection) and the number of units varied from three to nine; in particular there were two models with three units, three models with five units, three models with six units, six models with seven units, five models with eight units, and one model with nine units. Overall, the most common architecture was one hidden layer with seven or eight units.

For the best identified seed, the ROC curve suggested that the MLP had good classification performance (Figures 1 and 2 for the testing sample; Appendix D in the online supplemental materials for the training sample). For the prediction of MD, the AUC was .818, (p < .001, 95% C.I. .769–.867). Thus, there is an 81.8% chance that the model will distinguish between those with and without MD.

Next, based on the ROC analysis, we calculated 1-specificity, specificity, balanced accuracy, and positive and negative predictive values when the sensitivity was as close to .80 as possible. Based on the coordinates of the ROC curve, the closest sensitivity value to .80 was .672. Based on the ROC analysis, when the sensitivity was .672, the 1-specificity ranged from .238 to .246, and the specificity ranged from .754 to .762 (there was a range for the specificity because, as shown in the ROC curve, for the same sensitivity, we had a range of 1-specificity values). That is, when 67.2% of those with MD were correctly predicted by the model as having MD (sensitivity), 75.4%–76.2% of those without MD were correctly predicted by the model as not having MD (specificity). In addition, the balanced accuracy ranged from .713 to .717 (the range for the balanced accuracy was because it was calculated based on the sensitivity and specificity values), suggesting moderate performance of the model in terms of identifying those with and without MD. In addition, the positive predictive value (precision) ranged from .245

to .252, while the negative predictive value ranged from .954 to .955. That is, 24.5%–25.2% of individuals predicted by the model as having MD had MD, and 95.4%–95.5% of individuals predicted by the model as not having MD did not have MD. Given the low prevalence of MD, the low positive predictive values and the high negative predictive values are to be expected. The chances are high that someone does not have MD. So, even if the percentage of false positives (those predicted as having MD when actually they did not have MD) is very small among all the people who did not have MD, their number in comparison to the true positives (those predicted as having MD and who indeed had MD) is likely to be quite high when > 90% of the sample does not have MD. See Table 3 for the corresponding sensitivity, specificity, and precision estimates for different cut-off values and Table 4 for the specificity and precision values corresponding to specific sensitivity values.

Prediction Model for MD: Predictive Features

To identify the most important kindergarten predictors, an independent variable importance analysis was conducted using all 49 measures as features (i.e., data inputted into the model, such as cognitive skills and home learning environment measures). The top-ranked (with average normalized importance higher than 50%) features for the prediction of MD were RAN (71.16%), spatial relations (61.49%), arithmetic skills in kindergarten (56.23%), number concepts (53.89%), attention-orientation (50.94%), and counting (spring) (50.00%; Figure 3). The correlation coefficients (Appendix A in the online supplemental materials) between the math scores in Grade 6 (arithmetic fluency; sum score of arithmetic and multiplication) and RAN, spatial relations, number concepts, counting (spring), and the arithmetic skills in kindergarten were positive, suggesting that lower scores in terms of these skills were associated with lower performance regarding math skills. Attention-orientation, on the other hand, was negatively associated with math scores, suggesting that attention-orientation problems were associated with lower math scores in Grade 6. All the correlations for these six features were statistically significant. It should be noted, though, that these are linear correlations, while the MLP, in addition to linear effects can also identify non-linear interactions, or combinations.

Finally, a follow-up analysis to predict MD using an identical procedure for the MLP to the one described above but including as features only the top-ranked features (i.e., measures with average normalized importance higher than 50%), was conducted. This was conducted to examine whether a model based on only the six top-ranked features, instead of the 49 used in the initial model, continued to achieve good performance. The follow-up analysis with a model using RAN, spatial relations, arithmetic skills in kindergarten, number concepts, counting (spring), and attention-orientation, as features for the prediction of MD, achieved good classification accuracy.

The ROC analysis indicated that the AUC was .776, (p < .001, 95% C.I. .719–.834). Based on the ROC analysis, when the sensitivity was .586, the 1-specificity ranged from .228 to .232, the specificity ranged from .768 to .772, the precision ranged from .231 to .234, and the balanced accuracy ranged from .677 to .679. Thus, even by including only the six top-ranked features, the model achieved relatively similar performance for the prediction of MD in our sample as the model with the broader set of measures. Other features included in the initial model account for a 3.60%–3.80% improvement in the balanced accuracy in the model for the prediction of MD.

Discussion

MD are multifaceted, and several factors are involved in their manifestation. This makes the accurate early prediction of MD particularly challenging. Neural networks models have been gaining considerable attention due to their advantages over classical statistical methods that focus on the estimation of linear regression models and statistical significance testing with regard to whether an effect exists in the sample. In the present study, we applied a neural networks model, which can identify linear and non-linear associations, interaction effects, or combinations of these between the predictors and the target variable while allowing for the inclusion of many correlated variables that more traditional regression-based approaches cannot handle well. To predict Grade 6 MD, we used a broad set of measures assessed during kindergarten, including early arithmetic, cognitive skills, home learning environment, parental, motivational, and behavioral measures, and gender. The results confirmed our initial hypothesis: neural networks can provide good accuracy in the prediction of MD, and they can be a useful tool for MD prediction.

Our findings showed proof of concept for using a neural networks model in early MD prediction. In a school setting, early prediction results could be used to identify children at potential risk of developing MD and to provide access to early support. The results show that the model classified participants with and without MD above chance in our data. The model achieved an AUC of .818, demonstrating excellent performance (Mandrekar, 2010). This suggests an 81.8% chance that the model will correctly distinguish those with and without MD based on kindergarten-age measures. When 67.2% of those with MD were correctly identified as having MD, 75.4%–76.2% of those without MD were correctly identified as not having MD. The most important measures for the prediction of MD in Grade 6 were RAN, spatial relations, number concepts, counting (spring), attention-orientation, and arithmetic skills.

There is always a tradeoff between sensitivity and specificity, and tests with high sensitivity have low specificity. That is, it is typical that when models are good at identifying actual cases, they also have a fairly high rate of false positives. To further assess an individual's probability of having MD, we calculated two additional metrics: positive and negative predictive values (Altman & Bland, 1994). These values are influenced by the prevalence of the condition in the population that is being tested, while the sensitivity and specificity are characteristics of the test, and the population does not affect the results. In our study, the positive predictive value ranged from .245 to .252, while the negative predictive value ranged from .954 to .955. This suggests that 24.5%–25.2% of the individuals predicted by the model to have MD had MD, and 95.4%–95.5% of the individuals predicted by the model as not having MD did not have MD. The model was thus very precise in the prediction of those who did not have MD, but among those predicted by the model as having MD, many did not actually have MD. This finding is of course to be expected as > 90% of the sample did not have MD. The chance is therefore that someone does not have MD. However, as the prediction is not 100% accurate, some individuals predicted as not having MD actually had MD (false negatives), and others predicted as having MD did not actually have MD (false positives).

In general, when the prevalence of a condition is low, the positive predictive value will also be low, even when using a test with high sensitivity and specificity, and, consequently, a proportion of those with a positive result may not necessarily have the condition (Akobeng, 2007). To handle the imbalanced data, we used an oversampling technique. We first oversampled the MD group in the training sample and as such created a new balanced training dataset which included approximately 1,100 cases in the MD and no MD groups. We then trained the MLP model in this new training dataset and evaluated its performance using the unchanged testing dataset. By evaluating the model on an unchanged testing dataset, we were able to assess how well the model generalizes to new data. The same procedure has been followed as when the unbalanced dataset was used (for the results, see Appendix E in the online supplemental materials). As shown by the results, the model with balanced data did not perform better than the model with the unbalanced data indicating that the unbalanced data might have better represented the real-world distribution of the target variable. If the true distribution of the target variable is imbalanced, then artificially balancing the training data might have introduced bias into the model. Another reason could be that the original unbalanced data might have contained valuable information that was lost during the balancing process. By oversampling the minority group, some of the original variability in the data might have been lost, making it harder for the model to discern patterns in the majority class. Of course, though, a possible reason could also be that the technique used to handle the imbalanced data is not optimal. Additionally, given that the reliability of the math sum score used in the MLP was not 1 (it was .73), it is possible that the model may encounter difficulties accurately distinguishing between individuals with MD and those without.

Other reasons for the high rate of misclassification, particularly of those not having MD and predicted as having MD, can be related to the nature of the phenomenon being studied. First, we aim to predict who is at risk of manifesting MD in Grade 6 with skills and factors assessed in kindergarten. However, even individuals who share similar risk factors may follow different pathways making it challenging for the model to accurately distinguish between those with and without MD. For instance, a recent study on the developmental profiles of arithmetic skills in the same sample as the current study found that a group with persistent arithmetic difficulties and a group with delayed onset but average performance later on shared similar risk factors in kindergarten. The distinguishing factors between them were limited to four cognitive skills (letter knowledge, counting, number concept, and RAN) and task avoidant behavior (Psyridou et al., 2023b). Furthermore, learning is a dynamic phenomenon, and there are other factors that can influence the development of math skills during primary school which are not considered in the current model. One crucial factor is the support individuals receive at school. In Finland, access to extra support is extensive (e.g., over 20% of comprehensive school pupils receive intensified or special support; Statistics Finland, 2021) and can have a significant impact on the development of math skills throughout the school years. This support can, for instance, compensate for early difficulties and contribute to improved performance. An additional reason could be that math skills in the general population follow a normal distribution. There is not a clear threshold above which individuals have good math skills, so an arbitrary cut-off is needed to determine who has MD. In our study, the lowest 10% were classified as having MD. However, those scoring just above or just below the cut-off do not present large differences. While we can identify individual differences, we can still identify individuals scoring just above or just below the cut-off that have similar risk factors. This lack of clear distinction between those with and without MD may also contribute to the high rate of misclassifications in our study. Although the current approach of identifying MD based on an arbitrary cut-off can result in such misclassifications, where we identify too many children at-risk for developing MD, the identification of children in need of early intervention is important. While this may not be cost-effective, it is ethically justified if the ultimate goal is to provide support to children in need.

To the best of our knowledge, this is the first study that used a neural networks model to predict the manifestation of MD. Consequently, we cannot directly compare our results with those of other similar studies. Even though previous studies have examined the developmental trajectories and growth paths of math skills, using data from both the current sample (Zhang et al., 2020) and different samples (e.g., Little et al., 2021), suggesting that there are groups with different growth trajectories, there is still a lack of studies that use predictive methods, such as the neural networks model, for the early identification of those at risk of developing MD. It seems, though, that the results confirm our initial hypothesis that neural networks can be a useful tool for the early prediction of MD. In this respect, the results conform to the results of the previous study that used a neural networks model to predict reading difficulties (Psyridou et al., 2023a), as well as those of other studies that have used deep learning models for the prediction of various disorders, such as Alzheimer’s disease (Lu et al., 2018), ADHD (Kuang & He, 2014), autism (Heinsfeld et al., 2018), Parkinson’s disease (Choi et al., 2017), and heart failure (Olsen et al., 2020), and that reported high classification accuracy. Possible reasons for this high performance are that they can identify complex patterns in data, allowing them to make more accurate predictions (Durstewitz et al., 2019), that they can work with a large number of variables and in cases in which many weak factors correlated with each other influence the phenomenon (Urban & Gates, 2021), and that they can identify linear, non-linear, and interactional effects (e.g., several combinations of the independent variables that increase the cumulative risk), or a combination of these, which, for example, a linear model cannot identify.

The findings suggest that specific cognitive and arithmetic skills and behavioral measures are among the top-ranked measures for the prediction of MD. The top-ranked (with average normalized importance higher than 50%) features for the prediction of MD were RAN (71.16%), spatial relations (61.49%), arithmetic skills in kindergarten (56.23%), number concepts (53.89%), attention-orientation (50.94%), and counting (spring) (50.00%; Figure 3). This is in line with previous studies showing the association of these skills with later math skills and MD (e.g., Bernabini et al., 2021; Cirino et al., 2018; Geary et al., 2018; Koponen et al., 2019; Kroesbergen et al., 2022; Nelson & Powell, 2017; Psyridou et al., 2023b; Zhang et al., 2020). Arithmetic skills in kindergarten were among the most important predictors of MD. This was expected as previous studies have shown that children's math skills in kindergarten are associated with later math skills and that children with lower math skills in kindergarten tend to also have lower skills during the later grades (Aunola et al., 2004; Morgan et al., 2009).

RAN was the most important measure for the prediction of MD. Previous studies of RAN as a predictor of difficulties have mainly focused on reading. However, some findings suggest that RAN is associated with MD (Donker et al., 2016; Kroesbergen et al., 2022) and that those with dysfluent arithmetic skills in Grade 3 have slow RAN at the end of the first grade (and onwards) (Pulkkinen et al., 2022). The present study extends previous studies by showing that RAN measured before entering school can be an important predictor of MD six years later and even more important than the included math domain specific measures. Our findings provide further evidence that RAN is an important risk indicator for later difficulties in arithmetic fluency and that children with naming speed problems should receive particular attention regarding their calculation fluency development during their primary school years, along with timely and targeted support, when delays in the development of calculation skills are identified.

The importance of spatial relations in arithmetic skills has also been shown in previous studies (e.g., LeFevre et al., 2010; Zhang & Lin, 2015; Zhang et al., 2020). Children with better spatial skills also have better arithmetic skills (e.g., Zhang et al., 2020; see Mix & Cheng, 2012 for a review). Four possible explanatory accounts have been suggested for the link between visuospatial skills and arithmetic/MD: the spatial representation of numbers, shared neural processing, spatial modeling, or working memory (Hawes & Ansari, 2020). Counting and number concepts are also important features for the prediction of MD. Several previous studies have shown that counting (e.g., Bernabini et al., 2021; Cirino et al., 2018; Desoete & Grégoire, 2006; Geary et al., 2009; Koponen et al., 2019; Nelson & Powell, 2017; Psyridou et al., 2023b) and number concepts (Geary et al., 2009, 2018; Kroesbergen et al., 2022; Psyridou et al., 2023b) are strong predictors of later arithmetic skills. Counting and number concept skills form a foundation for learning basic arithmetic skills during the first school years, and they are also necessary for learning more complex math later. Previous findings have also suggested that the mapping between Arabic digits and numerical magnitudes is important for mathematical development and that this mapping process might be weakened in children with MD (Brankaer et al., 2014; De Smedt et al., 2013).

In addition to the cognitive skills in kindergarten, attention-orientation was among the most important early predictors of MD. Problems in attention-orientation led to higher chances of having MD. Children’s behaviors have been shown to play an important role in their performance in math (e.g., Fitzpatrick & Pagani, 2013; Hirvonen et al., 2012; Merrell & Tymms, 2001; Onatsu-Arvilommi & Nurmi, 2000; Sims et al., 2016). Attention problems have been shown to predict later math skills (Duncan et al., 2007; Hassinger-Das et al., 2014; Sims et al., 2016) and the slower development of such skills (Fuchs et al., 2005). Math skills development requires attention during learning activities. Consequently, children with attention problems may experience more difficulties compared to their peers who do not face such challenges. In addition, children with attention difficulties may have more problems with regard to ignoring irrelevant information during cognitive tasks (Marzocchi et al., 2002). Our attention-orientation task also included questions on the child's flexibility moving from one task to another (i.e., whether the child only starts after coaxing and if they get stuck in an old solution model or on a previous task). Such difficulties could possibly slow down a child during an arithmetic fluency task, such as that used in this study, leading to lower math scores.

Interestingly, parental MD was not among the top-ranked features for the prediction of MD, although previous studies have suggested a genetic basis of MD (Soares et al., 2018), and previous studies using the current sample and the same measure for parental MD have shown an association between parental MD and children's arithmetic skills (Khanolainen et al., 2020; Psyridou et al., 2023b). A possible reason is that the inclusion of cognitive skills in the model hides parental MD's effect on MD. Another reason could be the measure used to assess family risk in this study: a child was considered as having family risk if either the mother or father reported some or clear difficulties in math using one item. This assessment method is not as sensitive as a more formal assessment, considering the heritability of MD (Soares et al., 2018).

The home learning environment and motivation measures were not among the top-ranked predictors of MD either. The normalized importance of the home learning environment measures ranged from 28.83% to 43.26%. Interestingly, the two most important measures (teaching reading at home and teaching letters at home) for the prediction of MD among the home learning environment measures were related to home literacy measures rather than home numeracy measures. The normalized importance of the motivation measures ranged from

20.95% to 34.13%, with self-concept for literacy being the most important measure for the prediction of MD among the motivation measures. Even though self-concept of ability in math has been found to be associated with math performance (Cai et al., 2018; Psyridou et al., 2023b) in our study self-concept of literacy ability seemed to be a somewhat more important feature for the prediction of MD than self-concept of math ability (34.13% vs. 24.89%).

As shown by the model based on the top-ranked features, the balanced accuracy, sensitivity, and specificity of the models change as the features used for the training of the model change. However, the inclusion of the top features alone (6 features instead of 49) only led to a 3.60%-3.80% decline in balanced accuracy. Thus, other features can also contribute to the accurate prediction of MD. Future studies with different features are needed to identify the best minimal group of features that can be used for the prediction of MD, and this may lead to even more accurate models. From the current study, though, it seems that cognitive skills (RAN, spatial relations, counting, number concepts), behavioral measures (attention-orientation), and early arithmetic skills are warranted. Notably, the model with only six measures as predictors used in the current study indicated also good performance, providing an important step forward for the possible transfer of such models to everyday practice. Assessing all 49 measures included in the initial model for every child is quite challenging in everyday life, whereas evaluating only six skills is manageable. The current study suggested that even by including these six measures in the predictive model, the performance of the model is not hugely compromised as there was a 3.60%-3.80% drop in the balanced accuracy, and the AUC dropped from .818 to .776, suggesting that there is an 77.6% instead of an 81.8% chance that the model will distinguish between those with and without MD. Consequently, such a model with a small group of measures as predictors will be easier to use during, for example, kindergarten or early school years, for the early identification of those at risk of developing MD in order to allow early access to support. More research is needed, though, as there is still high rate of misclassifications, especially for those who are predicted as having MD despite not having MD.

This study has a number of limitations. First, not all the measures that are relevant to math skills were included. For example, magnitude comparison, especially symbolic number comparison, has been suggested to be a core predictor of math skills as well as of MD (e.g., Schneider et al., 2017). Including all the relevant measures could lead to even better predictive models in future studies. This will also allow the identification of the best minimal group of features that can be used and may lead to even more accurate models for the prediction of MD than that identified in the current study. Second, the assessment of some of our kindergarten measures was not optimal. The variables for parental math and reading difficulties were based on self-reports with a single question. This assessment may not have provided an accurate evaluation of parental difficulties and, thus, may have possibly underestimated the predictive power of parental difficulties. Moreover, the reliability estimate for listening comprehension was quite low. Such low reliability could lead to the underestimation of this skill in the prediction of MD. Third, because of the longitudinal nature of the testing, some of the participants had missing values. We balanced the imputed cases between the training and testing models, but not having missing values would have been optimal. Fourth, we did not assess for or obtain a definitive diagnosis of MD, and, instead, MD were defined as scoring in the lowest 10% of the arithmetic fluency distribution. The selection of the cut-off matters as it is always somewhat arbitrary, and the outcomes of the studies might be affected by this (see Psyridou et al., 2020). However, although the use of cut-offs is likely to lead to uncertainties in research findings because of measurement error (Branum-Martin et al., 2013; Francis et al., 2005; Psyridou et al., 2020), they are also a practical tool for the identification of children with MD. The large sample of the present study allowed the selection of a rather strict cut-off for the identification of MD. Fifth, the top-ranked features for the prediction of MD were selected with a rather arbitrary cut-off (those with normalized importance of 50% or higher). As shown in the follow-up analysis with the top-ranked features, other features also contribute to the accurate prediction of MD. Additionally, while evaluating the performance of the model using only the top-ranked features, we used data from the same longitudinal study as the one used to identify the features. Ideally, independent datasets should be used to test the generalizability of the model. However, due to the unavailability of other longitudinal studies containing the same variables, we used data from the same study.

Finally, there is one limitation related to the neural networks models. Neural networks models have limited interpretability, which can be a significant challenge for diagnosing and addressing learning difficulties. Their lack of transparency in the decision-making process and difficulty in understanding how and why a particular conclusion was reached can make it challenging to interpret results and identify potential errors or biases. For example, although the importance of each predictor could be evaluated, it is challenging to disentangle whether non-linear effects of single variables or interactions are in place, and in which exact ways the different variables impact the outcome. Future research should examine how this could be improved, or how other models (e.g., random forest, LASSO regression) compare with neural networks models in this respect as well as in prediction accuracy.

In conclusion, this study demonstrated the potential of neural networks models to facilitate the early identification of those at risk of developing MD. The present study is, to the best of our knowledge, the first to apply neural network models for predicting MD, and our approach involved an average model approach to evaluate the performance of multiple model architectures. Our analysis revealed that for the simple numerical variables used in this study, although there was some variation in the number of units in the hidden layer, the most common structure involved seven or eight units. With respect to practical implications, these findings provide educators with the impetus to be mindful of the high likelihood of MD in children, particularly with respect to those with low arithmetic, number, naming fluency, or visuo-spatial skills in kindergarten and behavioral issues such as problems in attention-orientation as it seems to predict also learning outcome in math.

References

Akobeng, A. K. (2007). Understanding diagnostic tests 1: sensitivity, specificity and predictive values. *Acta Paediatrica, 96*(3), 338-341. https://doi.org/10.1111/j.1651-2227.2006.00180.x

Altman, D. G., & Bland, J. M. (1994). Statistics Notes: Diagnostic tests 2: predictive values. *Bmj, 309*(6947), 102. https://doi.org/10.1136/bmj.309.6947.102

Amland, T., Lervåg, A., & Melby-Lervåg, M. (2021). Comorbidity between math and reading problems: Is phonological processing a mutual factor? *Frontiers in Human Neuroscience, 14*(592), 1–11. https://doi.org/10.3389/fnhum.2020.577304

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

Aunola, K., Leskinen, E., Lerkkanen, M.-K., & Nurmi, J.-E. (2004). Developmental dynamics of math performance from preschool to grade 2. *Journal of Educational Psychology, 96*, 699-713. https://doi.org/10.1037/0022-0663.96.4.699

Barnes, M. A., Clemens, N. H., Fall, A. M., Roberts, G., Klein, A., Starkey, P., ... & Flynn, K. (2020). Cognitive predictors of difficulties in math and reading in pre-kindergarten children at high risk for learning disabilities. *Journal of Educational Psychology, 112*(4), 685. https://doi.org/10.1037/edu0000404

Becker, B. E., & Luthar, S. S. (2002). Social-emotional factors affecting achievement outcomes among disadvantaged students: Closing the achievement gap. *Educational Psychology, 37*, 197-214. https://doi.org/10.1207/S15326985EP3704_1

Bernabini, L., Bonifacci, P., & de Jong, P. F. (2021). The relationship of reading abilities with the underlying cognitive skills of math: A dimensional approach. *Frontiers in Psychology*, *12*, 577488. https://doi.org/10.3389/fpsyg.2021.577488

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), e93565. https://doi.org/10.1371/journal.pone.0093565

Branum-Martin, L., Fletcher, J. M., & Stuebing, K. K. (2013). Classification and identification of reading and math disabilities: The special case of comorbidity. *Journal of Learning Disabilities*, *46*(6), 490-499. https://doi.org/10.1177/0022219412468767

Brodersen, K. H., Ong, C. S., Stephan, K. E., & Buhmann, J. M. (2010, August). The balanced accuracy and its posterior distribution. In 2010 *20th international conference on pattern recognition* (pp. 3121-3124). IEEE.

Bzdok, D., Altman, N. & Krzywinski, M. (2018). Statistics versus machine learning. *Nature Methods*, *15*(4), 233-234. https://doi.org/10.1038/nmeth.4642.

Bzdok, D., Krzywinski, M., & Altman, N. (2017). Machine learning: a primer. *Nature Methods*, *14*(12), 1119. https://doi.org/10.1038/nmeth.4526

Cai, D., Viljaranta, J., & Georgiou, G. K. (2018). Direct and indirect effects of self-concept of ability on math skills. *Learning and Individual Differences*, *61*, 51-58. https://doi.org/10.1016/j.lindif.2017.11.009

Calhoun, V. D., & Sui, J. (2016). Multimodal fusion of brain imaging data: a key to finding the missing link(s) in complex mental illness. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, *1*(3), 230-244. https://doi.org/10.1016/j.bpsc.2015.12.005

Caravolas, M., Lervåg, A., Mikulajová, M., Defior, S., Seidlová-Málková, G., & Hulme, C. (2019). A cross-linguistic, longitudinal study of the foundations of decoding and reading comprehension ability. *Scientific Studies of Reading, 23*(5), 386–402. https://doi.org/10.1080/10888438.2019.1580284

Choi, H., Ha, S., Im, H. J., Paek, S. H., & Lee, D. S. (2017). Refining diagnosis of Parkinson's disease with deep learning-based interpretation of dopamine transporter imaging. *NeuroImage: Clinical, 16*, 586–594. https://doi.org/10.1016/j.nicl.2017.09.010

Chu, F.W., van Marle, K., Rouder, J., & Geary, D. (2018). Children's early understanding of numbers predicts their later problem-solving sophistication in addition. *Journal of Experimental Child Psychology, 169*, 73-92. https://doi.org/10.1016/j.jecp.2017.12.010

Cichy, R. M., & Kaiser, D. (2019). Deep neural networks as scientific models. T*rends in Cognitive Sciences, 23*(4), 305–317. https://doi.org/10.1016/j.tics.2019.01.009

Cirino, P. T., Child, A. E., & Macdonald, K. T. (2018). Longitudinal predictors of the overlap between reading and math skills. *Contemporary Educational Psychology, 54*, 99– 111. https://doi.org/10.1016/j.cedpsych.2018.06.002

Cirino, P. T., Fletcher, J. M., Ewing-Cobbs, L., Barnes, M. A., Fuchs, L. S. (2007). Cognitive arithmetic differences in learning difficulty groups and the role of behavioral inattention. *Learning Disabilities Research & Practice, 22*(1), 25–35. https://doi.org/10.1111/j.1540-5826.2007.00228.x

Clayton, F. J., West, G., Sears, C., Hulme, C., & Lervåg, A. (2020). A longitudinal study of early reading development: Letter-sound knowledge, phoneme awareness and RAN, but not letter-sound integration, predict variations in reading development. *Scientific Studies of Reading, 24* (2), 91–107. https://doi.org/10.1080/10888438.2019.1622546

Darney, D., Reinke, W. M., Herman, K. C., Stormont, M., & Ialongo, N. S. (2013). Children with co-occurring academic and behavior problems in first grade: Distal outcomes in twelfth grade. *Journal of School Psychology, 51*, 117-128. https://doi.org/10.1016/j.jsp.2012.09.005

De Smedt, B., & Boets, B. (2010). Phonological processing and arithmetic fact retrieval: Evidence from developmental dyslexia. *Neuropsychologia, 48*(14), 3973–3981. https://doi.org/10.1016/j.neuropsychologia.2010.10 .018

De Smedt, B., Noël, M. P., Gilmore, C., & Ansari, D. (2013). How do symbolic and non-symbolic numerical magnitude processing skills relate to individual differences in children's mathematical skills? A review of evidence from brain and behavior. *Trends in Neuroscience and Education, 2*, 48–55. https://doi.org/10.1016/j.tine.2013.06.001

Denckla, M. B., & Rudel, R. G. (1976). Rapid "automatized" naming (R.A.N): Dyslexia differentiated from other learning disabilities. *Neuropsychologia, 14*(4), 471–479. https://doi.org/10.1016/0028-3932(76)90075-0

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

Donker, M., Kroesbergen, E., Slot, E., Van Viersen, S., & De Bree, E. (2016). Alphanumeric and non-alphanumeric Rapid Automatized Naming in children with reading and/or spelling difficulties and mathematical difficulties. *Learning and Individual Differences, 47*, 80-87. https://doi.org/10.1016/j.lindif.2015.12.011

Duncan, G. J., Dowsett, C. J., Claessens, A., Magnuson, K., Huston, A. C., Klebanov, P., ... & Japel, C. (2007). School readiness and later achievement. *Developmental Psychology, 43*(6), 1428. https://doi.org/10.1037/0012-1649.43.6.1428

Dunn, L. M., & Dunn, L. M. (1981). *Peabody picture vocabulary test—Revised.* American Guidance Service.

Dunst, C. J., Hamby, D. W., Wilkie, H., & Dunst, K. S. (2017). Meta-analysis of the relationship between home and family experiences and young children's early numeracy learning. In Phillipson, S., Gervasoni, A., Sullivan, P. (Eds) *Engaging Families as Children's First Mathematics Educators (pp. 105-125).* Springer, Singapore. https://doi.org/10.1007/978-981-10-2553-2_7

Durstewitz, D., Koppe, G., & Meyer-Lindenberg, A. (2019). Deep neural networks in psychiatry. *Molecular Psychiatry, 24*(11), 1583–1598. https://doi.org/10.1038/s41380-019-0365-9

Eccles, J. S., & Wigfield, A. (2002). Motivational beliefs, values, and goals. *Annual Review of Psychology, 53*(1), 109-132. https://doi.org/10.1146/annurev.psych.53.100901.135153

Eccles, J. S., Wigfield, A., & Schiefele, U. (1998). Motivation to succeed. In W. Damon & N. Eisenberg (Ed.), *Handbook of child psychology: Social, emotional, and personality development* (pp. 1017–1095). John Wiley & Sons, Inc.

Elliot, A. J., McGregor, H. A., & Gable, S. (1999). Achievement goals, study strategies, and exam performance: a mediational analysis. *Journal of Educational Psychology, 91*(3), 549. https://doi.org/10.1037/0022-0663.91.3.549

Fisher, P. H., Dobbs-Oates, J., Doctoroff, G. L., & Arnold, D. H. (2012). Early math interest and the development of math skills. *Journal of Educational Psychology, 104*(3), 673. https://doi.org/10.1037/a0027756

Fitzpatrick, C., & Pagani, L. S. (2013). Task-oriented kindergarten behavior pays off in later childhood. *Journal of Developmental & Behavioral Pediatrics, 34*(2), 94-101. https://doi.org/10.1097/DBP.0b013e31827a3779

Francis, D. J., Fletcher, J. M., Stuebing, K. K., Lyon, G. R., Shaywitz, B. A., & Shaywitz, S. E. (2005). Psychometric approaches to the identification of LD: IQ and achievement scores are not sufficient. *Journal of Learning Disabilities*, *38*(2), 98-108. https://doi.org/10.1177/00222194050380020101

Fuchs, L. S., Compton, D. L., Fuchs, D., Paulsen, K., Bryant, J. D., & Hamlett, C. L. (2005). The prevention, identification, and cognitive determinants of math difficulty. *Journal of Educational Psychology*, *97*(3), 493. https://doi.org/10.1037/0022-0663.97.3.493

Garon-Carrier, G., Boivin, M., Lemelin, J. P., Kovas, Y., Parent, S., Séguin, J. R., ... & Dionne, G. (2018). Early developmental trajectories of number knowledge and math achievement from 4 to 10 years: Low-persistent profile and early-life predictors. *Journal of School Psychology*, *68*, 84-98. https://doi.org/10.1016/j.jsp.2018.02.004

Geary, D. C. (1993). Mathematical disabilities: cognitive, neuropsychological, and genetic components. *Psychological Bulletin*, *114*(2), 345. https://doi.org/10.1037/0033-2909.114.2.345

Geary, D. C. (2011a). Consequences, characteristics, and causes of mathematical learning disabilities and persistent low achievement in mathematics. *Journal of developmental and behavioral pediatrics: JDBP*, *32*(3), 250. https://doi.org/10.1097/DBP.0b013e318209edef

Geary, D. C. (2011b). Cognitive predictors of achievement growth in mathematics: a 5-year longitudinal study. *Developmental Psychology*, *47*(6), 1539. https://doi.org/10.1037/a0025510

Geary, D. C., Bailey, D. H., Littlefield, A., Wood, P., Hoard, M. K., & Nugent, L. (2009). First-grade predictors of mathematical learning disability: A latent class trajectory analysis. *Cognitive Development*, *24*, 411–429. https://doi.org/10.1016/j.cogdev.2009.10.001

Geary, D. C., Hoard, M. K., Byrd-Craven, J., Nugent, L., Numtee, C. (2007). Cognitive mechanisms underlying achievement deficits in children with mathematical learning disability. *Child Development, 78*(4), 1343–1359. https://doi.org/10.1111/j.1467-8624.2007.01069.x

Geary, D. C., Nicholas, A., Li, Y., & Sun, J. (2017). Developmental change in the influence of domain-general abilities and domain-specific knowledge on mathematics achievement: An eight-year longitudinal study. *Journal of Educational Psychology, 109*(5), 680. https://doi.org/10.1037/edu0000159

Geary, D. C., van Marle, K., Chu, F., Rouder, J., Hoard, M. K., & Nugent, L. (2018). Early conceptual understanding of cardinality predicts superior school-entry number system knowledge. *Psychological Science, 29,* 191-205. https://doi.org/10.1177/0956797617729817

Georgiou, G. K., Tziraki, N., Manolitsis, G., & Fella, A. (2013). Is rapid automatized naming related to reading and mathematics for the same reason (s)? A follow-up study from kindergarten to Grade 1. *Journal of Experimental Child Psychology, 115*(3), 481-496. https://doi.org/10.1016/j.jecp.2013.01.004

Gold, A. B., Ewing-Cobbs, L., Cirino, P., Fuchs, L. S., Stuebing, K. K., & Fletcher, J. M. (2013). Cognitive and behavioral attention in children with math difficulties. *Child Neuropsychology, 19*(4), 420-437. https://doi.org/10.1080/09297049.2012.690371

Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep learning. MIT Press. http://www.deeplearningbook.org

Graves, A., Mohamed, A. R., & Hinton, G. (2013). Speech recognition with deep recurrent neural networks. In 2013 *IEEE International Conference on Acoustics, Speech and Signal Processing* (p. 6645–6649). https://doi.org/10.1109/ICASSP.2013.6638947

Hakkarainen, A. M., Holopainen, L. K., & Savolainen, H. K. (2015). A five-year follow-up on the role of educational support in preventing dropout from upper secondary education in Finland. *Journal of Learning Disabilities, 48*(4), 408-421. https://doi.org/10.1177/0022219413507

Hassinger-Das, B., Jordan, N. C., Glutting, J., Irwin, C., & Dyson, N. (2014). Domain-general mediators of the relation between kindergarten number sense and first-grade mathematics achievement. *Journal of Experimental Child Psychology, 118*, 78-92. https://doi.org/10.1016/j.jecp.2013.09.008

Hawes, Z., & Ansari, D. (2020). What explains the relationship between spatial and mathematical skills? A review of evidence from brain and behavior. *Psychonomic Bulletin & Review, 27*(3), 465-482. https://doi.org/10.3758/s13423-019-01694-7

Heinsfeld, A. S., Franco, A. R., Craddock, R. C., Buchweitz, A., & Meneguzzi, F. (2018). Identification of autism spectrum disorder using deep learning and the ABIDE dataset. *NeuroImage: Clinical, 17*, 16–23. https://doi.org/10.1016/j.nicl.2017.08.017

Hidi, S., & Renninger, K. A. (2006). The four-phase model of interest development. *Educational Psychologist, 41*(2), 111–127. https://doi.org/10.1207/s15326985ep4102_4

Hinton, G. E. (2007). Learning multiple layers of representation. *Trends in Cognitive Sciences, 11*(10), 428–434. https://doi.org/10.1016/j.tics.2007.09.004

Hirvonen, R., Tolvanen, A., Aunola, K., & Nurmi, J. E. (2012). The developmental dynamics of task-avoidant behavior and math performance in kindergarten and elementary school. *Learning and Individual Differences, 22*(6), 715-723. https://doi.org/10.1016/j.lindif.2012.05.014

Huijsmans, M. D. E., Kleemans, T., van der Ven, S. H. G., & Kroesbergen, E. H. (2020). The relevance of subtyping children with mathematical learning disabilities. *Research in Developmental Disabilities, 104*, 103704. https://doi.org/10.1016/j. ridd.2020.103704

Hutchison, J. E., Lyons, I. M., & Ansari, D. (2019). More similar than different: Gender differences in children's basic numerical skills are the exception not the rule. *Child Development, 90*(1), e66-e79. https://doi.org/10.1111/cdev.13044

Joyner, R. E., & Wagner, R. K. (2020). Co-occurrence of reading disabilities and math disabilities: a meta-analysis. *Scientific Studies of Reading, 24*(1), 14-22. https://doi.org/10.1080/10888438.2019.1593420

Khanolainen, D., Psyridou, M., Silinskas, G., Lerkkanen, M. K., Niemi, P., Poikkeus, A. M., & Torppa, M. (2020). Longitudinal effects of the home learning environment and parental difficulties on reading and math development across grades 1–9. *Frontiers in Psychology, 11*, 577981. https://doi.org/10.3389/fpsyg.2020.577981

Kleemans, T., Peeters, M., Segers, E., & Verhoeven, L. (2012). Child and home predictors of early numeracy skills in kindergarten. *Early Childhood Research Quarterly, 27*(3), 471-477. https://doi.org/10.1016/j.ecresq.2011.12.004

Kleemans, T., Segers, E., & Verhoeven, L. (2018). Individual differences in basic arithmetic skills in children with and without developmental language disorder: Role of home numeracy experiences. *Early Childhood Research Quarterly, 43*, 62–72. https://doi.org/10.1016/j.ecresq.2018.01.005

Knafo, A., & Jaffee, S. R. (2013). Gene–environment correlation in developmental psychopathology. *Development and Psychopathology, 25*(1), 1-6. https://doi.org/10.1017/S0954579412000855

Koponen, T., Aunola, K., Ahonen, T., & Nurmi, J. E. (2007). Cognitive predictors of single-digit and procedural calculation skills and their covariation with reading skill. *Journal* *of Experimental Child Psychology*, *97*(3), 220–241. https://doi.org/10.1016/j.jecp.2007.03.001

Koponen, T., Aro, M. Poikkeus, A-M., Niemi, P., Lerkkanen, M-K., Ahonen, T., Nurmi, J-E (2018). Comorbidity of difficulties in fluency of reading and arithmetic skills: Stability by grades. *Exceptional Children*, *81*(3), 298–311. https://doi.org/10.1177/0014402918756269

Koponen, T., Aunola, K., & Nurmi, J. E. (2019). Verbal counting skill predicts later math performance and difficulties in middle school. *Contemporary Educational Psychology*, *59*, 101803. https://doi.org/10.1016/j.cedpsych.2019.101803

Koponen, T., Georgiou, G., Salmi, P., Leskinen, M., & Aro, M. (2017). A meta-analysis of the relation between RAN and mathematics. *Journal of Educational Psychology*, *109*(7), 977. https://doi.org/10.1037/edu0000182

Kotsiantis, S. B., Zaharakis, I., & Pintelas, P. (2007). Supervised machine learning: A review of classification techniques. *Emerging Artificial Intelligence Applications in Computer engineering*, *160*(1), 3-24.

Kroesbergen, E. H., Huijsmans, M. D., & Friso-van den Bos, I. (2022). A Meta-Analysis on the Differences in Mathematical and Cognitive Skills Between Individuals With and Without Mathematical Learning Disabilities. *Review of Educational Research*, https://doi.org/10.3102/00346543221132773

Kuang, D., & He, L. (2014, November). Classification on ADHD with deep learning. In 2014 *International Conference on Cloud Computing and Big Data* (pp. 27–32). IEEE.

Kuhl, U., Sobotta, S., Legascreen Consortium, & Skeide, M. A. (2021). Mathematical learning deficits originate in early childhood from atypical development of a frontoparietal brain network. *PLoS Biology*, *19*(9), e3001407.

Landerl, K., Fussenegger, B., Moll, K., & Willburger, E. (2009). Dyslexia and dyscalculia: Two learning disorders with different cognitive profiles. *Journal of Experimental Child Psychology, 103*(3), 309–324. https://doi.org/10.1016/j.jecp .2009.03.006

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature, 521*(7553), 436–444. https://doi.org/10.1038/nature14539

LeFevre, J. A., Fast, L., Skwarchuk, S. L., Smith-Chant, B. L., Bisanz, J., Kamawar, D., & Penner-Wilger, M. (2010). Pathways to mathematics: Longitudinal predictors of performance. *Child Development, 81*(6), 1753-1767. https://doi.org/10.1111/j.1467-8624.2010.01508.x

Lehrl, S., Ebert, S., Blaurock, S., Rossbach, H. G., & Weinert, S. (2020). Long-term and domain-specific relations between the early years home learning environment and students' academic outcomes in secondary school. *School Effectiveness and School Improvement, 31*(1), 102-124. https://doi.org/10.1080/09243453.2019.1618346

Lerkkanen, M.-K., Poikkeus, A.-M., & Ketonen, R. (2006). *ARMI—Luku- ja kirjoitustaidon arviointimateriaali 1: Luokalle [ARMI–Reading and writing assessment material for 1st grade]*. Helsinki: WSOY.

Little, R. J. A. (1988). A test of missing completely at random for multivariate data with missing values. *Journal of the American Statistical Association, 83*(404), 1198–1202. https://doi.org/10.1080/01621459.1988.10478722

Little, C. W., Lonigan, C. J., & Phillips, B. M. (2021). Differential patterns of growth in reading and math skills during elementary school. *Journal of Educational Psychology, 113*(3), 462–476. https://doi.org/10.1037/edu0000635

Liu, S., Cheng, C., Wu, P., Zhang, L., Wang, Z., Wei, W., Chen, Y., & Zhao, J. (2022). Phonological processing, visuospatial skills, and pattern understanding in Chinese developmental dyscalculia. *Journal of Learning Disabilities, 55*(6), 499–512. https://doi.org/10.1177/00222194211063650

Lu, D., Popuri, K., Ding, G. W., Balachandar, R., & Beg, M. F. (2018). Multimodal and multiscale deep neural networks for the early diagnosis of Alzheimer's disease using structural MR and FDG-PET images. *Scientific Reports, 8*(1), 1–13. https://doi.org/10.1038/s41598-018-22871-z

Lundetræ, K., Gabrielsen, E., & Mykletun, R. (2010). Do basic skills predict youth unemployment (16-to 24-year-olds) also when controlled for accomplished upper-secondary school? A cross-country comparison. *Journal of Education and Work, 23*(3), 233-254. https://doi.org/10.1080/13639081003745439

Magnuson, K., Duncan, G. J., Lee, K. T., & Metzger, M. W. (2016). Early school adjustment and educational attainment. *American Educational Research Journal, 53*(4), 1198-1228. https://doi.org/10.3102/0002831216634658

Maloney, E. A., Ramirez, G., Gunderson, E. A., Levine, S. C., & Beilock, S. L. (2015). Intergenerational effects of parents' math anxiety on children's math achievement and anxiety. *Psychological Science, 26*(9), 1480-1488.

Mamoshina, P., Vieira, A., Putin, E., & Zhavoronkov, A. (2016). Applications of deep learning in biomedicine. *Molecular Pharmaceutics, 13*(5), 1445–1454. https://doi.org/10.1021/acs.molpharmaceut.5b00982

Mandrekar, J. N. (2010). Receiver operating characteristic curve in diagnostic test assessment. *Journal of Thoracic Oncology, 5*(9), 1315–1316. https://doi.org/10.1097/JTO.0b013e3181ec173d

Manolitsis, G., Georgiou, G. K., & Tziraki, N. (2013). Examining the effects of home literacy and numeracy environment on early reading and math acquisition. *Early Childhood Research Quarterly, 28*(4), 692-703. https://doi.org/10.1016/j.ecresq.2013.05.004

Marzocchi, G. M., Lucangeli, D., De Meo, T., Fini, F., Cornoldi, C. (2002). The disturbing effect of irrelevant information on arithmetic problem solving in inattentive children. *Developmental Neuropsychology*, *21*, 73-92. https://doi.org/10.1207/S15326942DN2101_4

Mazzocco, M. M. M., & Myers, G. F. (2003). Complexities in identifying and defining mathematics learning disability in the primary school-age years. *Annals of Dyslexia, 53*, 218–253. https://doi.org/10.1007/s11881-003-0011-7

Merrell, C., & Tymms, P. B. (2001). Inattention, hyperactivity and impulsiveness: Their impact on academic achievement and progress. *British Journal of Educational Psychology*, *71*, 43-56. https://doi.org/10.1348/000709901158389

Missall, K., Hojnoski, R. L., Caskie, G. I., & Repasky, P. (2015). Home numeracy environments of preschoolers: Examining relations among mathematical activities, parent mathematical beliefs, and early mathematical skills. *Early Education and Development*, *26*(3), 356-376. https://doi.org/10.1080/10409289.2015.968243

Mix, K. S., & Cheng, Y. L. (2012). The relation between space and math: Developmental and educational implications. *Advances in Child Development and Behavior*, *42*, 197-243. https://doi.org/10.1016/B978-0-12-394388-0.00006-X

Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., ... & Hassabis, D. (2015). Human-level control through deep reinforcement learning. *Nature*, *518*(7540), 529–533. https://doi.org/10.1038/nature14236

Moll, K., Göbel, S. M., Gooch, D., Landerl, K., Snowling, M. J. (2014a). Cognitive risk factors for specific learning disorder processing speed, temporal processing, and working memory. *Journal of Learning Disabilities*, *49*(3), 272–281. https://doi.org/10.1177/0022219414547221

Moll, K., Kunze, S., Neuhoff, N., Bruder, J., & Schulte-Körne, G. (2014b). Specific learning disorder: Prevalence and gender differences. *PLoS one, 9*(7), e103537. https://doi.org/10.1371/journal.pone.0103537

Morgan, P. L., Farkas, G., & Wu, Q. (2009). Five-year growth trajectories of kindergarten children with learning difficulties in mathematics. *Journal of Learning Disabilities, 42*(4), 306-321. https://doi.org/10.1177/0022219408331037

Mulcahy, C., Maccini, P., Wright, K., & Miller, J. (2014). An examination of intervention research with secondary students with EBD in light of common core state standards for mathematics. *Behavioral Disorders, 39*, 146-164. https://doi.org/10.1177/019874291303900304

Mulcahy, C., Krezmien, M., & Travers, J. (2016). Improving mathematics performance among secondary students with EBD: A methodological review. *Remedial and Special Education, 37*, 113-128. https://doi.org/10.1177/0741932515579275

Murphy, P. K., & Alexander, P. A. (2000). A motivated exploration of motivation terminology. *Contemporary Educational Psychology, 25*(1), 3-53. https://doi.org/10.1006/ceps.1999.1019

Napoli, A. R., & Purpura, D. J. (2018). The home literacy and numeracy environment in preschool: Cross-domain relations of parent–child practices and child outcomes. *Journal of Experimental Child Psychology, 166,* 581-603. https://doi.org/10.1016/j.jecp.2017.10.002

Nelson, G., & Powell, S. R. (2018). A systematic review of longitudinal studies of mathematics difficulty. *Journal of Learning Disabilities, 51*(6), 523-539. https://doi.org/10.1177/0022219417714773

Niklas, F., & Schneider, W. (2014). Casting the die before the die is cast: The importance of the home numeracy environment for preschool children. *European Journal of Psychology of Education, 29*(3), 327-345. https://doi.org/10.1007/s10212-013-0201-6

Olsen, C. R., Mentz, R. J., Anstrom, K. J., Page, D., & Patel, P. A. (2020). Clinical applications of machine learning in the diagnosis, classification, and prediction of heart failure. *American Heart Journal, 229*, 1-17. https://doi.org/10.1016/j.ahj.2020.07.009

Onatsu-Arvilommi, T., & Nurmi, J. E. (2000). The role of task-avoidant and task-focused behaviors in the development of reading and mathematical skills during the first school year: A cross-lagged longitudinal study. *Journal of Educational Psychology*, *92*(3), 478. https://doi.org/10.1037/0022-0663.92.3.478

Parhiala, P., Torppa, M., Vasalampi, K., Eklund, K., Poikkeus, A. M., & Aro, T. (2018). Profiles of school motivation and emotional well-being among adolescents: Associations with math and reading performance. *Learning and Individual Differences, 61*, 196-204. https://doi.org/10.1016/j.lindif.2017.12.003

Peng, P., Lin, X., Ünal, Z. E., Lee, K., Namkung, J., Chow, J., & Sales, A. (2020). Examining the mutual relations between language and mathematics: A meta-analysis. *Psychological Bulletin, 146*(7), 595. https://doi.org/10.1037/bul0000231

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

Pennington, B. F. (2006). From single to multiple deficit models of developmental disorders. *Cognition, 101*(2), 385–413. https://doi.org/10.1016/j.cognition.2006.04.008

Psyridou, M., Tolvanen, A., de Jong, P. F., Lerkkanen, M.-K., Poikkeus, A.-M., & Torppa, M. (2021). Developmental profiles of reading fluency and reading comprehension from grades 1 to 9 and their early identification. *Developmental Psychology*, *57*(11), 1840. https://doi.org/10.1037/dev0000976

Psyridou, M., Tolvanen, A., Lerkkanen, M. K., Poikkeus, A. M., & Torppa, M. (2020). Longitudinal stability of reading difficulties: Examining the effects of measurement error, cut-offs, and buffer zones in identification. *Frontiers in Psychology, 2841*. https://doi.org/10.3389/fpsyg.2019.02841

Psyridou, M., Tolvanen, A., Patel. P., Khanolainen. D., Lerkkanen, M.-K., Poikkeus, A.-M., Torppa, M. (2023a). Reading difficulties identification: A comparison of neural networks, linear, and mixture models. *Scientific Studies of Reading, 27*(1), 39-66 https://doi.org/10.1080/10888438.2022.2095281

Psyridou, M., Torppa, M., Tolvanen, A., Poikkeus, A., Lerkkanen, M., & Koponen, T. (2023b). Developmental profiles of arithmetic skills from Grades 1 to 9 and their early identification. *Developmental Psychology*. https://doi.org/10.1037/dev0001622

Pulkkinen, J., Eklund, K., Koponen, T., Heikkilä, R., Georgiou, G., Salminen, J., ... & Aro, M. (2022). Cognitive skills, self-beliefs and task interest in children with low reading and/or arithmetic fluency. *Learning and Individual Differences*, *97*, 102160. https://doi.org/10.1016/j.lindif.2022.102160

Räsänen, P., & Aunola, K. (2007). *The basic arithmetic test. Jyväskylä,* Finland: Niilo Mäki Institute and University of Jyväskylä.

Rubinsten, O. & Henik, A. (2009).Developmental Dyscalculia: heterogeneity might not mean different mechanisms, *Trends in Cognitive Sciences, 13*(2), 92-99. https://doi.org/10.1016/j.tics.2008.11.002.

Schneider, M., Beeres, K., Coban, L., Merz, S., Susan Schmidt, 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*(3), e12372. https://doi.org/10.1111/desc.12372

Sénéchal, M., Lefevre, J. A., Thomas, E. M., & Daley, K. E. (1998). Differential effects of home literacy experiences on the development of oral and written language. *Reading Research Quarterly, 33*(1), 96–116. https://doi.org/10.1598/RRQ.33.1.5

Shalev, R. S., Manor, O., & Gross-Tsur, V. (2005). Developmental dyscalculia: A prospective six-year follow-up. *Developmental Medicine and Child Neurology, 47*, 121–125. https://doi.org/10.1017/S001216220500

Silinskas, G., Torppa, M., Lerkkanen, M. K., & Nurmi, J. E. (2020b). The home literacy model in a highly transparent orthography. *School Effectiveness and School Improvement, 31*(1), 80–101. https://doi.org/10.1080/09243453.2019.1642213

Silver, D., Huang, A., Maddison, C. J., Guez, A., Sifre, L., Van Den Driessche, G., ... & Hassabis, D. (2016). Mastering the game of Go with deep neural networks and tree search. *Nature, 529*(7587), 484–489. https://doi.org/10.1038/nature16961

Silver, A. M., & Libertus, M. E. (2022). Environmental influences on mathematics performance in early childhood. *Nature Reviews Psychology*, 1-12. https://doi.org/10.1038/s44159-022-00061-z

Sims, D. M., Purpura, D. J., & Lonigan, C. J. (2016). The relation between inattentive and hyperactive/impulsive behaviors and early mathematics skills. *Journal of Attention Disorders, 20*(8), 704-714. https://doi.org/10.1177/1087054712464390

Singh, A., Thakur, N., & Sharma, A. (2016). A review of supervised machine learning algorithms. In *2016 3rd International Conference on Computing for Sustainable Global Development* (pp. 1310-1315). IEEE.

Skwarchuk, S. L., Sowinski, C., & LeFevre, J. A. (2014). Formal and informal home learning activities in relation to children’s early numeracy and literacy skills: The development of a home numeracy model. *Journal of Experimental Child Psychology*, *121*, 63-84. https://doi.org/10.1016/j.jecp.2013.11.006

Soares, N., Evans, T., & Patel, D. R. (2018). Specific learning disability in mathematics: a comprehensive review. *Translational Pediatrics,* *7*(1), 48. https://doi.org/10.21037/tp.2017.08.03

Statistics Finland. (2007). *Statistical databases.* Retrieved March 1, 2019, from www.stat.fi/tup/tilastotietokannat/index_en.html

Statistics Finland. (2021). Support for learning. Retrieved May 1, 2023, from https://stat.fi/en/publication/cktyiw7xc2e8w0c586gqxm122

Susperreguy, M. I., Di Lonardo Burr, S., Xu, C., Douglas, H., & LeFevre, J. A. (2020). Children's home numeracy environment predicts growth of their early mathematical skills in kindergarten. *Child Development,* *91*(5), 1663-1680. https://doi.org/10.1111/cdev.13353

Taylor, J., Roehrig, A. D., Hensler, B. S., Connor, C. M., & Schatschneider, C. (2010). Teacher quality moderates the genetic effects on early reading. *Science, 328*(5977), 512-514. https://doi.org/10.1126/science.1186149

Trout, A., Nordness, P., Pierce, C., & Epstein, M. (2003). Research on the academic status of children with emotional and behavioral disorders: A review of the literature from 1961 to 2000. *Journal of Emotional and Behavioral Disorder,* *11*, 198-210. https://doi.org/10.1177/10634266030110040201

Urban, C. J., & Gates, K. M. (2021, April 1). Deep Learning: A Primer for Psychologists. *Psychological Methods.* http://dx.doi.org/10.1037/met0000374

Vanbinst, K., Ghesquière, P., & De Smedt, B. (2014). Arithmetic strategy development and its domain-specific and domain-general cognitive correlates: A longitudinal study in children with persistent mathematical learning difficulties. *Research in Developmental Disabilities, 35*(11), 3001-3013. https://doi.org/10.1016/j.ridd.2014.06.023

Vauras, M., Mäki, H., Dufva, M., & Hämäläinen, S. (1995). *Diagnostiset testit 2. Kuullun ja luetun ymmärtäminen.* [Diagnostic tests 2: Listening and reading comprehension]. Oppimistutkimuksen keskus, Turun yliopisto.

Vieira, S., Pinaya, W. H., & Mechelli, A. (2017). Using deep learning to investigate the neuroimaging correlates of psychiatric and neurological disorders: Methods and applications. *Neuroscience & Biobehavioral Reviews, 74*, 58–75. https://doi.org/10.1016/j.neubiorev.2017.01.002

Viljaranta, J., Lerkkanen, M.-K., Poikkeus, A.-M., Aunola, K., & Nurmi, J.-E. (2009). Cross-lagged relations between task motivation and performance in arithmetic and literacy in kindergarten. *Learning and Instruction, 19*(4), 335-344. https://doi.org/10.1016/j.learninstruc.2008.06.011

Wasserstein, R. L., & Lazar, N. A. (2016). The ASA statement on p-values: context, process, and purpose. *The American Statistician, 70*(2), 129-133. https://doi.org/10.1080/00031305.2016.1154108

Wei, T., Liu, X., & Barnard-Brak, L. (2015). Gender differences in mathematics and reading trajectories among children from kindergarten to eighth grade. *Research in Education, 93*(1), 77-89. https://doi.org/10.7227/RIE.0015

Wigfield, A., & Cambria, J. (2010). Students' achievement values, goal orientations, and interest: Definitions, development, and relations to achievement outcomes. *Developmental Review, 30*(1), 1-35. http://dx.doi.org/10.1016/j.dr.2009.12.001

Willcutt, E. G., Petrill, S. A., Wu, S., Boada, R., Defries, J. C., Olson,R. K., & Pennington, B. F. (2013). Comorbidity between reading disability and math disability: Concurrent psychopathology, functional impairment, and neuropsychological functioning. *Journal of Learning Disabilities*, *46*, 500 –516. http://dx.doi.org/10.1177/0022219413477476

Woodcock, R. W., & Johnson, M. B. (1977). *Woodcock–Johnson psycho-educational battery*. Itasca, IL: Riverside.

Yang, X., Yan, M., Ruan, Y., Ku, S. Y. Y., Lo, J. C. M., Peng, P., & McBride, C. (2022). Relations among phonological processing skills and mathematics in children: A meta-analysis. *Journal of Educational Psychology*, *114*(2), 289–307. https://doi.org/10.1037/edu0000710

Zhang, X., & Lin, D. (2015). Pathways to arithmetic: The role of visual-spatial and language skills in written arithmetic, arithmetic word problems, and nonsymbolic arithmetic. *Contemporary Educational Psychology*, *41*, 188-197. https://doi.org/10.1016/j.cedpsych.2015.01.005

Zhang, X., Nurmi, J.-E., Kiuru, N., Lerkkanen, M.-K., & Aunola, K. (2011). A teacher-report measure of children’s task-avoidant behavior: A validation study of the behavioral strategy rating scale. *Learning and Individual Differences*, *21*(6), 690–698. https://doi.org/10.1016/j.lindif.2011.09.007

Zhang, X., Räsänen, P., Koponen, T. Aunola, K., Lerkkanen, M., & Nurmi, J. (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

Zippert, E. L., & Rittle-Johnson, B. (2020). The home math environment: More than numeracy. *Early Childhood Research Quarterly*, *50*(3), 4-15. https://doi.org/10.1016/j.ecresq.2018.07.009

Table 1

Arithmetic and multiplication measures and kindergarten-age factors used in the study

Measure	Task	Assessment Year	Description	Scoring	Reliability
Arithmetic	Basic Arithmetic Test (Räsänen & Aunola, 2007)	Spring 2013	28 items in total with increased difficulty across the test. In this time-limited, group-administered paper and-pencil test, the participant is required to complete as many arithmetic operations as possible within a 3-min time limit. Performance in the test requires both accuracy and speed (automatization of basic calculation routines). In Grade 6, the test included addition, subtraction, or their combination (e.g., 84+13-27=), multiplication (e.g., 12 × 28 =), division (e.g., 57 ÷ 5 =), or calculation with decimals (e.g., 106,2-30,04 =).	A score of 1 was given for every correct answer. Max 28.	-
Multiplication		Spring 2013	120 items in total across the test. In this group-administered paper and-pencil test, the participant is required to complete as many single-digit multiplication operations as possible (e.g., 2*7=, 2*2=, 5*8=). Children were given 2 minutes to complete as many items as possible.	A score of 1 was given for every correct answer. Max 120.	-
Kindergarten Arithmetic	Arithmetic (Basic Arithmetic Test (Räsänen & Aunola, 2007))	Spring 2007	28 items in total with increased difficulty across the test. In this time-limited, group-administered paper and-pencil test, the participant is required to complete as many arithmetic operations as possible within a 3-min time limit. Performance in the test requires both accuracy and speed (automatization	A score of 1 was given for every correct answer. Max 28.	-

			of basic calculation routines). In kindergarten, the test included addition (e.g., 2+1=) and subtraction (e.g., 11-2=).		
	Arithmetic strategies	Spring 2007	Children's use of arithmetic strategies was examined with a 6-item questionnaire from the testers (Strategy 1: provides answer immediately "from memory", Strategy 2: nods or uses gaze, Strategy 3: uses fingers, Strategy 4: makes lines or dots and counts them, Strategy 5: counts aloud, Strategy 6: any other strategy). Each item was answered using a 3-point Likert scale (0= Does not use this strategy at all, 1= Sometimes uses this strategy, 1-2 times, 2= Very often uses this strategy, 3 times or more)	One item per strategy. Each item was examined individually. Very few children used strategy 4 so it was removed from the present study.	-
Kindergarten Cognitive Skills	Counting (for similar tasks, see Koponen et al., 2007)	Fall 2006, Spring 2007	There were four tasks in which children were asked to count aloud forward (from 1 to 31and from 6 to 13) and backward (from 12 to 7 and from 23 to 1).	Scored using a 3-point scale: 2 = no errors, 1 = one small error, 0 = two or more errors. Max 8.	Cronbach's alpha= .48 (fall), .64 (spring)/ Revelle's omega= .52 (fall), .87 (spring)
	Number concepts	Spring 2007	A combined measure of ordinal and cardinal number knowledge as well as knowledge of basic mathematical concepts. The child saw a number and was asked to draw a corresponding number of balls or, alternatively, was shown balls and was asked to select the corresponding number from five choices. The child was asked to draw balls according the instructions "as many," "one more," and "one less" and mark the "first," "fourth," and "seventh" ball.	A score of 1 was given for every correct answer. Max 9.	Cronbach's alpha=.72
	Spatial relations (Woodcock and Johnson (1977))	Spring 2007	The test requires the child to identify the subset of pieces needed to form a complete shape with multiple point scored items (i.e., "Two of these	A score of 1 was given for every correct answer. Max 31.	-

		pieces () go together to make this (). Tell me which two pieces.”). It involves complicated, multistep manipulations of spatial information (i.e., detecting multiple spatial forms or shapes, rotating or manipulating them in the imagination, and matching). Children were given 3 minutes to complete as many items of the 31 items as possible.		
RAN (Denckla & Rudel, 1976)	Spring 2007	The children were asked to name as fast as possible a series of five pictures of objects arranged in semi random order in five rows of 10. There was a practice trial before the test to ensure the child’s familiarity with names of the objects.	Total matrix completion time in seconds.	-
Initial phoneme identification (ARMI; Lerkkanen et al., 2006)	Fall 2006, Spring 2007	The children were shown one set a time 10 sets of 4 pictures, each picture depicting an object. Students were first asked to name aloud the objects and then identify the object with the same initial phoneme as the one spoken aloud by the assessor. All sounds were single phonemes.	A score of 1 was given for every correctly selected object. Max 10.	Cronbach’s alpha= .75 (fall), .71 (spring)
Letter knowledge (ARMI; Lerkkanen et al., 2006)	Fall 2006, Spring 2007	The children were shown 29 uppercase letters arranged in random order across three rows. Students were asked to name them aloud. Either a phoneme or letter name was regarded as correct. The test was discontinued after 6 incorrect responses.	A score of 1 was given for every correct response. Max 29.	Cronbach’s alpha= .94 (fall), .93 (spring)
Receptive Vocabulary (PPVT-R, Form L; Dunn & Dunn, 1981)	Spring 2007	A 30-item version of the Peabody Picture Vocabulary Test-Revised. Students were required to select the picture, out of 4 options, that correctly depicts a spoken word.	A score of 1 was given for every correct response. Max 30.	Cronbach’s alpha= .60

	Reading Words (ARMI; Lerkkanen et al., 2006)	Fall 2006, Spring 2007	Students were administered a word list containing 6 words at the fall assessment and 10 words at the spring assessment. Students were asked to read aloud the words. At the fall assessment, there were 4 two-syllabic words, 1 three-syllabic word and 1 five-syllabic word. At the spring assessment, there were 7 two-syllabic words, 2 three-syllabic words, and 1 five-syllabic word.	A score of 1 was given for every correctly read word. Max 10.	Cronbach's alpha= .84 (fall), .85 (spring)
	Listening comprehension (Vauras et al., 1995)	Spring 2007	Groups of 6 students were read aloud a story (130 words), twice, and then asked six multiple-choice questions based on the story, one question at a time. In four of the questions there were three choices, and in two questions there were four choices Each question was accompanied by 3 or 4 pictures and student responded by marking the picture that correctly matched the story in their own test booklet.	2 points were given for every correct answer. Max 12.	Cronbach's alpha= .30/ Revelle's omega= .42
Motivational Measures	Interest in reading and math	Spring 2007	Children's interest was assessed with an individually administered interview addressing how much a child likes reading/math. Each question was answered using a 5-point Likert scale (1= Does not like at all, 2= Does not like very much, 3= In-between, 4= Likes quite a lot, 5= Likes very much).	One question for reading and one for math. Each item was examined individually. Max 5 for reading interest and 5 for math interest.	-
	Self-concept in reading and math	Spring 2007	Self-concept of ability was assessed with an individually administered interview addressing how good a child thinks he/she is in reading/math in comparison to other children	One question for reading and one for math. Each item was examined individually. Max. 10 for each skill.	-
	Task values of numeracy and literacy	Spring 2007	Task value measured children's task motivation in literacy and math activities in preschool. 6 items were measured for preschool activities (literacy -	Each item was examined individually.	-

	activities (see (Aunola et al., 2006)	"how much fun", letter tasks - "how much fun", numbers and counting - "how much fun", literacy - "how much you like", letter tasks - "how much you like", numbers and counting tasks, "how much you like") and 3 for at home activities (literacy, "how much you like", letter tasks - "how much you like", number and counting tasks - "how much you like"). Each question was answered using a 5-point Likert scale (1= Does not like at all, 2= Does not like very much, 3= In-between, 4= Likes quite a lot, 5= Likes very much)		
Behavioral Measures	Attention	Trained testers evaluated the behavior of each student in the class by rating them on 3 questions based on how the child typically behaved in classroom situations (e.g., Does the child get tired of tasks easily (works well in the beginning, but the ability to concentrate deteriorates significantly with the tasks)?). Ratings were done on a 7-point Likert scale (1 = Not at all this kind of behavior; 7 = Always / almost always this kind of behavior).	A sum score of the 3 items was calculated.	Cronbach's alpha= .86
	Hyperactivity	Trained testers evaluated the behavior of each student in the class by rating them on 3 questions based on how the child typically behaved in classroom situations (e.g., Does the child show difficulty sit still in a chair (squirms, swings legs, gets up from chair)?). Ratings were done on a 7-point Likert scale (1 = Not at all this kind of behavior; 7 = Always / almost always this kind of behavior).	A sum score of the 3 items was calculated.	Cronbach's alpha=.67
	Impulsiveness	Trained testers evaluated the behavior of each student in the class by rating them on 3 questions based on how the child typically behaved in classroom situations (e.g., Does the child confirm	A sum score of the 3 items was calculated.	Cronbach's alpha=.83

	the answer before the whole question is asked?). Ratings were done on a 7-point Likert scale (1 = Not at all this kind of behavior; 7 = Always / almost always this kind of behavior).		
Attention-orientation	Trained testers evaluated the behavior of each student in the class by rating them on 3 questions based on how the child typically behaved in classroom situations (e.g., Is the child rigid and inflexible in solving the task (easily sticks to an old solution model or previous type of task or gets stuck)?). Ratings were done on a 7-point Likert scale (1 = Not at all this kind of behavior; 7 = Always / almost always this kind of behavior).	A sum score of the 3 items was calculated.	Cronbach's alpha=.66
Planning	Trained testers evaluated the behavior of each student in the class by rating them on 3 questions based on how the child typically behaved in classroom situations (e.g., Is for the child difficult to perform tasks that require multi-step progress from one step to another (needs a lot of adult help in structuring and moving the task forward)?). Ratings were done on a 7-point Likert scale (1 = Not at all this kind of behavior; 7 = Always / almost always this kind of behavior).	A sum score of the 3 items was calculated.	Cronbach's alpha=.78
Disruptive behavior	Trained testers evaluated the behavior of each student in the class by rating them on 4 questions based on how the child typically behaved in classroom situations (e.g., Does the child speak even when told to be quiet). Ratings were done on a 7-point Likert scale (1 = Not at all this kind of behavior; 7 = Always / almost always this kind of behavior).	A sum score of the 4 items was calculated.	Cronbach's alpha=.71

	Task-avoidant behavior (Behavior Strategy Rating Scale; Zhang et al., 2011)		Kindergarten teachers evaluated the behavior of each student in the class by rating them on 5 questions based on how the child typically behaved in classroom situations (e.g., Does the child have a tendency to find something else to do instead of focusing on the task at hand?; Does the child show persistence even in the more difficult tasks?). Ratings were done on a 5-point Likert scale (1 = not at all; 5 = to a great extent). (Note: the items were transposed so that for all items higher values represent less task avoidant behavior)	A sum score of the 5 items was calculated.	Cronbach's alpha= .92
Home Learning Environment	Home learning environment (Sénéchal et al., 1998; see Silinskas et al., 2020b)	Spring 2007	5-item questionnaire about at-home learning activities answered by the mothers and fathers. It included 1-item regarding shared reading which was answered using a 5-point Likert scale (1 = less than once a week, 2 = 1–3 times a week, 3 = 4–6 times a week, 4 = once a day, 5 = more than once a day), 2-items regarding in-home teaching of literacy (teaching letters & teaching reading) which was also answered in a 5-point scale (1 = never/very seldom to 5 = very often/daily), and 2-items regarding in-home teaching of math (teaching arithmetic & teaching numeracy) which were also answered in a 5-point scale (1 = never/very seldom to 5 = very often/daily).	Each item was examined individually.	-
Parental Measures	Parental education	Spring 2007	Mothers and fathers were asked to indicate their own education level on a 7-point scale: 1 = no vocational education, 2 = vocational courses (4 months), 3 = vocational school degree, 4 = vocational college degree, 5 = polytechnic degree or bachelor's degree, 6 = master's degree, and 7 = licentiate or doctoral degree.	Answers were recoded using a 3-point scale: basic education, vocational education, and university education.	-

	Parental reading difficulties	Parents were asked to indicate on a 3-point scale whether they had clear difficulties, some difficulties, or no difficulties in reading.	A child was considered as having family risk if the mother or the father reported that she or he had experienced some or clear RD.	-
	Parental math difficulties	Parents were asked to indicate on a 3-point scale whether they had clear difficulties, some difficulties, or no difficulties in math.	A child was considered as having family risk if the mother or the father reported that she or he had experienced some or clear MD.	-

Note. Max = maximum; MD = math difficulties; RAN = rapid automatized naming; PPVT-R = Peabody Picture Vocabulary Test – Revised; RD = reading difficulties

Table 2

Descriptive statistics for the kindergarten-age factors and the arithmetic and multiplication skills in Grade 6

Measures	N	Mean	S.D.	Min.	Max.	Skewness	Kurtosis
Phonological awareness fall	1867	7.46	2.45	0	10	-.81	-.21
Letter knowledge fall	1867	16.95	9.01	0	29	-.25	-1.27
Reading words fall	1867	1.00	1.99	0	6	1.81	1.60
Counting fall	1866	4.43	2.83	0	8	-.21	-1.33
Vocabulary fall	1839	19.82	3.38	7	29	-.38	.31
Phonological awareness spring	1836	8.93	1.72	0	10	-1.99	4.11
Letter knowledge spring	1836	23.21	6.61	0	29	-1.34	1.02
Reading words spring	1823	4.03	4.29	0	10	.44	-1.61
Counting spring	1836	6.06	2.20	0	8	-1.10	.25
Spatial relations spring	1830	14.26	2.36	0	24	-.38	1.41
Number concepts spring	1834	8.28	1.36	1	9	-2.35	6.05
Listening comprehension spring	1832	7.71	2.34	0	12	-.31	-.13
RAN spring	1835	173.71	17.78	34	210	-1.72	6.69
Interest in literacy	1837	3.62	1.43	1	5	-.68	-.87
Interest in math	1836	3.84	1.35	1	5	-.90	-.46
Task values: content areas in preschool, literacy, "how much fun"	1836	3.88	1.34	1	5	-1.02	-.18

Task values: content areas in preschool, letter tasks, "how much fun"	1836	3.67	1.33	1	5	-.73	-.62
Task values: content areas in preschool, numbers and counting, "how much fun"	1834	3.84	1.33	1	5	-.89	-.41
Task values: content areas in preschool, literacy, "how much you like"	1834	3.99	1.24	1	5	-1.09	.16
Task values: content areas in preschool, letter tasks, "how much you like"	1835	3.77	1.28	1	5	-.79	-.43
Task values: content areas in preschool, numbers and counting tasks, "how much you like"	1834	3.85	1.29	1	5	-.89	-.33
Task values: content areas in the home, literacy, "how much you like"	1835	4.03	1.33	1	5	-1.21	.20
Task values: content areas in the home, letter tasks, "how much you like"	1835	3.75	1.36	1	5	-.82	-.55
Task values: content areas in the home, number and counting tasks, "how much you like"	1835	3.85	1.32	1	5	-.89	-.43
Self-concept in literacy	1835	3.37	2.34	1	10	1.08	.80
Self-concept in numbers and counting	1835	2.99	2.35	1	10	1.30	1.19
Arithmetic kindergarten spring	1803	2.95	2.21	0	17	1.15	2.45
Strategy 1, provides answer immediately "from memory"	1666	.82	.63	0	2	.16	-.59

Strategy 2, nods or uses gaze	1640	.56	.76	0	2	.92	-.68
Strategy 3, uses fingers	1686	1.27	.87	0	2	-.55	-1.46
Strategy 5, counts aloud	1578	.79	.88	0	2	.43	-1.58
Strategy 6, any other strategy	535	.52	.83	0	2	1.10	-.64
Attention	1832	5.12	3.26	2	21	2.12	4.84
Hyperactivity	1832	5.34	3.08	2	21	1.71	2.95
Impulsiveness	1831	4.66	2.70	3	20	2.30	6.03
Attention-orientation	1832	4.05	2.00	2	20	2.85	10.68
Planning	1831	4.61	2.67	2	21	2.30	6.25
Disruptive behavior	1831	4.69	1.94	3	22	4.32	22.50
Task avoidance	1814	18.37	5.17	5	25	-.55	-.55
Parental reading difficulties	1505	.33	.47	0	1	.74	-1.46
Parental math difficulties	1501	.34	.47	0	1	.68	-1.54
Maternal education	2087	1	3	2.32	.60	-.27	-.64
Paternal education	2068	1	3	2.25	.61	-.20	-.58
Shared reading	1603	4.45	2.12	1	10	.40	-.63
Teaching letters at home	1606	4.71	1.80	1	10	.18	-.53
Teaching reading at home	1607	3.76	1.73	1	10	.34	-.41
Teaching numeracy at home	1607	4.86	1.86	1	10	.05	-.68
Teaching arithmetic at home	1607	3.85	1.76	1	10	.30	-.47
Gender	1884	1.52	.50	1	2	-.10	-1.99
Arithmetic Grade 6	1817	16.29	3.71	1	26	-.30	.26
Multiplication Grade 6	1817	40.47	16.89	4	117	.88	.88
Math (sum score) Grade 6	1817	0	1.78	-6.11	6.61	.32	.15

Note. Min. = Minimum; Max. = Maximum; RAN = rapid automatized naming

Table 3

Corresponding sensitivity, specificity, and precision estimates for different cut-off values for the prediction of math difficulties

Cut-off	Sensitivity	1-specificity	Specificity	Precision	F1-score
10%	1	.889	.111	.118	.211
15%	1	.832	.168	.125	.222
20%	1	.749	.251	.137	.241
25%	.983	.723	.277	.139	.244

Note. For the testing data out of the 545 cases, there were 58 positive cases.

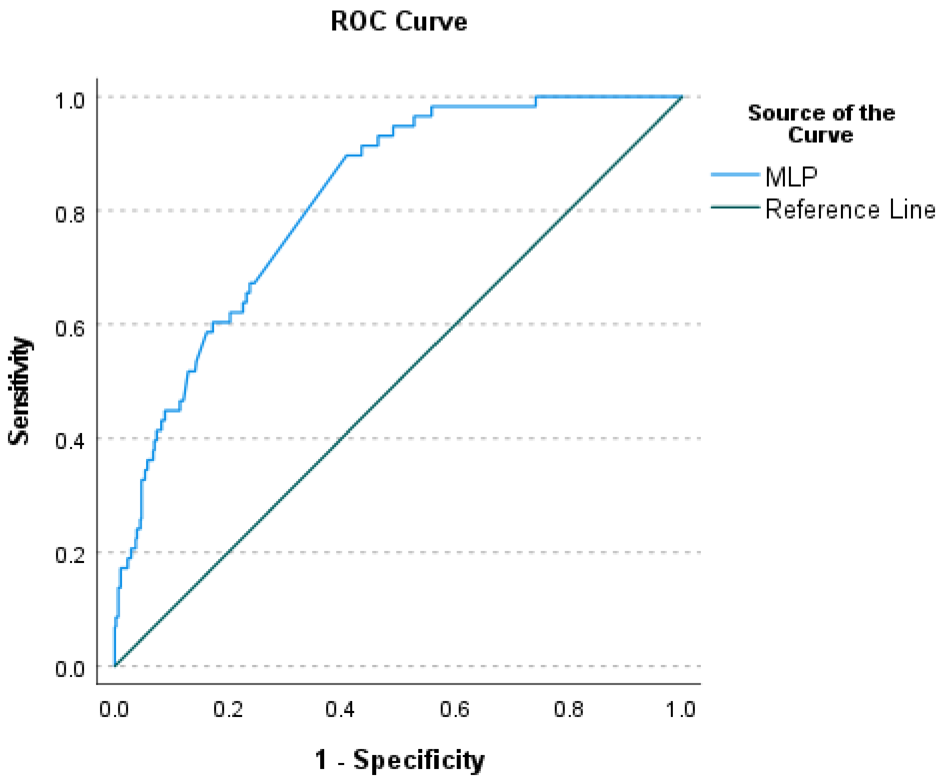
Table 4

Specificity and precision values corresponding to specific sensitivity values for the prediction of math difficulties

Sensitivity	1-specificity	Specificity	Precision
.638	.226 - .232	.768 - .774	.247 - .252
.672	.238 - .246	.754 - .762	.245 - .252
.897	.409 - .435	.565 - .591	.197 - .207

Figure 1

ROC curve for the prediction of math difficulties for the testing models

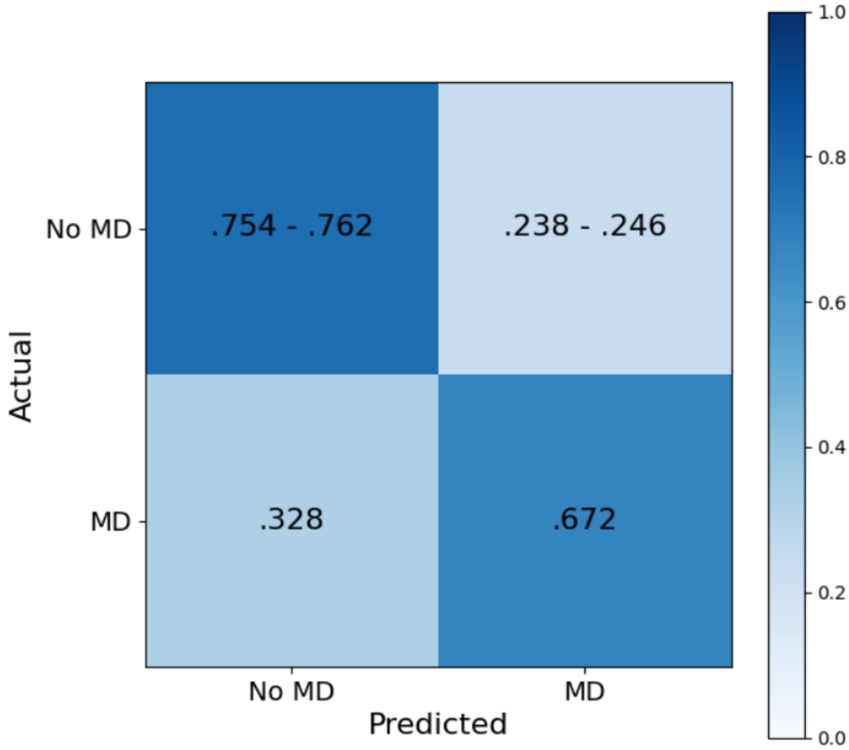


Note. ROC curve for the prediction of math difficulties. Sensitivity (true positive rate) is the rate of the model classifying an individual with math difficulties as having math difficulties, whereas 1-specificity (false positive rate) is the rate of the model classifying an individual with math difficulties as not having math difficulties. AUC denotes the area under the ROC curve (blue line).

ROC = receiver operating charactirstic curve; AUC = area under the curve; MLP = multilayer perceptron network.

Figure 2

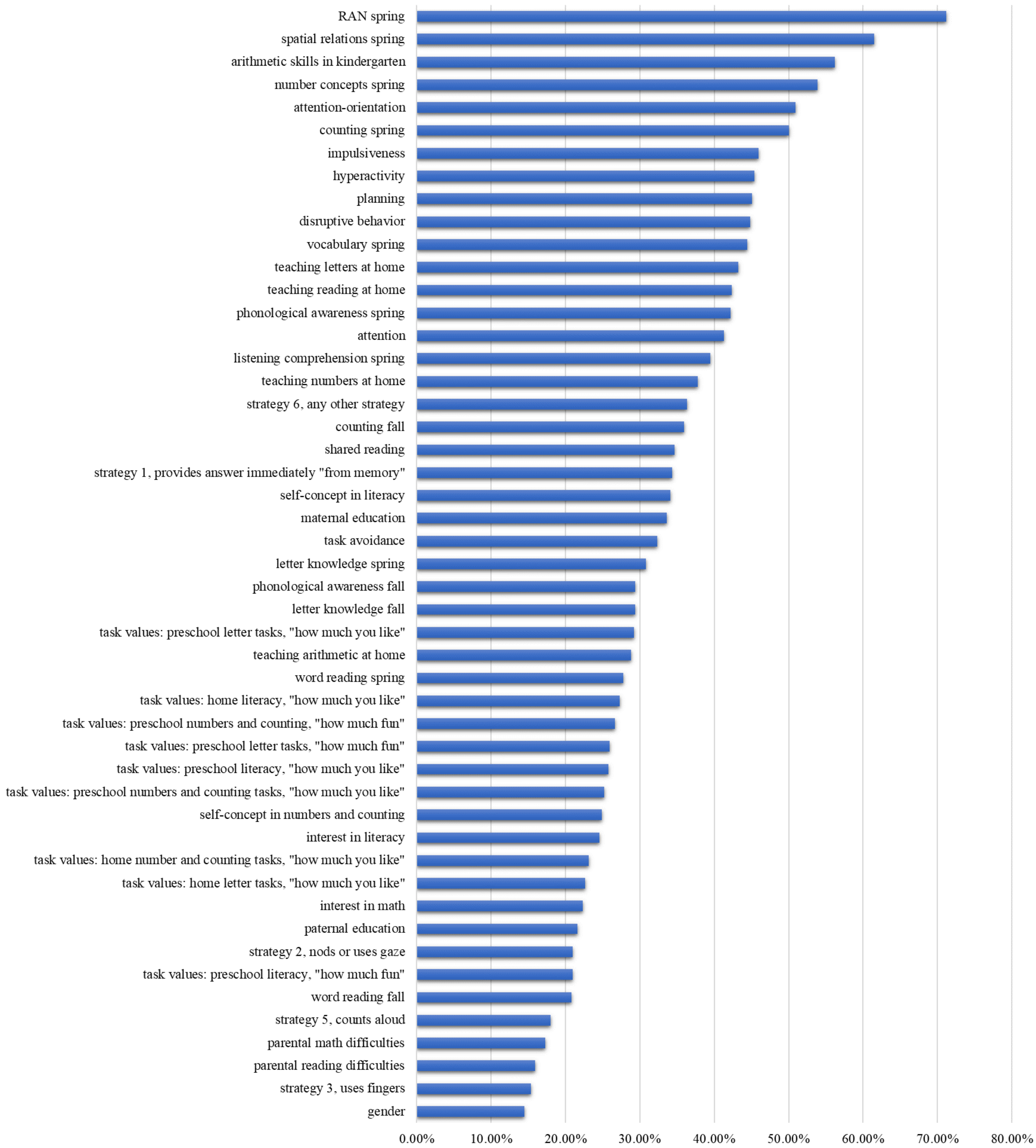
Confusion matrix for predicting math difficulties in the testing models when the sensitivity is .672



Note. MD = math difficulties.

Figure 3

Mean of normalized importance for the kindergarten-age factors for the prediction of math difficulties



Appendix A

Correlations among the kindergarten-age factors and the arithmetic and multiplication skills in Grade 6

	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.
1. Phonological awareness fall	1	.61***	.56***	.43***	.30***	.56***	.54***	.62***	.35***	.25***	.27***	.19***	.29***
2. Letter knowledge fall	.58***	1	.63***	.59***	.32***	.55***	.81***	.71***	.48***	.27***	.29***	.19***	.37***
3. Reading words fall	.46***	.55***	1	.41***	.24***	.37***	.50***	.62***	.31***	.21***	.21***	.17***	.28***
4. Counting fall	.42***	.60***	.38***	1	.26***	.39***	.56***	.50***	.66***	.30***	.29***	.15***	.30***
5. Vocabulary fall	.32***	.34***	.22***	.28***	1	.26***	.27***	.25***	.24***	.27***	.25***	.30***	.17***
6. Phonological awareness spring	.55***	.52***	.28***	.38***	.29***	1	.59***	.63***	.35***	.24***	.24***	.14***	.31***
7. Letter knowledge spring	.52***	.79***	.37***	.55***	.29***	.62***	1	.72***	.51***	.27***	.27***	.15***	.37***
8. Reading words spring	.57***	.67***	.62***	.48***	.24***	.50***	.60***	1	.41***	.27***	.27***	.17***	.38***
9. Counting spring	.37***	.50***	.29***	.69***	.27***	.40***	.55***	.39***	1	.31***	.31***	.11***	.30***
10. Spatial relations spring	.26***	.28***	.22***	.30***	.29***	.27***	.28***	.27***	.34***	1	.24***	.20***	.28***
11. Number concepts spring	.31***	.30***	.18***	.31***	.30***	.32***	.34***	.25***	.38***	.28***	1	.19***	.18***
12. Listening comprehension spring	.18***	.18**	.16***	.15***	.29***	.16***	.16***	.17***	.14***	.20***	.21***	1	.14***
13. RAN spring	.26***	.34***	.25***	.30***	.19***	.28***	.32***	.33***	.31***	.29***	.24***	.11***	1
14. Interest in literacy	.04	.10***	.07***	.05*	.06*	.10***	.12***	.10***	.10***	.08***	.10***	.04	.06*
15. Interest in math	.04	.08***	.06**	.11***	.03	.05	.11***	.06**	.14***	.10***	.06**	.03	.07***

16. Task values: content areas in preschool, literacy, "how much fun"	-.02	.04	.01	.02	-.01	.00	.08***	.00	.06*	.07**	.03	.00	.08***
17. Task values: content areas in preschool, letter tasks, "how much fun"	.07***	.11***	.07**	.06**	.09***	.07***	.14***	.10***	.10***	.08***	.08***	.04	.06**
18. Task values: content areas in preschool, numbers and counting, "how much fun"	.04	.11***	.06**	.15***	.06**	.05*	.12***	.06**	.17***	.12***	.11***	.06*	.11***
19. Task values: content areas in preschool, literacy, "how much you like"	.00	.04	.03	.00	.02	.04	.05*	.02	.05*	.06**	.08***	.04	.08***
20. Task values: content areas in preschool, letter tasks, "how much you like"	.06**	.12***	.08***	.07***	.09***	.08***	.13***	.08***	.12***	.07***	.10***	.02	.10***
21. Task values: content areas in preschool, numbers and counting tasks, "how much you like"	.04	.14***	.07***	.14***	.07***	.05*	.13***	.08***	.18***	.07***	.13***	.09***	.11***

22. Task values: content areas in the home, literacy, "how much you like"	-.05*	.01	.00	-.05*	-.02	.00	.01	-.04	-.03	.01	.01	-.02	.00
23. Task values: content areas in the home, letter tasks, "how much you like"	.01	.04	.03	-.02	.01	.04	.06**	.04	.04	.04	.03	.00	.03
24. Task values: content areas in the home, number and counting tasks, "how much you like"	.02	.07***	.04	.07***	.05*	.02	.08***	.03	.10***	.04	.07***	.05*	.05*
25. Self-concept in literacy	-.12***	-.18***	-.14***	-.15***	-.11***	-.14***	-.20***	-.17***	-.20***	-.09***	-.13***	-.07***	-.18***
26. Self-concept in numbers and counting	-.08***	-.15***	-.10***	-.20***	-.05*	-.10***	-.16***	-.11***	-.22***	-.07***	-.10***	-.06***	-.09***
27. Arithmetic kindergarten spring	.29***	.42***	.33***	.50***	.24***	.28***	.41***	.39***	.46***	.32***	.28***	.15***	.27***
28. Strategy 1, provides answer immediately "from memory"	.22***	.28***	.23***	.38***	.14***	.20***	.27***	.24***	.34***	.24***	.19***	.12***	.19***
29. Strategy 2, nods or uses gaze	.02	.03	.02	.03	.01	.04	.04	.04	.05	.03	.08***	.00	.04
30. Strategy 3, uses fingers	.01	.01	-.02	-.05*	.03	.02	.02	.01	-.01	.01	.03	.01	.05*
31. Strategy 5, counts aloud	-.05*	-.04	-.07***	-.07***	.04	-.06*	-.07***	-.05*	-.06*	-.03	-.01	.00	.00

32. Strategy 6, any other strategy	.04	-.03	.00	-.03	-.04	-.06	-.03	.02	-.02	.05	.01	-.01	.02
33. Attention	-.24***	-.27***	-.16***	-.21***	-.20***	-.29***	-.29***	-.23***	-.23***	-.29***	-.25***	-.10***	-.25***
34. Hyperactivity	-.15***	-.15***	-.09***	-.11***	-.10***	-.17***	-.20***	-.14***	-.12***	-.16***	-.17***	-.04	-.19***
35. Impulsiveness	-.15***	-.13***	-.08***	-.09***	-.14***	-.17***	-.17***	-.14***	-.13***	-.19***	-.17***	-.05*	-.13***
36. Attention-orientation	-.20***	-.22***	-.12***	-.22***	-.25***	-.30***	-.25***	-.19***	-.27***	-.23***	-.21***	-.08***	-.25***
37. Planning	-.27***	-.29***	-.17***	-.26***	-.24***	-.31***	-.33***	-.26***	-.31***	-.29***	-.28***	-.10***	-.28***
38. Disruptive behavior	-.12***	-.14***	-.06**	-.10***	-.08***	-.12***	-.15***	-.10***	-.12***	-.16***	-.15***	-.05*	-.19***
39. Task avoidance	.31***	.35***	.24***	.30***	.22***	.32***	.36***	.33***	.32***	.27***	.30***	.17***	.26***
40. Parental reading difficulties	-.11***	-.16***	-.10***	-.12***	-.03	-.16***	-.17***	-.18***	-.14***	-.07**	-.07**	-.06*	-.10***
41. Parental math difficulties	-.09***	-.13***	-.07**	-.11***	-.09***	-.10***	-.12***	-.10***	-.12***	-.13***	-.12***	-.08***	-.15***
42. Maternal education	.18***	.18***	.16***	.13***	.17***	.13***	.13***	.14***	.12***	.12***	.13***	.11***	.10***
43. Paternal education	.14***	.16***	.11***	.11***	.15***	.11***	.12***	.11***	.08***	.14***	.11***	.10***	.11***
44. Shared reading	.14***	.20***	.14***	.08***	.24***	.10***	.15***	.16***	.06**	.03	.07**	.16***	.06**
45. Teaching letters at home	.07**	.15***	.09***	.08***	.11***	.09***	.15***	.10***	.09***	.02	.04	.05	.04
46. Teaching reading at home	.13***	.23***	.15***	.15***	.13***	.18***	.25***	.23***	.13***	.05*	.06*	.07**	.08***
47. Teaching numeracy at home	-.01	.07***	.05	.06*	.08***	.01	.07**	.01	.08***	.00	.02	.03	.00
48. Teaching arithmetic at home	.02	.14***	.08***	.13***	.13***	.08***	.14***	.06*	.14***	.04	.07***	.07***	.02
49. Gender	-.10***	-.13***	-.11***	.13***	.00	-.17***	-.15***	-.15***	.09***	-.02	-.07***	-.07**	-.05*

50. Arithmetic Grade 6	.21***	.29***	.23***	.43***	.14***	.18***	.28***	.26***	.40***	.29***	.22***	.12***	.29***
51. Multiplication Grade 6	.11***	.25***	.21***	.30***	.00	.14***	.24***	.21***	.29***	.20***	.12***	.01	.28***
52. Math (sum score) Grade 6	.18***	.30***	.25***	.41***	.08**	.18***	.29***	.26***	.39***	.27***	.19***	.07**	.32***

(correlation table continue)

	14.	15.	16.	17.	18.	19.	20.	21.	22.	23.	24.	25.	26.
1. Phonological awareness fall	.02	.02	-.04	.05*	.03	-.03	.05*	.02	-.08***	-.01	.00	-.08***	-.03
2. Letter knowledge fall	.08***	.06**	.02	.09***	.10***	.01	.10***	.11***	-.02	.02	.05*	-.15***	-.09***
3. Reading words fall	.05*	.05*	-.03	.05*	.04	.01	.06**	.06**	-.02	.02	.03	-.13***	-.06**
4. Counting fall	.03	.09***	.00	.05*	.13***	-.02	.06**	.13***	-.07***	-.03	.05*	-.11***	-.15***
5. Vocabulary fall	.03	.01	-.03	.08***	.05*	-.01	.07***	.05*	-.05	.00	.04	-.08***	-.01
6. Phonological awareness spring	.07***	.04	-.02	.05**	.04	-.01	.06**	.03	-.02	.02	-.02	-.10***	-.04
7. Letter knowledge spring	.09***	.09***	.04	.10***	.10***	.01	.11***	.11***	-.02	.04	.05	-.15***	-.12***
8. Reading words spring	.11***	.08***	.01	.11***	.07***	.02	.10***	.10***	-.04	.04	.03	-.17***	-.09***
9. Counting spring	.07***	.13***	.04	.07***	.14***	.01	.11***	.15***	-.05*	.02	.08***	-.13***	-.15***
10. Spatial relations spring	.06**	.08***	.03	.05*	.10***	.02	.03	.05*	-.02	.01	.02	-.06**	-.02
11. Number concepts spring	.07***	.05*	.00	.07***	.10***	.04	.06**	.11***	-.02	.00	.06**	-.05*	-.04
12. Listening comprehension spring	.03	.02	-.01	.02	.03	.02	.00	.06**	-.04	-.01	.03	-.05*	-.03

13. RAN spring	.07***	.07***	.07***	.06**	.08***	.08***	.10***	.10***	.00	.03	.04	-.16***	-.06**
14. Interest in literacy	1	.42***	.54***	.43***	.32***	.47***	.37***	.31***	.20***	.28***	.27***	-.19***	-.08***
15. Interest in math	.42***	1	.46***	.38***	.43***	.38***	.38***	.41***	.20***	.29***	.35***	-.16***	-.13***
16. Task values: content areas in preschool, literacy, "how much fun"	.58***	.46***	1	.43***	.35***	.54***	.37***	.35***	.25***	.27***	.25***	-.19***	-.07***
17. Task values: content areas in preschool, letter tasks, "how much fun"	.45***	.40***	.48***	1	.39***	.38***	.55***	.42***	.23***	.41***	.36***	-.21***	-.11***
18. Task values: content areas in preschool, numbers and counting, "how much fun"	.33***	.43***	.37***	.40***	1	.30***	.41***	.55***	.18***	.31***	.44***	-.16***	-.19***
19. Task values: content areas in preschool, literacy, "how much you like"	.49***	.40***	.56***	.39***	.32***	1	.36***	.31***	.25***	.28***	.25***	-.15***	-0.04
20. Task values: content areas in preschool, letter tasks, "how much you like"	.38***	.38***	.38***	.56***	.42***	.38***	1	.38***	.24***	.41***	.35***	-.16***	-.13***

21. Task values: content areas in preschool, numbers and counting tasks, "how much you like"	.34***	.42***	.36***	.43***	.54***	.31***	.38***	1	.19***	.33***	.48***	-.19***	-.20***
22. Task values: content areas in the home, literacy, "how much you like"	.21***	.19***	.25***	.23***	.17***	.26***	.22***	.19***	1	.38***	.31***	-.11***	-.07***
23. Task values: content areas in the home, letter tasks, "how much you like"	.28***	.28***	.26***	.40***	.31***	.28***	.40***	.33***	.40***	1	.44***	-.16***	-.08***
24. Task values: content areas in the home, number and counting tasks, "how much you like"	.28***	.35***	.25***	.36***	.42***	.25***	.36***	.47***	.34***	.45***	1	-.15***	-.20***
25. Self-concept in literacy	-.22***	-.17***	-.20***	-.24***	-.16***	-.17***	-.18***	-.22***	-.09***	-.17***	-.15***	1	.29***
26. Self-concept in numbers and counting	-.09***	-.15***	-.08***	-.12***	-.21***	-.05*	-.13***	-.22***	-.07***	-.08***	-.20***	.28***	1
27. Arithmetic kindergarten spring	.08***	.11***	.06**	.10***	.16***	.03	.07***	.16***	-.03	.01	.09***	-.19***	-.20***

28. Strategy 1, provides answer immediately "from memory"	.00	-.01	-.05	.02	.07***	-.01	.00	.08***	-.06*	-.07**	.02	-.13***	-.14***
29. Strategy 2, nods or uses gaze	.02	.01	.03	.02	.00	.00	.03	.00	.04	.02	.00	.03	-.01
30. Strategy 3, uses fingers	.03	.02	.02	.02	.03	.00	.01	.04	-.01	.06*	.04	-.01	.03
31. Strategy 5, counts aloud	.03	.02	.02	.04	.01	.01	.05	-.02	.01	.03	.03	.00	.00
32. Strategy 6, any other strategy	.01	-.05	.00	.06	.04	.07	-.01	.05	.07	.10*	.04	-.02	.00
33. Attention	-.12***	-.08***	-.10***	-.17***	-.10***	-.18***	-.16***	-.10***	-.05*	-.10***	-.10***	.13***	.02
34. Hyperactivity	-.10***	-.06*	-.08***	-.12***	-.04	-.14***	-.13***	-.07***	-.05*	-.07***	-.07***	.08***	.00
35. Impulsiveness	-.08***	-.03	-.04	-.10***	-.04	-.10***	-.11***	-.02	-.03	-.07**	-.02	.04	-.04
36. Attention orientation	-.09***	-.08***	-.04	-.08***	-.08***	-.11***	-.11***	-.07***	-.04	-.08***	-.10***	.12***	.06**
37. Planning	-.11***	-.06***	-.07***	-.14***	-.11***	-.15***	-.14***	-.10***	-.02	-.09***	-.08***	.15***	.04
38. Disruptive behavior	-.09***	-.03	-.10***	-.13***	-.09***	-.15***	-.15***	-.07**	-.03	-.07***	-.06**	.09***	-.04
39. Task avoidance	.16***	.12***	.12***	.16***	.12***	.11***	.13***	.12***	.03	.07***	.07***	-.10***	-.08***
40. Parental reading difficulties	.00	-.01	.00	-.03	.00	.00	-.01	.00	.05*	.03	.05	.02	.01

41. Parental math difficulties	-.04	-.01	-.01	-.02	-.03	.03	-.02	-.05*	.04	.02	.01	.00	.00
42. Maternal education	-.05*	-.01	-.04	-.04	.00	-.05*	-.04	-.03	-.02	-.01	-.03	-.04	-.03
43. Paternal education	-.03	-.01	-.02	-.04	.03	-.05*	-.04	-.03	-.03	-.02	-.02	.00	.00
44. Shared reading	.02	.00	.00	.05	.03	.00	.03	.03	.00	.04	.07***	-.04	.00
45. Teaching letters at home	.05	.00	.03	.06**	.05	-.01	.05*	.09***	.00	.02	.09***	-.03	-.03
46. Teaching reading at home	.04	-.01	.02	.07**	.05*	-.01	.05*	.09***	-.03	.03	.09***	-.07***	-.05*
47. Teaching numeracy at home	.04	.02	.05*	.07**	.06*	.02	.05*	.08***	-.01	.01	.08***	-.05	-.05
48. Teaching arithmetic at home	.06**	.03	.05	.08***	.09***	.02	.05*	.10***	-.03	.01	.09***	-.08***	-.10***
49. Gender	-.14***	-.04	-.10***	-.11***	.02	-.12***	-.12***	-.03	-.14***	-.11***	-0.02	.03	-.10***
50. Arithmetic Grade 6	.10***	.15***	.09***	.10***	.17***	.08**	.05	.12***	.00	.02	.09***	-.17***	-.17***
51. Multiplication Grade 6	.07**	.14***	.09***	.11***	.11***	.08***	.10***	.10***	.03	.05	.08***	-.11***	-.13***
52. Math (sum score) Grade 6	.09***	.16***	.10***	.12***	.15***	.09**	.09**	.12***	.02	.04	.10***	-.16***	-.17***

(correlation table continue)

	27.	28.	29.	30.	31.	32.	33.	34.	35.	36.	37.	38.	39.
1. Phonological awareness fall	.31***	.23***	.01	.00	-.05*	.02	-.21***	-.13***	-.13***	-.17***	-.23***	-.11***	.30***
2. Letter knowledge fall	.45***	.29***	.02	.00	-.05	-.02	-.26***	-.15***	-.11***	-.21***	-.27***	-.13***	.34***
3. Reading words fall	.32***	.23***	.01	-.02	-.07**	.00	-.19***	-.12***	-.08***	-.16***	-.22***	-.08***	.26***
4. Counting fall	.53***	.39***	.02	-.06**	-.07**	-.03	-.21***	-.10***	-.07***	-.19***	-.25***	-.11***	.30***
5. Vocabulary fall	.24***	.13***	.00	.02	.05*	-.03	-.16***	-.07***	-.10***	-.20***	-.16***	-.06**	.20***
6. Phonological awareness spring	.31***	.21***	.03	.00	-.08***	-.05	-.24***	-.15***	-.14***	-.21***	-.24***	-.13***	.30***
7. Letter knowledge spring	.45***	.28***	.02	.01	-.07***	.01	-.25***	-.16***	-.13***	-.22***	-.26***	-.13***	.31***
8. Reading words spring	.42***	.26***	.03	.02	-.05*	.02	-.25***	-.16***	-.15***	-.20***	-.28***	-.13***	.35***
9. Counting spring	.49***	.32***	.05	-.04	-.05*	-.01	-.21***	-.13***	-.11***	-.22***	-.25***	-.13***	.29***
10. Spatial relations spring	.33***	.24***	.02	.00	-.04	.04	-.23***	-.12***	-.15***	-.17***	-.22***	-.14***	.25***
11. Number concepts spring	.30***	.19***	.05*	.02	-.04	.03	-.20***	-.13***	-.14***	-.15***	-.20***	-.15***	.26***
12. Listening comprehension spring	.16***	.13***	.00	.01	.00	-.02	-.11***	-.04	-.04	-.09***	-.08***	-0.03	.16***

13. RAN spring	.30***	.19***	.05	.06*	-.02	.02	-.21***	-.17***	-.10***	-.21***	-.23***	-.13***	.26***
14. Interest in literacy	.07***	-.01	.01	.04	.03	.02	-.09***	-.08***	-.05*	-.08***	-.09***	-.07***	.14***
15. Interest in math	.10***	-.01	.00	.03	.02	-.06	-.08***	-.05*	.00	-.07***	-.06**	-.05	.11***
16. Task values: content areas in preschool, literacy, "how much fun"	.04	-.06*	.02	.03	.01	.02	-.08***	-.07***	-.04	-.05*	-.05*	-.09***	.10***
17. Task values: content areas in preschool, letter tasks, "how much fun"	.10***	.01	.00	.03	.04	.07	-.15***	-.11***	-.09***	-.06**	-.11***	-.11***	.14***
18. Task values: content areas in preschool, numbers and counting, "how much fun"	.17***	.07**	-.01	.03	.02	.05	-.09***	-.03	-.03	-.10***	-.10***	-.06**	.10***
19. Task values: content areas in preschool, literacy, "how much you like"	.02	-.03	-.03	.00	.00	.10*	-.11***	-.11***	-.07**	-.07***	-.09***	-.12***	.09***
20. Task values: content areas in preschool, letter tasks, "how much you like"	.06*	-.02	.01	.01	.05*	.02	-.13***	-.11***	-.09***	-.10***	-.11***	-.12***	.11***

21. Task values: content areas in preschool, numbers and counting tasks, "how much you like"	.16***	.09***	-.01	.04	-0.01	.04	-.10***	-.06**	-0.03	-.09***	-.11***	-.06*	.11***
22. Task values: content areas in the home, literacy, "how much you like"	-.06*	-.07**	.03	.00	.00	.09*	-.04	-.03	-.03	-.03	-.01	-.02	.00
23. Task values: content areas in the home, letter tasks, "how much you like"	.00	-.07***	.01	.05*	.03	.10*	-.07***	-.04	-.04	-.08***	-.08***	-.06***	.03
24. Task values: content areas in the home, number and counting tasks, "how much you like"	.07***	.01	.00	.04	.03	.04	-.08***	-.06**	-.02	-.11***	-.07***	-.02	.06**
25. Self-concept in literacy	-.15***	-.11***	.03	-.01	0	-.02	.09***	.05*	0	.08***	.10***	.05*	-.06**
26. Self-concept in numbers and counting	-.16***	-.12***	.02	.04	-.01	-.01	.00	-.03	-.05*	.05*	.01	-.08***	-.02
27. Arithmetic kindergarten spring	1	.45***	.07***	.07**	.02	.07	-.22***	-.09***	-.09***	-.22***	-.22***	-.11***	.30***

28. Strategy 1, provides answer immediately "from memory"	.46***	1	.01	-.21***	-.13***	-.07	-.14***	-.06**	-.04	-.10***	-.14***	-.07**	.15***
29. Strategy 2, nods or uses gaze	.07**	.01	1	-.08***	-.04	.01	.02	-.01	.01	.08***	.03	-.05	.07**
30. Strategy 3, uses fingers	.01	-.22***	-.09**	1	.23***	-.24***	-.01	.06*	.03	-.04	.01	.02	-.02
31. Strategy 5, counts aloud	.01	-.14***	-.03	.23***	1	-.31***	.10***	.11***	.05*	-.02	.08***	.08***	-.08**
32. Strategy 6, any other strategy	.06	-.07	.02	-.22***	-.30***	1	-.06	-.10*	-.07	.02	-.07	-.05	.03
33. Attention	-.20***	-.12***	-.01	.00	.08***	-.06	1	.63***	.59***	.48***	.63***	.49***	-.32***
34. Hyperactivity	-.11***	-.04	-.03	.05	.08***	-.09*	.71***	1	.54***	.28***	.50***	.46***	-.27***
35. Impulsiveness	-.09***	-.02	-.02	.04	.05*	-.07	.66***	.62***	1	.29***	.51***	.45***	-.23***
36. Attention orientation	-.21***	-.14***	.04	-.03	-.01	.04	.54***	.33***	.33***	1	.50***	.23***	-.23***
37. Planning	-.23***	-.14***	-.01	-.02	.05*	-.05	.75***	.59***	.63***	.64***	1	.45***	-.32***
38. Disruptive behavior	-.10***	-.05	-.04	.01	.06*	-.05	.64***	.58***	.53***	.30***	.54***	1	-.29***
39. Task avoidance	.28***	.15***	.08***	-.02	-.07**	.04	-.36***	-.30***	-.27***	-.23**	-.35***	-.27***	1
40. Parental reading difficulties	-.06*	-.05	-.02	.04	.07*	.05	.04	.01	.01	.03	.06*	.02	-.12***

41. Parental math difficulties	-.10***	-.05	-.05*	.00	.00	.05	.05*	.03	.02	.07**	.09***	.07**	-.15***
42. Maternal education	.08***	.08***	.07**	-.04	-.03	-.08	-.05*	-.03	-.05*	-.05*	-.09***	-.05	.11***
43. Paternal education	.09***	.05*	.01	-.02	.02	-.09*	-.08***	-.03	-.08***	-.09***	-.11***	-.05*	.10***
44. Shared reading	.03	.04	.00	.01	.00	-.03	-.05	-.01	-.05*	-.04	-.07**	-.03	.06**
45. Teaching letters at home	.01	.04	.03	.01	-.02	.05	-.06*	-.03	-.06**	-.06*	-.08***	-.04	.05
46. Teaching reading at home	.08***	.08***	.03	.00	-.03	.02	-.08***	-.03	-.07**	-.09***	-.10***	-.03	.10***
47. Teaching numeracy at home	.01	.03	.04	-.03	-.03	.05	-.02	.00	-.04	-.02	-.03	-.02	.02
48. Teaching arithmetic at home	.12***	.11***	.00	-.01	-.01	.06	-.04	.01	-.04	-.06**	-.09***	.01	.05*
49. Gender	.06*	.12***	-.01	-.09**	.03	-.03	.19***	.19***	.14***	.04	.13***	.11***	-.24***
50. Arithmetic Grade 6	.45***	.35***	.02	-.05	-.04	-.09	-.14***	-.07**	-.04	-.13***	-.14***	-.07**	.24***
51. Multiplication Grade 6	.35***	.24***	.03	-.04	-.05	-.04	-.16***	-.14***	-.08**	-.11***	-.15***	-.11***	.22***
52. Math (sum score) Grade 6	.45***	.33***	.03	-.05	-.05	-.07	-.17***	-.12***	-.06*	-.13***	-.16***	-.10***	.26***

(correlation table continue)

	40.	41	43.	44.	45.	46.	47.	48.	49.	50.	51.	52.	53
1. Phonological awareness fall	-.11***	-.10***	.19***	.14***	.15***	.06**	.13***	-.01	.02	-.11***	.22***	.12***	.19***
2. Letter knowledge fall	-.15***	-.13***	.19***	.17***	.21***	.15***	.23***	.07***	.14***	-.12***	.30***	.25***	.31***
3. Reading words fall	-.10***	-.07**	.18***	.13***	.15***	.08***	.15***	.03	.06**	-.13***	.25***	.19***	.25***
4. Counting fall	-.12***	-.11***	.13***	.11***	.08***	.08***	.15***	.06**	.13***	.13***	.43***	.30***	.41***
5. Vocabulary fall	-.02	-.09***	.18***	.15***	.25***	.10***	.12***	.08***	.12***	.01	.12***	-.02	.06*
6. Phonological awareness spring	-.14***	-.10***	.15***	.13***	.10***	.08***	.18***	.00	.07**	-.16***	.17***	.16***	.18***
7. Letter knowledge spring	-.17***	-.11***	.14***	.11***	.16***	.14***	.24***	.05*	.12***	-.12***	.30***	.28***	.33***
8. Reading words spring	-.17***	-.10***	.14***	.10***	.16***	.10***	.24***	.01	.07***	-.17***	.27***	.22***	.27***
9. Counting spring	-.12***	-.11***	.11***	.09***	.07***	.09***	.12***	.07**	.14***	.09***	.38***	.30***	.39***
10. Spatial relations spring	-.07***	-.14***	.13***	.15***	.04	.02	.05	-.01	.04	-.01	.28***	.18***	.26***
11. Number concepts spring	-.06*	-.12***	.12***	.10***	.08***	.06**	.07***	.05	.08***	-.08***	.19***	.13***	.18***
12. Listening comprehension spring	-.04	-.08***	.11***	.11***	.16***	.04	.06*	.03	.07**	-.07**	.11***	.01	.07*

13. RAN spring	-.10***	-.15***	.10***	.11***	.06**	.03	.06**	-.03	.01	-.08***	.31***	.28***	.33***
14. Interest in literacy	.00	-.03	-.06**	-.03	.01	.03	.02	.03	.05	-.14***	.08**	.07**	.08**
15. Interest in math	-.01	.00	-.03	-.03	-.01	.00	-.01	.02	.02	-.04	.13***	.14***	.15**
16. Task values: content areas in preschool, literacy, "how much fun"	.00	.00	-.06**	-.04	-.01	.01	.00	.04	.03	-.10***	.06*	.08***	.08**
17. Task values: content areas in preschool, letter tasks, "how much fun"	-.03	-.01	-.05	-.05*	.04	.06*	.07**	.06**	.08***	-.10***	.08**	.11***	.11***
18. Task values: content areas in preschool, numbers and counting, "how much fun"	.01	-.02	-.02	.02	.01	.04	.05	.05*	.08***	.03	.15***	.09***	.13***
19. Task values: content areas in preschool, literacy, "how much you like"	.00	.02	-.07**	-.07***	.00	-.02	-.02	.01	.01	-.13***	.04	.08**	.07*
20. Task values: content areas in preschool, letter tasks, "how much you like"	-.01	-.01	-.06*	-.05*	.02	.05	.04	.05	.05	-.11***	.03	.09***	.08*

21. Task values: content areas in preschool, numbers and counting tasks, "how much you like"	.00	-.04	-.04	-.04	.03	.07**	.08***	.07**	.09***	-.02	.10***	.10***	.10***
22. Task values: content areas in the home, literacy, "how much you like"	.05	.03	-.04	-.05*	-.01	-.01	-.03	-.01	-.03	-.12***	-.04	-.01	-.02
23. Task values: content areas in the home, letter tasks, "how much you like"	.04	.03	-.02	-.03	.03	.01	.03	.00	.00	-.09***	.00	.04	.02
24. Task values: content areas in the home, number and counting tasks, "how much you like"	.06*	.00	-.03	-.02	.05*	.07***	.07***	.07***	.08***	-.01	.06*	.06*	.07*
25. Self-concept in literacy	-.01	-.03	-.01	.02	.00	-.01	-.05*	-.02	-.05*	.02	-.14***	-.09***	-.13***
26. Self-concept in numbers and counting	-.01	-.02	.00	.03	.02	-.01	-.03	-.03	-.08***	-.12***	-.13***	-.09***	-.12***
27. Arithmetic kindergarten spring	-.08***	-.12***	.08***	.10***	.04	.02	.10***	.01	.12***	.04	.44***	.30***	.42***

28. Strategy 1, provides answer immediately "from memory"	-.05	-.05	.08***	.06*	.04	.05	.09***	.02	.11***	.12***	.34***	.24***	.32***
29. Strategy 2, nods or uses gaze	-.01	-.05	.06*	.01	.00	.03	.03	.03	.01	-.01	.03	.03	.04
30. Strategy 3, uses fingers	.03	.01	-.04	-.03	.00	.01	.00	-.02	.00	-.09***	-.03	-.02	-.02
31. Strategy 5, counts aloud	.07**	.00	-.02	.02	.00	-.02	-.03	-.03	-.01	.03	-.04	-.05	-.05
32. Strategy 6, any other strategy	.05	.04	-.08	-.09*	-.02	.06	.02	.06	.08	-.03	-.08	-.07	-.09
33. Attention	.03	.03	-.04	-.06**	-.06*	-.06**	-.08**	-.04	-.05*	.19***	-.12***	-.14***	-.14***
34. Hyperactivity	.00	.02	-.03	-.02	-.01	-.04	-.04	-.02	-.01	.18***	-.07**	-.14***	-.11***
35. Impulsiveness	.01	.01	-.02	-.04	-.04	-.04	-.05*	-.03	-.02	.14***	-.01	-.05	-.04
36. Attention orientation	.00	.08***	-.07**	-.08***	-.06*	-.06*	-.10***	-.02	-.05*	.06**	-.13***	-.09***	-.12***
37. Planning	.06*	.10***	-.11***	-.10***	-.08***	-.06**	-.09***	-.03	-.07***	.12***	-.13***	-.15***	-.15***
38. Disruptive behavior	.03	.03	-.04	-.05	.00	-.01	-.02	-.01	.02	.12***	-.07**	-.12***	-.11***
39. Task avoidance	-.12***	-.15***	.11***	.10***	.08***	.04	.10***	.01	.05*	-.23***	.23***	.23***	.26***
40. Parental reading difficulties	1	.39***	-.14***	-.14***	-.06*	-.01	-.03	.00	-.01	.06**	-.08**	-.07*	-.09**

41. Parental math difficulties	.39***	1	-.20***	-.18***	-.05	.01	.00	.00	-.01	.04	-.14***	-.06	-.10***
42. Maternal education	-.14***	-.20***	1	.59***	.25***	.06**	.09***	.06*	.08***	.04	.16***	.14***	.15***
43. Paternal education	-.14***	-.18***	.59***	1	.17***	.03	.06**	.02	.02	.04	.11***	.13***	.12***
44. Shared reading	-.06*	-.04	.25***	.17***	1	.51***	.46***	.48***	.45***	.00	.09***	.06	.09**
45. Teaching letters at home	-.01	.01	.06**	.03	.49***	1	.80***	.87***	.71***	.00	.04	.04	.05
46. Teaching reading at home	-.04	-.01	.09***	.06*	.44***	.80***	1	.73***	.71***	-.04	.08**	.07*	.08*
47. Teaching numeracy at home	.00	.00	.07**	.02	.46***	.86***	.72***	1	.75***	.07***	.05	.06	.06*
48. Teaching arithmetic at home	-.01	-.02	.08***	.02	.43***	.71***	.70***	.75***	1	.06**	.09***	.05	.08*
49. Gender	.06**	.04	.04	.05*	.01	.00	-.05	.07***	.07**	1	.08**	-.04	.02
50. Arithmetic Grade 6	-.07*	-.12***	.17***	.12***	.09***	.03	.07*	.04	.09***	.08***	1	.57***	.89***
51. Multiplication Grade 6	-.07*	-.07*	.13***	.13***	.06*	.02	.05	.05	.06	-.01	.58***	1	.87***
52. Math (sum score) Grade 6	-.08*	-.10***	.16***	.13***	.09*	.03	.07*	.05	.08*	.04	.88***	.89***	1

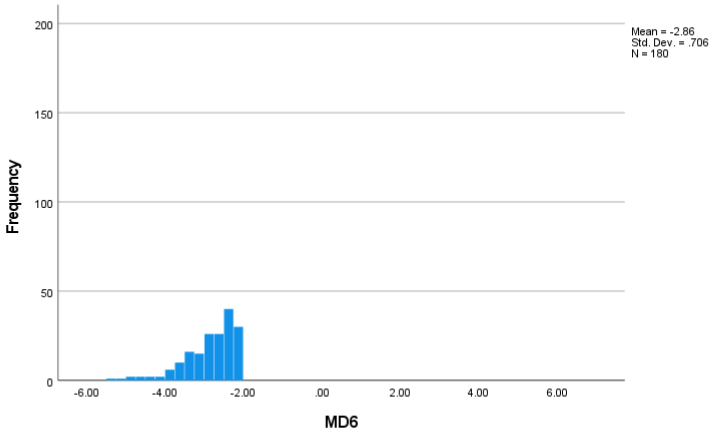
Note. Below the diagonal there is Pearson correlation coefficient (in black) and above the diagonal Spearman’s (in blue)

*p < .05, **p < .01, ***p < .001

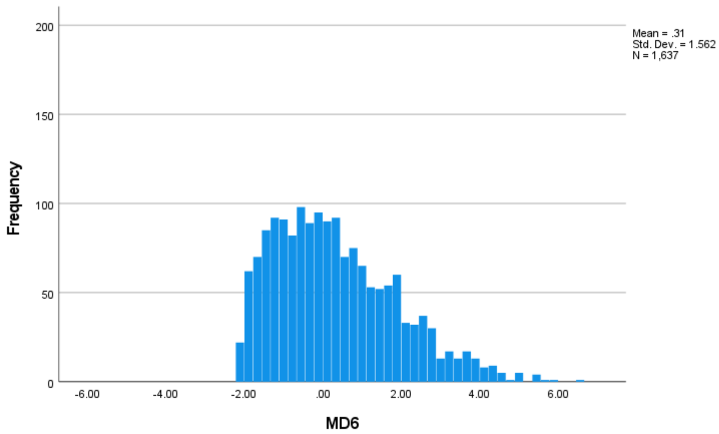
Appendix B

Histograms for the math scores for those belonging to the lowest 10%, those belonging to the remaining 90% and for the whole sample.

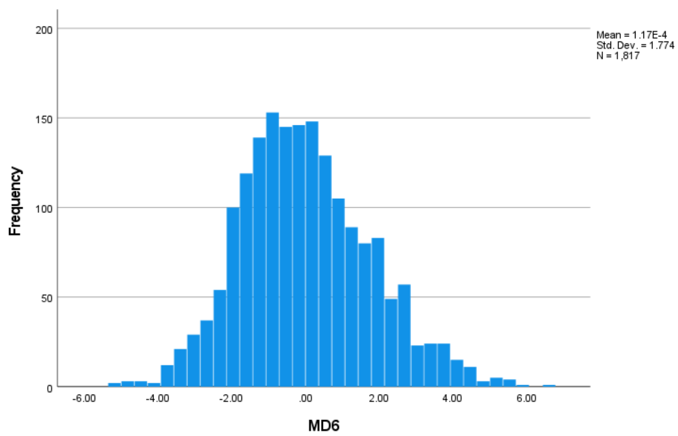
a) Lowest 10%



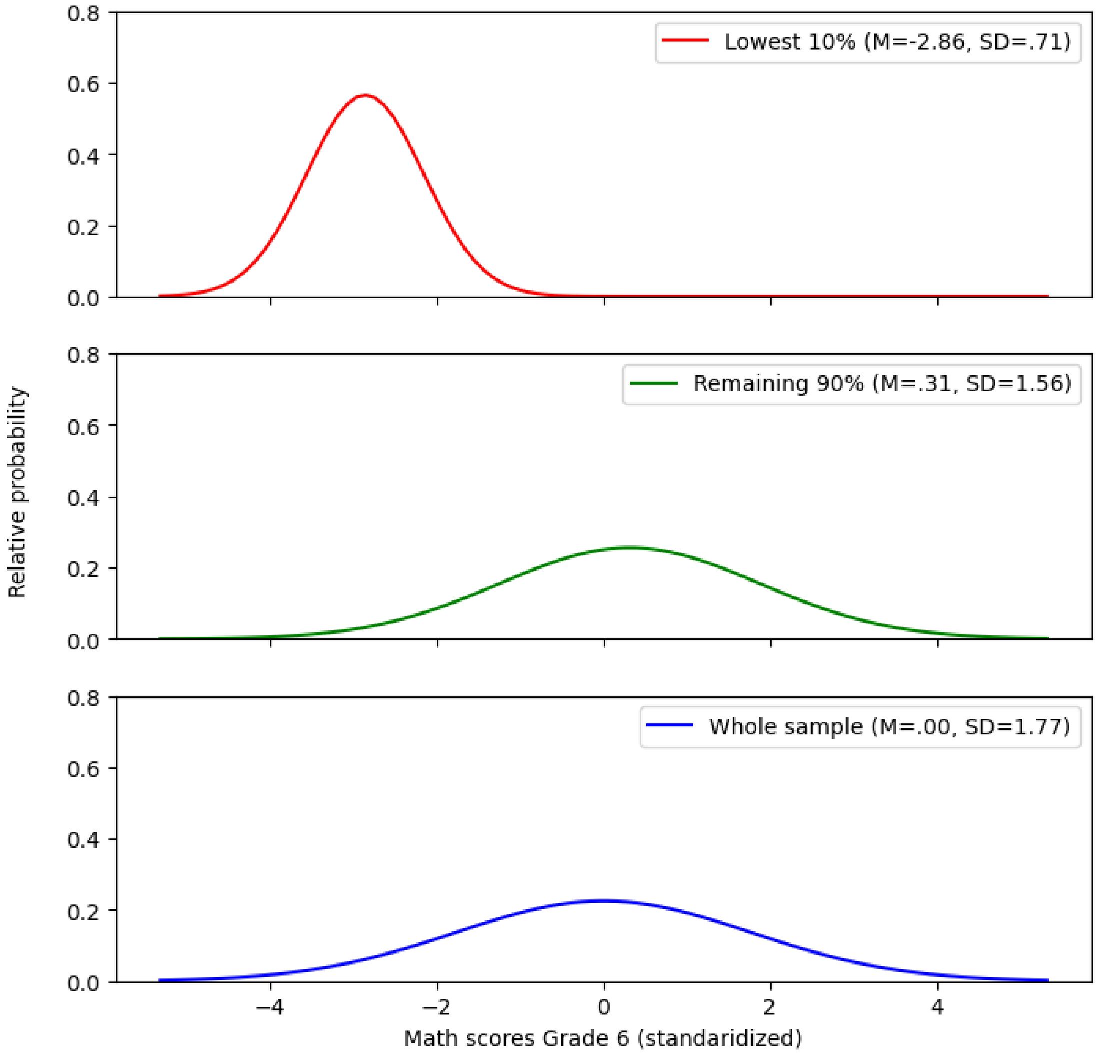
b) Remaining 90%



c) Whole group



Distribution of math scores for those belonging to the lowest 10%, those belonging to the remaining 90% and for the whole sample



Appendix C

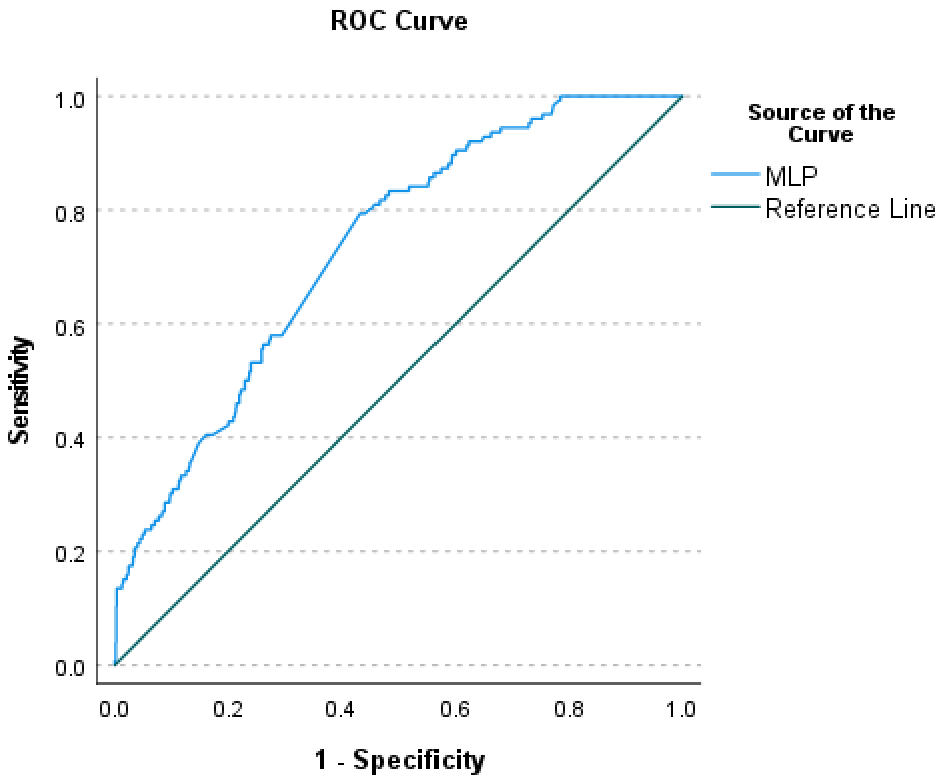
Difference in the training and the testing samples of the percentage of imputed cases in the MD and noMD groups per seed

Seed value	Testing sample		Training sample		Testing and training sample difference (absolute value) of imputed cases in the MD group	Testing and training sample difference (absolute value) of imputed cases in the noMD group	Testing and training sample difference (absolute value) of imputed cases in the MD and noMD group
	noMD	MD	noMD	MD			
7291626	84.02%	90.48%	82.88%	90.91%	0.43%	1.14%	0.71%
7281626	81.33%	92.06%	84.01%	90.08%	1.98%	2.68%	0.70%
7271626	**82.96%**	**91.38%**	**83.33%**	**90.48%**	**0.90%**	**0.38%**	**0.53%**
7261626	84.88%	93.88%	82.50%	89.63%	4.25%	2.38%	1.87%
7251626	81.19%	94.64%	84.09%	89.06%	5.58%	2.90%	2.68%
7241626	83.40%	84.21%	83.14%	93.70%	9.49%	0.26%	9.23%
7231626	82.46%	97.96%	83.55%	88.15%	9.81%	1.09%	8.72%
7221626	83.30%	94.44%	83.19%	89.23%	5.21%	0.11%	5.10%
7211626	84.04%	96.00%	82.86%	88.81%	7.19%	1.18%	6.01%
6291626	83.64%	91.94%	83.04%	90.16%	1.78%	0.60%	1.18%
6281626	81.97%	92.98%	83.76%	89.76%	3.22%	1.79%	1.43%
6271626	84.17%	87.69%	82.83%	92.44%	4.75%	1.34%	3.41%
6261626	82.11%	86.79%	83.70%	92.37%	5.58%	1.59%	3.99%
6251626	83.64%	82.00%	83.04%	94.03%	12.03%	0.60%	11.43%
6241626	82.62%	94.64%	83.48%	89.06%	5.58%	0.86%	4.72%
6231626	83.47%	93.44%	83.12%	89.43%	4.01%	0.35%	3.66%

6221626	79.76%	84.78%	84.74%	92.75%	7.97%	4.98%	2.99%
6211626	82.52%	84.91%	83.52%	93.13%	8.22%	1.00%	7.22%
6201626	83.47%	87.27%	83.11%	92.25%	4.98%	0.36%	4.62%
5201626	82.90%	87.50%	83.36%	91.91%	4.41%	0.46%	3.95%

Appendix D

ROC Curve for the prediction of MD for the training models. The area under the curve is .732 (p<.001, 95% C.I. .690 - .774). In the training sample there were 1,272 cases.



Appendix E

Results when oversampling the MD group in the training sample

Table 1

Corresponding sensitivity, specificity, and precision estimates for different cut-off values for the prediction of math difficulties

Cut-offs	Sensitivity	1-specificity	Specificity	Precision	F1-score
10%	.983	.891	.109	.116	.208
15%	.983	.832	.168	.123	.219
20%	.983	.778	.222	.131	.231
25%	.983	.723	.277	.139	.244

Table 2

Specificity and precision values corresponding to specific sensitivity values for the prediction of math difficulties

Sensitivity	1-specificity	Specificity	Precision
.690	.261 - .359	.641 - .739	.186 - .240
.707	.359 - .396	.604 - .641	.175 - .190
.931	.559 - .589	.411 - .441	.158 - .166

Figure 1

ROC curve for the prediction of math difficulties for the testing models. The area under the curve is .770 (p<.001, 95% C.I. .710 - .831). In the testing sample there were 545 cases.

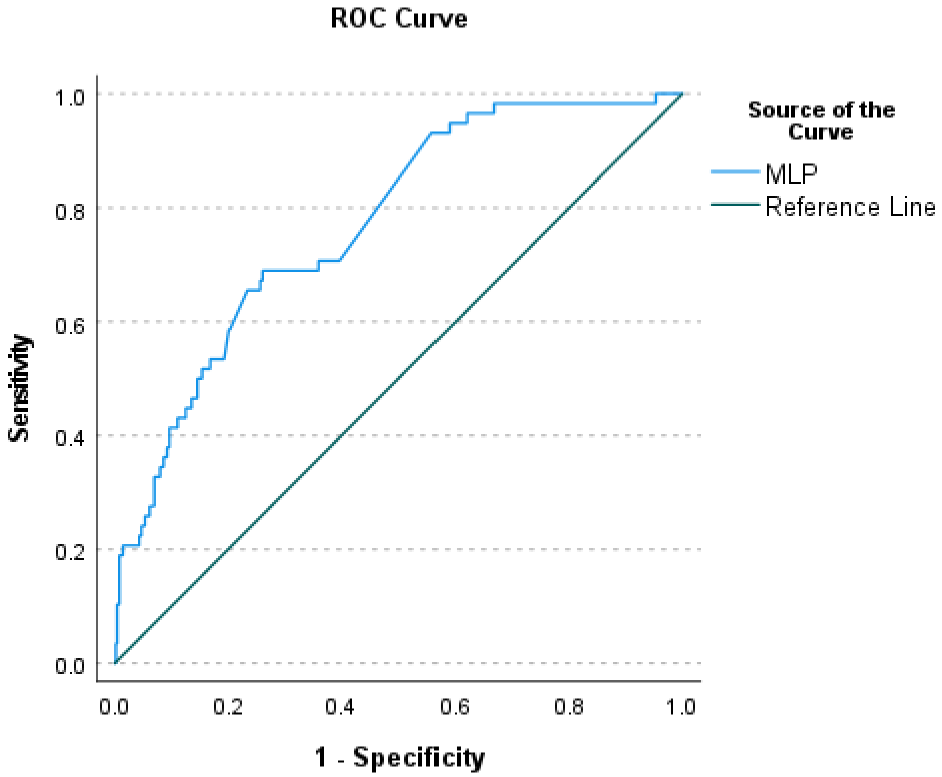


Figure 2

Confusion matrix for predicting math difficulties in the testing models when the sensitivity is .707