Rudimentary stages of the mathematical thinking and proficiency: Mathematical skills of low-performing pupils at the beginning of the first grade

Jari Metsämuuronen and Annette Ukkola

Finnish National Education Evaluation Centre (FINEEC)

A national-level dataset (n = 7770) at grade 1 of primary school is re-analyzed to study preconditions in proficiency in mathematical concepts, operations and mathematical abstractions and thinking. The focus is on those pupils whose preconditions are so low that they are below the first measurable level of proficiency in the common framework with reference to mathematics (CFM). At the beginning of school, these pupils may not be familiar with, e.g., the concepts of numbers 1–10, they may not be aware of the consecutive nature of numbers, and they have no or very limited understanding of the basic concepts of length, mass, volume, and time. A somewhat surprising finding is that the key factor explaining the absolute low proficiency in mathematics appeared to be a low proficiency in listening comprehension. This variable alone explains 41% of the probability of belonging to the group of pupils who are not able to show proficiency enough to reach the lowest level in any of the criteria. It is understandable that, if language skills are underdeveloped in general, a child is not expected to master the specific mathematical vocabulary either and, hence, the low score in a test of preconceptions in mathematics too. Other variables predicting the absolute low level or preconditions of mathematics are the decision on intensified or special support, status of Finnish or Swedish as second language, and negative attitudes toward mathematics.

Keywords: mathematical thinking, mathematically low-achieving students, national assessment in mathematics, pre-primary education, primary education

ARTICLE DETAILS

LUMAT Special Issue
Vol 10 No 2 (2022), 56–83

Pages: 28
References: 55

Correspondence:
jari.metsamuuronen@gmail.com

https://doi.org/10.31129/10.2.1632

1 Introduction

Mathematical competence is one of the key skills needed in modern society. From the viewpoint of socializing citizens to mathematical concepts and operations, as well as abstraction and thinking, teachers in schools are the key persons because pure mathematic is rarely a natural hobby of children, unlike sports, handicrafts, or reading. The main contents of mathematics are learnt, practically speaking, exclusively in or through the school: in the first grades, in mathematics lessons and while doing school homework (Metsämuuronen, 2013a). From this viewpoint, measuring the level of mathematical thinking and proficiency in mathematics in general makes sense at higher grades. Usually, in Finland, national assessments of learning outcomes are administered at grade 9 (Metsämuuronen, 2009) and the international PISA (Programme of International Student Assessment) and TIMSS (Trends in International Mathematics and Science Study) comparisons at grade 8 (e.g., PISA, 2019; TIMSS, 2020) or even later in adulthood (see the Programme for International Assessment of Adult Competencies [PIAAC], OECD, 2016). Hence, the end-product of the socializing the citizens in mathematics during the school years is well-known and well-followed-up.

Although the systematic socialization to mathematics happens mainly in and through school, children have learned a lot matters that are related to mathematics even before the school age—7 years in Finland. These preconditions on mathematics are in the focus of this article. At the national level, it is very rare to see measures of mathematical thinking and competencies at the beginning of schooling, that is, large studies on what are the first stages of development of mathematical thinking and what kind of proficiency are largely lacking (see, however, e.g., Lerkkanen et al. 2012, where mathematics skill was assessed as a part of the First Steps study by the fluency in counting forwards and backwards number sequencies). In 2018, the Finnish National Education Evaluation Centre (FINEEC) launched a longitudinal assessment to measure the achievement level of pupils and students at different stages of their school years in mathematics and mother language. The first measurement was administered in the first weeks at grade 1 with a minimal effect of school in mathematical thinking (see methods in Metsämuuronen & Ukkola, 2019 and results in Ukkola & Metsämuuronen, 2019; Ukkola, Metsämuuronen, & Paananen, 2020). The dataset gives quite a unique possibility to study the outcome of early childhood development from the mathematic development viewpoint.

In this article, this unique dataset of preconditions of mathematics at grade 1 (n = 7770) is re-analyzed from the viewpoint of mathematical thinking by using the common framework with reference to mathematics (CFM) suggested by Metsämuuronen (2018). CFM divides the mathematics skills into three criteria: proficiency in mathematical concepts, proficiency in mathematical operations, and proficiency in mathematical abstractions and thinking. In Section 2, the factors affecting the development of mathematical skills in the early childhood are discussed which is followed by discussing the characteristics of CFM in Section 3 and methodological matters for the empirical section in Section 4. Section 5 combines these and presents results of proficiency of mathematical concepts, procedures and thinking at the beginning of the grade 1 in schools in Finland. The focus is, specifically, in predicting and detecting the children in whom mathematical skills are underdeveloped or whose skill level is lower than the measurable level when they are at the age of starting school, that is, when they 7 years old. The main research question is, what variables characterize the pupils whose mathematical skills and thinking are very low—even below a measurable level, and which background factors could be used to detect such children. The main research question is divided to three sub-questions:

1. What kind is the overall distribution of preconditions in mathematics at the beginning of the first grade?
2. How do personal factors characterize the pupils with very low preconditions in mathematics at the beginning of the first grade?
3. How do family factors characterize the pupils with very low preconditions in mathematics at the beginning of the first grade?

2 Some known factors affecting the development of mathematical thinking in the early childhood

In comparison with many other countries, in Finland the children enter the school rather late, typically when they turn 7. Because the first 6 years may be radically different, children enter the school with a wide variety of mathematical skills (see Metsämuuronen, 2010; 2013a; Metsämuuronen & Tuohilampi, 2014; Ukkola & Metsämuuronen, 2019; Ukkola et al., 2020). This is caused by the fact that the preliminary concepts related to mathematics are learnt at home or during the preprimary education, and these conditions may vary dramatically (see Ukkola et al., 2020). Hence, some children enter the school with no or very limited knowledge of basic mathematical concepts while some may be already at the level of grade 3 (Ukkola & Metsämuuronen, 2019; 2021). The reasons for this deviance are discussed here, focusing on the factors related to the child and the home background.

2.1 Factors explaining the preconditions of mathematics in literature

Several individual factors have been shown to affect the development in general and in mathematics in specific. Some of these are sex (see, e.g., Metsämuuronen 2017a; Niemi et al., 2020, 2021), and language background including medium of instruction being the mother tongue (first language, L1) or the second language (L2) (see, e.g., Kuukka and Metsämuuronen, 2016). Other important factors found to explain the competence are relative age of starting the school (see, e.g., Dhuey et al., 2019; Kivinen, 2018; Ukkola et al. 2020), attitudes toward school and self-efficacy (see, e.g., Aunola, Leskinen, & Nurmi, 2006; Bandura, 2012; Lerkkanen et al., 2010, 2012; Tuohilampi and Hannula, 2013; see in-depth in Ukkola et al., 2020).

Three factors related to child's home background are found to be important in explaining the pupil's school performance: education of the parents (see, e.g., Kivinen, Hedman, & Kaipainen, 2012; OECD 2015), economic factors (see, e.g., Erola, Jalonen, & Lehti, 2016; Paju, 2020; Palomäki et al., 2016; Sirniö, 2016), and genetics including inheritance related to mathematics (see, e.g., Dilnot et al., 2016; Malanchini et al., 2020). The first two are commonly combined as factors related to socioeconomic status (SES), and the latter has been important factors in explaining learning disabilities, for example (see, e.g., Eklund, 2017).

All in all, many factors related to a child—of which many are given, and of which the child cannot affect at all—are related to the early childhood development in general and mathematics development in specific. In-depth discussion of all these matters is found in Ukkola and colleagues (2020). These are discussed further in the empirical section.

2.1 What is known of the combined factors explaining the low level of preconditions of mathematics?

Ukkola and colleagues (2020) collected quite a variety of possible variables explaining the high and low levels of preconditions in general at grade 1. They sought a simple model, a kind of check list type of presentation of the factors predicting the exceptionally low level of preconditions in the population. Based on logistic regression analysis (LRA) and decision tree analysis (DTA), they came up with five binary variables explaining the low performance in the test of preconditions in mathematics and language combined (Table 1).

The strongest predictor for the low performance in the test of preconditions is whether the child was decided to be on intensive or special support even before the school age. The risk of these children to belong to the lowest quartile (Q1) is 4.6 times higher and to the lowest decile (D1) 5.3 times higher than when it is not the case. Second strongest predictor is the L2 status with 3.3- and 4.2-times risk, respectively. Other factors such as learning disabilities of the parents, relatively young school starting age, and guardians' low education give 1.5- to 2.0-times risks for a child to belong to the group of exceptionally low preconditions in general. Notably, the explaining powers of the models are not very high ($R^2_{Adj} = 0.12 - 0.13$) referring to the fact that even though the tendency is clear, nothing is determined even if the child happens to be born with less advantageous genes and a "wrong" time of the year or in a "wrong" country.

Table 1. Five main factors explaining the low level of preconditions in mathematics and language combined at grade 1 (Ukkola et al., 2020)

Variables in the model[1]	B[2]	risk to be at Q1 (lowest quartile)	risk to be at D1 (lowest decile)
Constant	547		
Support in three levels (1 = decision on intensive or special support, 0 = general support meant for all pupils)	-65	4.58	5.30
Status for Finnish/Swedish as a second language (L2 status) (1 = registered L2 status, 0 = no L2 status)	-63	3.29	4.17
Learning disabilities in the close family (1 = at least one type of learning disability in parents, 0 = no learning disabilities in the close family)	-36	1.99	1.83
Relative age of starting school (1 = months 9–12, 0 = other months)	-35	1.76	2.01
Education of the guardians (1 = both or either of the guardians have basic education or vocational qualification, 0 = other alternatives)	-30	1.71	1.50
predicted level if in group 1 in every factor	320		
Explaining power R^2_{Adj}	0.12	0.13	0.13

1) Variables are ordered by the risk related to Q1
2) regresson weight

3 Quest for common standards for mathematics

Assessing the *absolute* level of mathematical skills or thinking is not as simple as sometimes it is thought to be. Specifically, the task is even more difficult when test takers are young and there is no obvious measurement stick which would tell what "good" or "high level" is. The most obvious challenge is that there is no commonly accepted general framework for proficiency in mathematics.

Metsämuuronen (2018) suggested a common framework for mathematics (CFM) based on the levels used in the common European framework of reference for languages (CEF or CEFR; https://www.coe.int/en/web/common-european-framework-reference-languages)—after all, mathematics is a kind of universal language, and mathematical skills tend to cumulate. In CFM, the domains are reduced to three elements: 1) proficiency in mathematical concepts (M1), 2) proficiency in mathematical operations (M2), and 3) proficiency in mathematical abstractions and thinking (M3). The rationale for the first two criteria is obvious: to master even the simplest and most mechanical mathematical operation, a certain level of proficiency in mathematical concepts is needed: the concepts of numbers and their representations, consecutive nature of the numbers, and certain basic shapes such as the triangle, square, and circle. The rationale for proficiency in mathematical abstractions and thinking is that the essence in mathematical proficiency (maybe except at a theoretical level) is to transform everyday life challenges into a mathematical form and solve the problems by using mathematical operations. Without proficiency in mathematical abstractions and thinking, the proficiencies in concepts and operations are largely useless; one may know how to do a mathematical operation (such as derivation) but have no idea when or why to use it.

The basic mathematical concepts, operations, as well as the elements of mathematical abstractions and thinking are usually hierarchically organized in the normal educational process. For example, to manage powers, the procedure of multiplication is needed and to master multiplication, the procedure of addition is needed. Hence, we understand that it is wise to start teaching and learning mathematics with concrete things such as addition and subtraction of the natural numbers before introducing decimals and rational numbers.

The standard levels in CFM are based on this logic which are divided into levels A, B, and C (Table 2). The level A refers to the elementary and basic level with the relevance to the *everyday life*, B refers to an advance level with relevance to the *further studies* in several professional areas like statistics, engineering, or economics, and C is the *professional* level mathematics needed either in practical fields (like that of statisticians, advanced researchers, economists, or engineers) or in the theoretically oriented fields (like that of professors or researchers of pure mathematics, physics, astronomy, or chemistry).

As far as this article is concerned, only level A1 is relevant at the beginning of the school even though there may be some prodigies among the pupils. The descriptions and stages in CFM are based on the national core curricula of mathematics in Finland (EDUFI, 2004, 2014 for the basic education; EDUFI, 2003, 2015 for the upper secondary general education).

Table 2. Brief descriptions of the CFM levels (Metsämuuronen, 2018)

CFM level (main stages)	CFM level	Short Description
A1 Elementary proficiency	A1.1	First stage of elementary proficiency
	A1.2	Developing elementary proficiency
	A1.3	Functional elementary proficiency
A2 Basic proficiency	A2.1	Developing of basic proficiency
	A2.2	Functional basic proficiency
B1 Advanced proficiency	B1.1	First stage of advanced proficiency
	B1.2	Developing advanced proficiency
B2 Functional advanced proficiency	B2.1	First stage of Functional advanced proficiency
	B2.2	Functional advanced proficiency
C Professional level	C1	Basic Professional level
	C2	Advanced Professional level

In CFM, the first measurable level is A1.1 (First stage of elementary proficiency) where the basic elements needed in mathematics such as the numbers and basic shapes related to geometry are identified. A1.1 refers to the level at which the rudimentary basic elements of mathematical proficiency are mastered. At this level, among others, one is familiar with the numbers, but the use in mathematical operations is very limited; one recognizes the basic two-dimensional shapes (circle, square, triangle) and their three-dimensional counterparts (ball, box, and pyramid) and can couple their name with pictures; one can express some limited mathematical expressions, such as the order of numbers; one knows the importance of numbers in stating amount and order; one knows how to write numbers but the proficiency in using formulated mathematic expressions is very limited.

The empirical section is specifically about pupils *below* level A1.1. These pupils have the most disadvantageous start for their mathematical career although they may also give the most joy to the teacher when noticing how well they advance in school despite the low level at the beginning.

If someone is at a level lower than A1.1, from the mathematical *concepts* viewpoint, he or she may not (adequately) know the numbers in the range 1–10; may not be aware of the consecutive nature of numbers; may not be able to name the basic forms of the circle, square, triangle, ball, box, and pyramid; and may have no understanding of the basic concepts of length, mass, volume, and time.

From the mathematical *operations* viewpoint, a person below level A1.1 may not be able to recognize or write the numbers; may not understand the consecutive order of numbers; may not be able to categorize the basic shapes into groups without messing with different sizes, colors, and positions; may not be able to couple the names of basic shapes with pictures; and may not know how to measure length, mass, and time in the everyday life.

From the mathematical *abstractions and thinking* viewpoint, a person below level A1.1 may not have the basic understanding of the concepts of adding, subtracting, dividing, or multiplying; may not have the basic understanding of unseen numbers (for example, what number is missing in the consecutive order); may not have the basic understanding how to place things in order, to find opposites for things, to classify things according to different attributes, or to state the location of object for example by using the words *above*, *below*, *on the right*, *on the left*, *behind*, and *between*.

Obviously, only a new-born baby may be at the stage where the mathematical thinking or understanding of concepts and operations would be non-existent—all children starting the school have *some* mathematical preconditions and skills as discussed above.

4 Methodology

The general methodological issues related to the dataset used in the empirical section are discussed in detail by Metsämuuronen and Ukkola (2019). Some points relevant to this article are highlighted in Section 4.1 about the sample and datasets while Section 4.2. is about the test of mathematic. Section 4.3 describes the procedure of standard setting and Sections 4.4 and 4.5 describe the practicalities related to the analysis itself.

4.1 Sampling and data

A dataset of n = 7770 pupils from grade 1 was collected in August 2018 from 264 schools selected by using stratified random sampling. The selected schools comprise 13% of all schools teaching grade 1 and the pupils are 19% of all grade 1 pupils in Finland. Swedish population was oversampled (28% of the Swedish-speaking schools) for a relevant analysis of this minority. Of the pupils in the target group, 97.5% participated in the test.

Part of the information concerning the child was provided by the guardians of the child. This dataset comprises n = 4,316 children (56% of the pupils) and it includes information of the parents as well as such information of the child that was difficult to extract from the child, for example, concerning their interests. Hence, some analyses can be done by using the whole dataset while some other interesting variables are restricted to a smaller number of pupils and, in the latter case, the dataset is slightly biased toward higher-educated families; as usual, guardians with a higher educational level seemed to have been more active in answering the questionnaire (see closer Metsämuuronen & Ukkola, 2019). Relevant characteristics of the dataset are collected in Table 3.

Using pupils' ID numbers, relevant information was added to the dataset from the national KOSKI-database. This included information such as home language, L2 status, and information concerning the 3-stage support.

4.2 Test items, validity, and reliability of the test score

The content of the mathematics test was based on content areas in the National core curricula for preprimary education (EDUFI, 2016) and for basic education (EDUFI, 2014). Based on these norms, the contents of the mathematic test comprised of three main areas: geometry and measurement, numbers and calculation, and mathematical thinking (Table 4). From the construct validity viewpoint, the test comprises all areas of the "theoretical framework" from the core curricula.

The sub-test of mathematics comprises 58 items totaling 62 points. The lower bound of reliability of the test was $\alpha_R = 0.88$ by coefficient alpha and, after correction for deflation by using Somers' *D* instead of Pearson correlation in the coefficient (see Metsämuuronen, 2020, 2021, 2022a, 2022b; Metsämuuronen & Ukkola, 2019), $\alpha_D = 0.94$. Hence, in general, the score is accurate enough to discriminate the test takers from each other.

After a pre-trial, two task types were selected to the final test: "press" and "move" (see Figures 1 and 2). The pupils did the assessment tasks in the school's language of instruction using a tablet or a computer. The tasks were speech-instructed. Each pupil logged into the testing system through a unique sequence of graphical symbols and selected an avatar (such as a robot) to lead into the test (and "speaking" the instructions). Children learnt quickly how to use these two task types by a training sequence before the test. Teachers were instructed to help the child if some technical challenges occurred but not to interfere with the answering process. After selecting the item, an arrow appeared automatically. The child pressed the arrow to move to the next task.

Table 3. Characteristics of the datasets of grade 1

Variable		**Whole dataset (*n*)**	**Pupils *n* = 7770 (%)**	**Guardians *n* = 4316 (%)**
Sex	Girl	3875	49.9	49.9
	Boy	3895	50.1	50.1
Instruction language	Finnish	6902	88.8	91.1
	Swedish	868	11.2	8.9
Syllabus	Finnish	6405	82.4	
	Swedish	834	10.7	
	Fin/Swe as second language (L2)	531	6.8	
Regional state administrative agency	South Finland	3015	38.8	38.2
	South-West Finland	917	11.8	11.8
	East Finland	732	9.4	8.7
	West and Middle Finland	1672	21.5	22.2
	North Finland	780	10	10.3
Type of municipality	City	5468	70.4	70.5
	Population density area	1184	15.2	15.7
	Rural	1118	14.4	13.8
L2 status	No	7239	93.2	95.1
	Yes	531	6.8	4.9
Three-stage support	General support	6971	89.7	92.2
	Intensive support	521	6.7	5.7
	Specific support	278	3.6	2.1
Learning disabilities in parents	No learning disabilities	3125		72.4
	One type of learning disability	720		16.7
	Several types of disabilities	471		10.9
Highest education in the family	Basic education	48		1.1
	Vocational education	927		21.5
	Matriculation examination	343		8.0
	Polytechnic education	1346		31.2
	University education	1535		35.6
	Else	111		2.6
Level of preconditions	Score in mathematic test	500		514

Table 4. Contents of the mathematics test

Domain	Topic	Number of items	Reliability (α_R)	Reliability (α_D)
Mathematics as whole		58	0.879	0.940
Geometry and measurement	Geometry	9		
	Measurement	10		
	Time and Clock	4		
Numbers and calculation	Calculation	17		
	Numbers	8		
Mathematical thinking	Oral tasks	17		
	Reasoning	8		
	Relations	17		

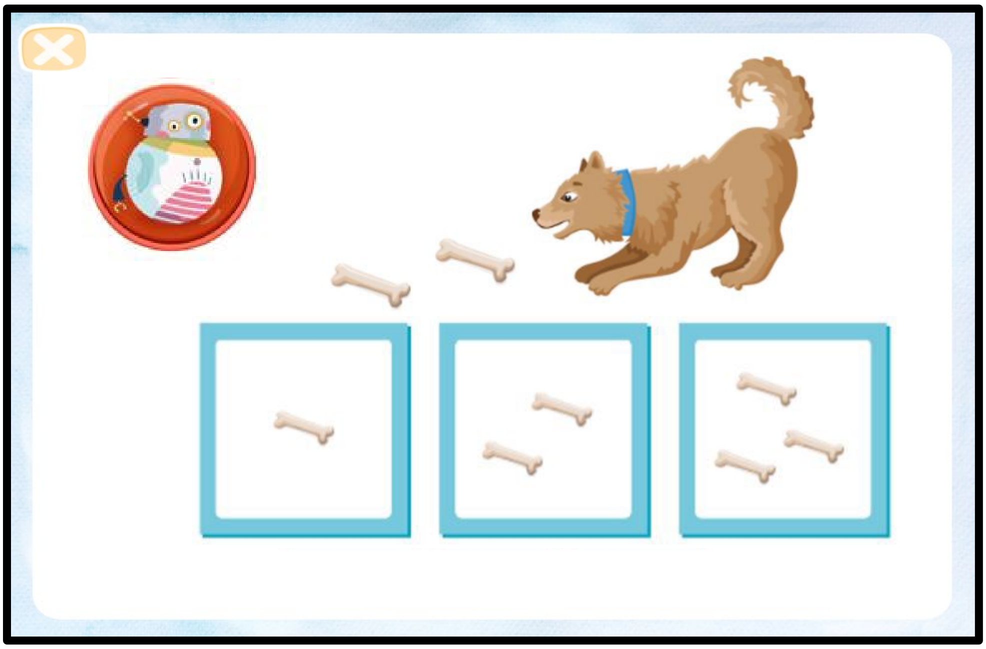

Figure 1. A “Press” type of task (“Press the figure which has one bone less than the dog has”)

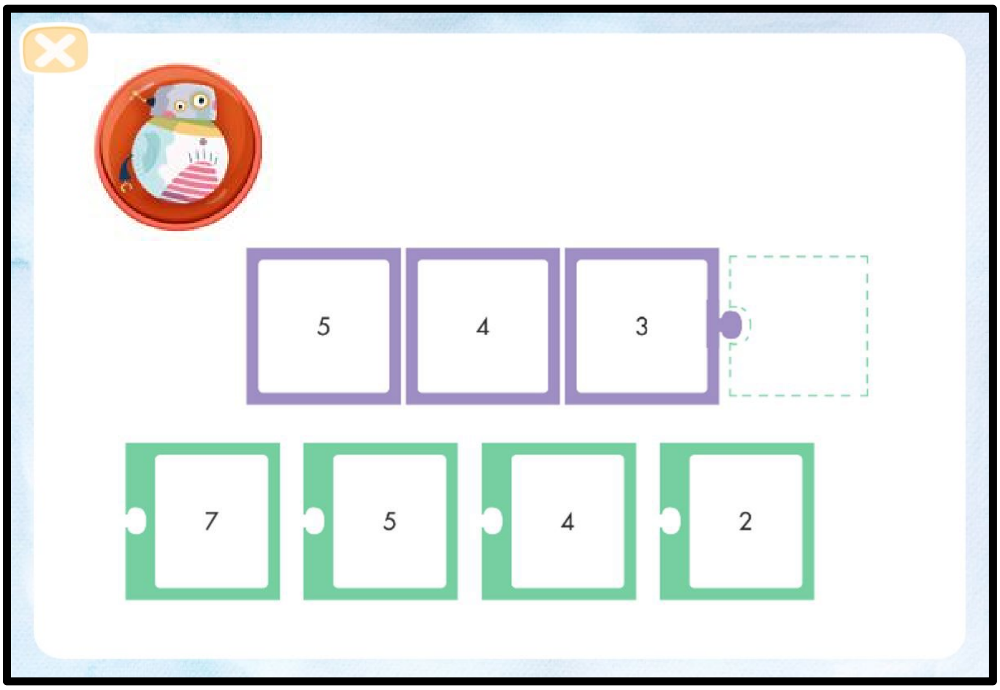

Figure 2. A "Move" type of task ("*Move the correct number to the queue*")

4.3 Standard setting

The original analysis (see Ukkola & Metsämuuronen, 2019; Ukkola et al. 2020) was based on norm-reference assessment. For the re-analysis, a standard setting was administered by using a method called 3TTW (Three-phased Theory-based and Test-centered method for the Wide range of proficiency levels, Metsämuuronen, 2013b).

At the first phase of 3TTW, items were classified on different bins of standard systemic based on the (theoretical) content of the item; first based on three criteria (concepts, operations, and thinking) and second, based on the standard level (A1.1, A1.2, and A1.3). Only in the criteria on mathematical abstraction and thinking, it was possible to find items that fit the level A1.3. These items were more demanding where a semi-complicated real-world problem was needed to be transformed into a mathematical form and to solve for instance in "*In the morning, the thermometer showed +2 Celsius. During the school day, it dropped six degrees. What is the temperature after the school day? Press the correct number.*" For a possible interest of a reader, in this type of item, 12.5% of the pupils of the grader 1 were able to give a correct answer at the beginning of school.

At the second phase, items belonging to a same bin were summed up. The sums were transformed into a form that indicated whether the test taker had reached the level of proficiency required for a specific level of standard in a specific criterion. For this, a proper cut-off in the sums was set to mark the needed level of proficiency for the standard level. Usually, these boundaries are named as “weak pass”, “pass”, and “strong pass” (e.g., Van der Schoot, 2009)—strong pass may mean that one needs to score at least 80% of the total score in the bin. A minimum boundary was set to 50% correct, that is, to “weak pass”: to belong to a certain standard level, the test taker needed to solve at least 50% of the tasks correctly (see Figure 3).

At the third phase, each test taker got his/her profile of passing and failing in the levels of standard systemic. In most cases, the profile was “pure” in a sense that if one was able to solve more-demanding tasks, also the less-demanding tasks were solved. Then, it is straightforward to conclude that if a test taker can show proficiency enough for levels A1.1 and A1.2 but not for A1.3, a credible proficiency level for the test taker is A1.2 (see detailed, Metsämuuronen, 2013b).

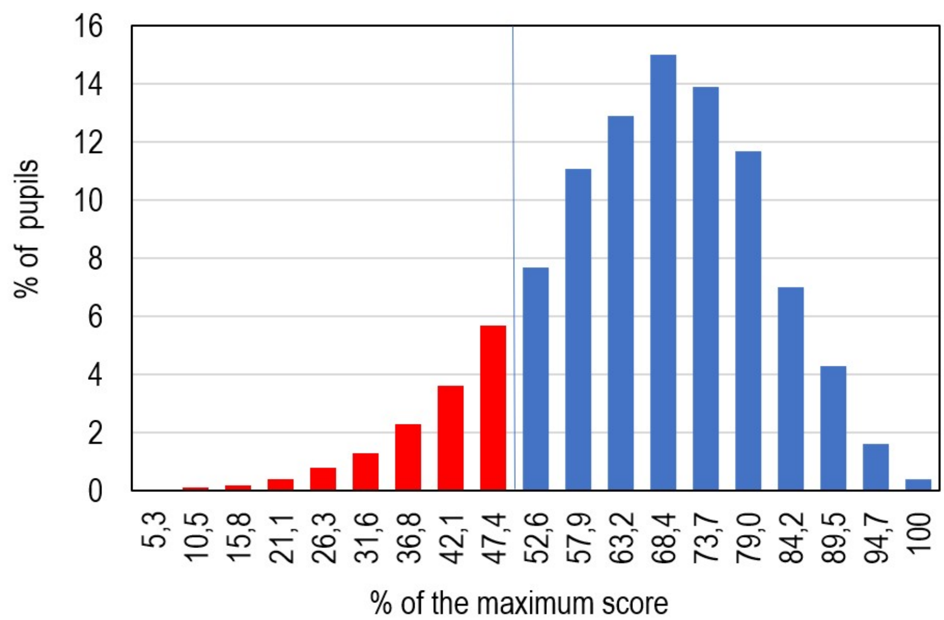

Figure 3. Distribution of proficiency in M3, level A1.1 items summed, and the boundary of a weak pass (50% correct)

4.4 Variables used in the analysis

The score of mathematics is formed of the raw score by one-parameter item response theory (IRT) modelling, that is, the Rasch modelling. The outcome (theta score) is a logistic transformation of the raw score. The original theta score is a standardized normal variable where the average scorer gets the value 0. This score is further transformed to a T10 form, that is, $Y = 100 \times X + 500$ leading to a score where the average test taker gets the score 500 and the standard deviation is 100. The same transformation and mechanics are used, for instance, in PISA- and TIMSS inquiries (see, e.g., PISA, 2019). In what follows, also some other test scores such as different sub-tests related to the medium of instruction of the school and attitude scales are used in analysis. Validity and reliability of these tests are described in Metsämuuronen and Ukkola (2019).

For a re-analysis of the dataset, a standard setting was administered (see above), and those pupils were detected who showed the lowest absolute level of preconditions of mathematics at grade 1 in all criteria (n = 608; 7.8% of the pupils). In theory, these pupils are below the lowest measurable level of proficiency in mathematics at the beginning of grade 1. Naturally, they have some proficiency in mathematics—in some cases maybe almost half of the task solved—but not enough to reach the lowest standard level A1.1 in any of the areas on the criterion systemic. This dummy variable (later "below A1.1") is mainly discussed in what follows.

The analysis is mainly exploratory in nature. Hence, relevant descriptive variables such as sex, relative age of school start, and family factors are used to profile these pupils with the least advantageous start of the mathematical studies in school. Finally, by combining the statistically significant predictors, a model parallel to Table 1 is formed to predict the grouping of the lowest level preconditions in an absolute sense, that is, the group "below A1.1".

4.5 Methods of analysis

Three main analytical tools are used in the analysis: a data mining tool decision tree analysis (DTA), traditional logistic regression analysis (LRA), and traditional general linear modeling (GLM) in IBM SPSS environment. These methods are generally known and, hence, there is no need to describe them further (see, e.g., Metsämuuronen, 2017b). In LRA, standard statistical procedure with conditional selection of variables is used with Nagelkerke's (1991) adjustment for the explaining power (R_{Adj}^2). GLM is used mainly as one-way ANOVA; in post hoc analysis, Šidák's (1967) procedure is used; in the case, it gives more plausible correction for p-values than the traditional Bonferroni correction (see discussion in Metsämuuronen, 2017b). For effect sizes, Cohen's f (Cohen, 1988) is used; the classic, rough boundaries for small-, medium-, and large effect size are $f < 0.1$, $f = 0.2–0.3$, and $f > 0.4$, respectively. In DTA, CHAID algorithm (Kass, 1980) is used, and child nodes with three levels were allowed as is default in SPSS (see detailed, Metsämuuronen, 2017c).

5 Results

5.1 Overall distribution of preconditions in mathematics at the beginning of the first grade

Overall, the distribution of test score of preconditions in mathematics forms a slightly widened normal distribution (Figure 4) referring to the fact that whole population may be comprised of three to four different normal populations with slightly different means. This seems to fit what was previously noted by Metsämuuronen (2017a; see also Metsämuuronen & Tuohilampi, 2014): children starting the school in Finland seem to form four populations. *Developing pupils* have no or very thin idea of mathematical concepts or thinking—this group is very small in Finland. *Beginner pupils* have some academic preconditions and understanding of mathematical concepts and thinking although those may be very limited—this group is also rather small in Finland. The target group in this article consists mainly of pupils in these two groups. *Normally developed pupils* form the main population. They recognize or master basic concepts such as natural numbers in a limited range and can name basic forms such as triangle, circle, and square; they may be able to solve simple mathematical problems by using adequate mathematical operations; and they may have basic understanding of measuring mass and time, for instance. *Advanced pupils* form the highest performing segment of the cohort. Their mathematical performance at the beginning of school may already be partly at the level of grade 3. Some pupils in this group may be categorized as *exceptionally advanced pupils*. Notably, in the dataset the ultimately highest and lowest-performing pupils were boys. Also, in both extremes (scores < 200 and >800), the number of boys is twice that of girls. This fits with the greater male variability hypothesis discussed by, e.g., Baye and Monseur (2016), Johnson, Carothers, and Deary (2008), Machin and Pekkarinen (2008), and O'Dea and collegues (2018). However, the number of ultimately performing pupils is rather small.

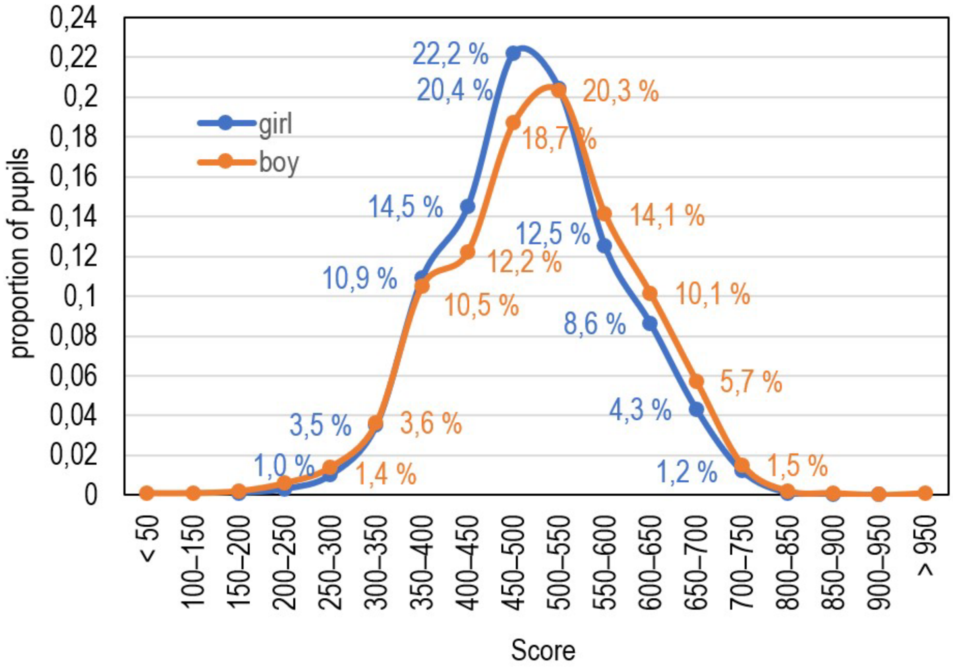

Figure 4. Distribution of the score of preconditions in mathematics at grade 1

After the standard setting, the distributions of proficiency levels are as in Figure 5. In criteria M1 and M3, 12–13% of the pupils fall in the category “below A1.1” while, in criterion M2, 51% of pupils fail to reach the level A1.1. This refers to the fact that, at the beginning of grade 1, pupils may know well the basic natural numbers and recognize and name basic shapes, and they may show some elementary mathematical thinking, but they cannot use much mathematical operations. This makes sense because the mathematical operations are not taught in the preprimary education in Finland; these are taught in school.

By labeling the standard levels with ordinal 0, 1, and 2 in M1 and M2 and 0, 1, 2, and 3 in M3 and summing up the levels of different criteria, we get a rough distribution of absolute proficiency levels for each pupil (Figure 6). While 7.8% of the pupils fall into the category “below A1.1”, 22.5% of the pupils appeared to be at the levels A1.2 or A1.3 in all criteria. Notably, the proportion of pupils in the highest category is exceptionally high because the category consists, factually, of *two* categories, “A1.2 in all criteria” and “A1.2 in criteria M1 and M2 and A1.3 in criterion M3”. The middle levels are formed by varied combinations of the standard levels and, hence, their interpretation is ambiguous.

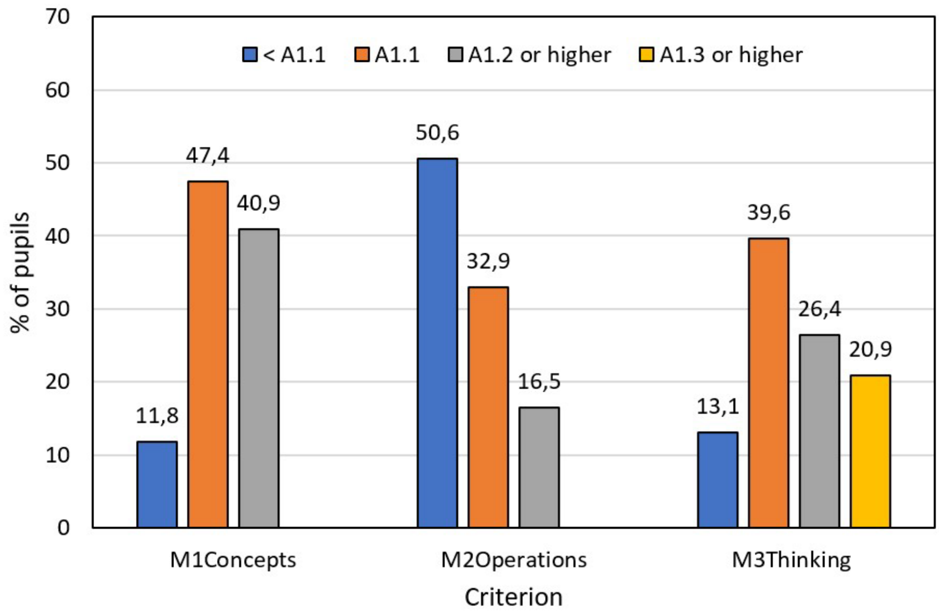

Figure 5. Distributions of standard levels of different criteria in mathematics at grade 1

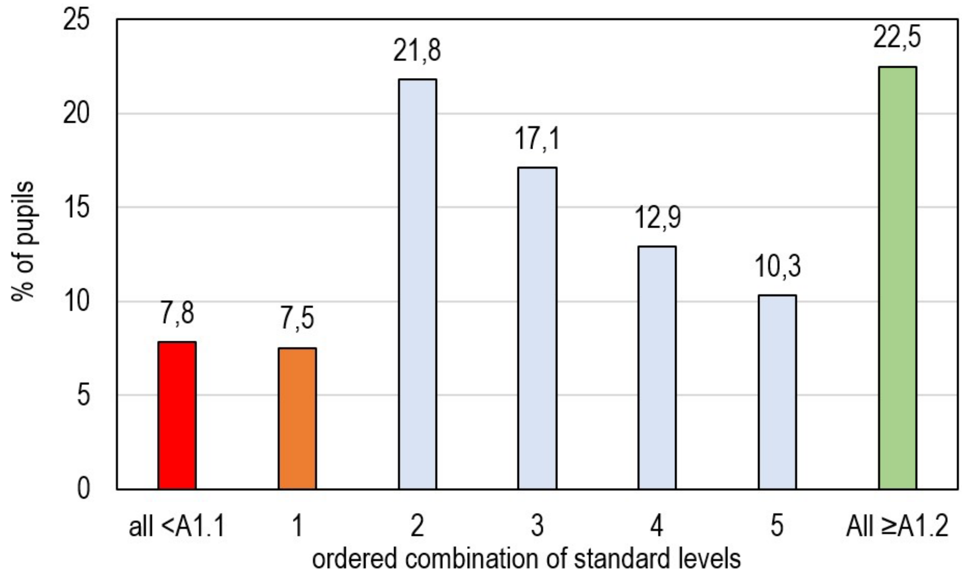

Figure 6. Distributions of combined standard levels in mathematics at grade 1

The average score in the group “below A1.1” was 329 (std. dev. 54.57) and in the highest achieving group “A1.2 or more in all criteria”, it was 628 (std. dev. 58.50) (Figure 7). The difference between the groups is, obviously, statistical significant ($F(6, 7763) = 5422.54$, $p < 0.001$) and the difference between the extreme groups is remarkable ($f = 2.04$). Also, the average scores in each level of ordered proficiency levels differ from each other statistically significantly (GLM, post hoc tests, all $p < 0.001$).

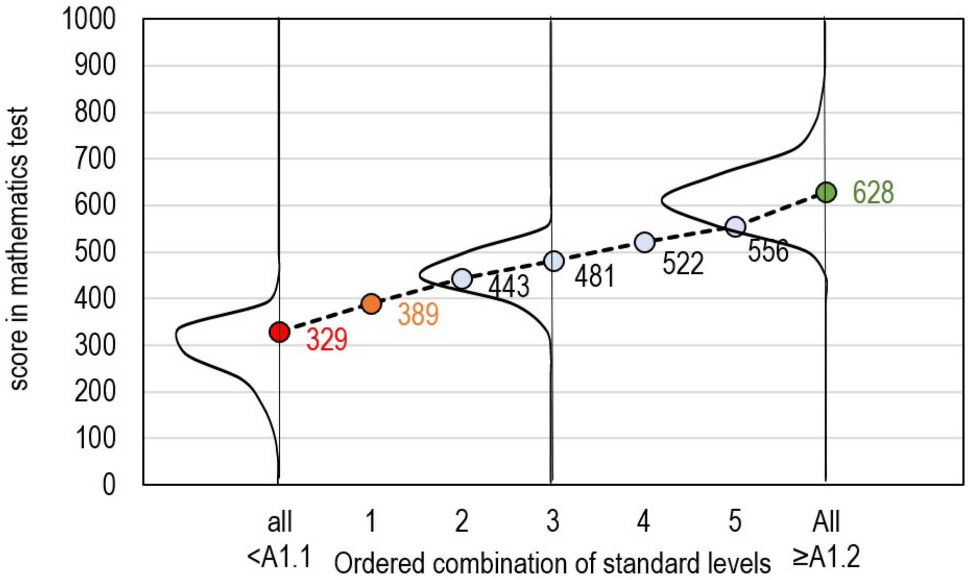

Figure 7. Relation of the test score and the combined standard levels in mathematics at grade 1 with observed distributions for selected levels

5.2 Personal factors characterizing the pupils with very low preconditions in mathematics at the beginning of the first grade

When focusing on the pupils below the lowest measurable standard level (n = 608), DTA suggests that these pupils scored low also in the general test that combined mathematics and the medium of instruction of the school (Finnish/Swedish). Hence, these pupils low-performed not only in mathematics, but they were at a lower level in preconditions for the school in general. Because the test score in the medium of instruction of the school correlates almost one to one with the total score (r = 0.996) caused by the fact that almost all items include a linguistic component (see Metsämuuronen & Ukkola, 2019; Ukkola et al., 2019), it is expected that deficiencies in proficiency in the medium of instruction may explain well the low performance in mathematics.

Of all 17 sub-tests related to proficiency in language (see Metsämuuronen & Ukkola, 2019), DTA suggests that the low score of proficiency in *listening comprehension* predicts inclusion in the low-achieving group the strongest; 68% (n = 415) of the pupils in the group “below A1.1” came from the group where the score of the listening comprehension was below 372.19. By LRA, the main effect of the listening comprehension (dummied into lower and higher than score 372,19) is remarkable: the risk of belonging to the group “below A1.1” is 44,2 times higher if the pupil scored 372 or lower, and the explaining power of the simple model is high

(R^2_{Adj} = 0.408). Also, in GLM, the effect size is high (f = 0.71). Notably, 72.6% of the pupils in this group were *not* immigrants with an L2 status. Logically, if the language skills are underdeveloped in general, the pupils are not expected to master the more demanding specific mathematical vocabulary either. Hence, they get a low score in mathematics too.

L2 status also has its own—although small—main effect in predicting belonging to the group "below A1.1". When L2 status is added to the model of LRA with the dummied score of listening comprehension, it still appeared to be a statistically significant independent predictor (p = 0.003), with the risk index 1.5. Then, it has an effect, but the effect is small in comparison with the low language comprehension.

One obvious possible factor explaining the low absolute achievement level is the decision regarding *intensified or special support*. In Finland, this decision could be made during preprimary education based on the obvious signs for slow learning. Intensified and special support are given already before school and, in many cases, the need is still there when the school starts. In the dataset, the level of 3-stage support appears to explain significantly (GLM, $F(2, 7767) = 219.14$, $p < 0.001$) and remarkably (f = 0.24) belonging to the group "below A1.1": in the group with a need for intensified support, 20% belonged to the group "below A1.1" and, to the group with a need for special support, 35%. For further analysis, these two groups are combined (Figure 8).

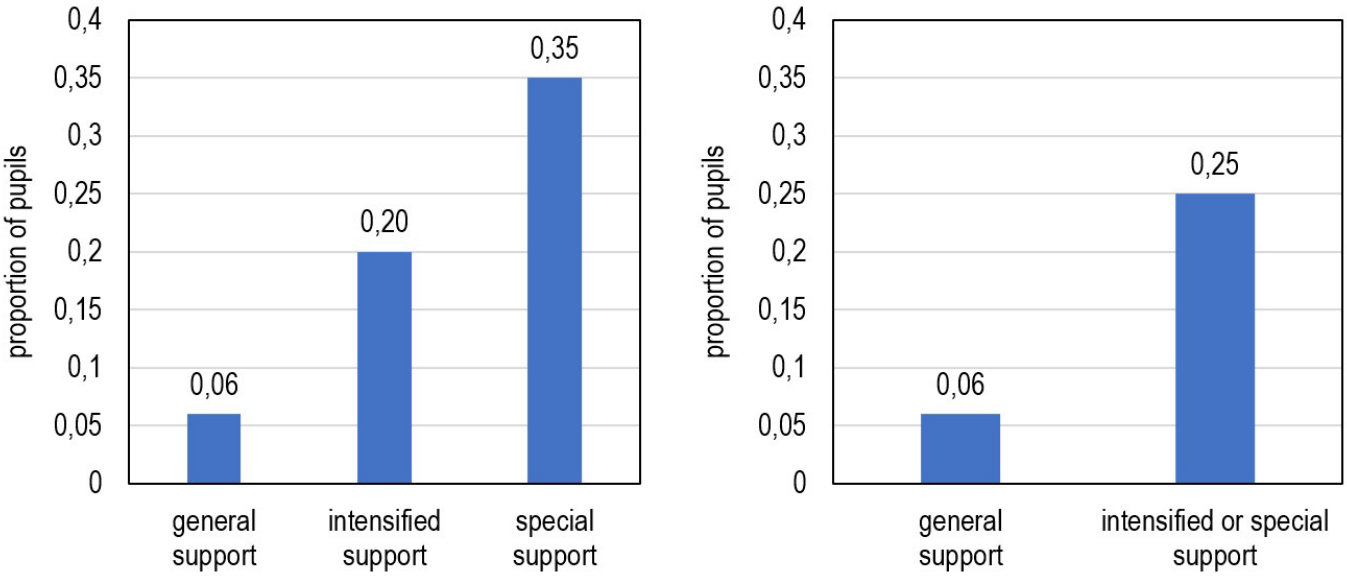

Figure 8. Proportions of pupils belonging to the group "below A1.1" at the levels of the 3-stage support

Attitudes and emotions are shown to be related to performance although the direction of the effect is not easy to determine unambiguously. Namely, we do not know whether high proficiency level creates a positive attitude, is it the other way around, or is it reciprocal. For grade 1 pupils, the attitude items related to mathematics were notably simple: "*Counting is…*" and "*I can count*" and "*The tasks were easy*" with smile faces with 5-point Likert type of scale anchored to 1–5 (see Metsämuuronen & Ukkola, 2019). Still, "counting" may be rather concrete for many young children: they might think, for example, listing numbers in numerical order.

Of the dimensions of attitude (the whole test, general attitude toward math and language, attitude toward mathematics, attitude toward language, and self-efficacy), the *attitude toward mathematics* appeared to be the factor explaining the best belonging to the group "below A1.1" in DTA. The lower is the mean in the attitude scale, the higher the probability to be found in the group "below A1.1". Division of the attitude scale into four groups explains the belonging statistically significantly (GLM, $F(3, 7457) = 53{,}62$, $p < 0.001$) although not remarkably ($f = 0.15$). Again, we do not know how much the result is related to a factual realistic understanding of the child: "I really was not able to solve the tasks, hence, the tasks were not easy to me". If so, it shows that the children even at the grade 1 seem to be able to do realistic evaluation of their capabilities. For the later use, the attitude variable is dichotomized from lower and higher than the score 3.333 as suggested by DTA.

Of other personal factors discussed in the introduction, neither *sex*, more *relative school starting age* nor any of the *hobbies*—not even programming or reading in home (see the variables in Metsämuuronen & Ukkola, 2019)—did explain the belonging to the group "below A1.1" in this dataset. This may be partly explained by the fact that the information concerning the latter variables were provided by the guardians, and this reduced the number of pupils belonging to the group "below A1.1" from $n = 608$ to $n = 246$ pupils.

5.3 Family factors characterizing the pupils with very low preconditions in mathematics at the beginning of the first grade

Two relevant sets of variables related to parents and guardians in explaining the pupils to belong to the lowest achieving group are discussed here: guardians' educations on the one hand and potentially inherited disabilities from the parents on the other. This information is obtained from the guardians' questionnaire and reduces the number of pupils to almost a half and in group "below A1.1" from $n = 608$ to $n =$

246.

The educational background of two guardians ("mother or other guardian" and "father or other guardian") was asked being at one of the five levels: basic education, vocational education and training, matriculation examination, polytechnic, and university degree. Few guardians also selected the alternative "other"—this alternative seems to be a more typical selection by guardians from immigrated families. From this information, several combinations of educational background were derived to explain the general level of preconditions in mathematics (see Ukkola et al., 2020). Here, the original variables are used.

Mother's education appeared to be a better predictor ($\chi^2(2) = 41.82$, $p < 0.001$) than that of father's ($\chi^2(2) = 29.37$, $p < 0.001$). Although DTA suggests three groups for mother education (basic, secondary, and tertiary education), after the correction in p-values by using Šidák's procedure, GLM suggests that only the group of mothers with just *basic education* (or "missing") differs from the other groups (post hoc, $p = < 0.001$; for the other groups, $p > 0.05$). Of the children with this background, 16.8% belonged to the group "below A1.1" while in two other groups the percentage is 5–6%. Effect size is small though ($f = 0.095$).

Another interesting family-related factor explaining the belonging to the lowest-achieving group is the possibly inherited learning disabilities. Five different types of disabilities were given as alternatives in the guardians' questionnaire: linguistic (such as dyslexia), mathematical (such as dyscalculia), concentration, perception, and social challenges. Of these, linguistic and mathematics disabilities did not explain the lowest performance although they may, in general, influence performance. However, pupils belonging to the group "below A1.1" were slightly more likely to have *parents with concentration problems* (23.6% vs. 11.4%, $\chi^2(1) = 31.99$, $p < 0.001$). In LRA, it shown 2.4 times risk although with low explaining power ($R^2_{Adj} = 0.017$). In GLM ($F(1, 4314) = 32.21$, $p < 0.001$), the effect size is small ($f = 0.08$).

5.4 Outline of the results

By combining the results from Sections 5.1–5.3, we may conclude that, of the variables used in the analysis, *deficiencies in language*—specifically a low level of understanding of spoken language, also indicating a lack of adequate vocabulary related to mathematics—is the most powerful factor explaining why the precondition level on mathematics in pupils remained lower than the lowest measurable level (below A1.1 in all criteria). This variable alone explains 41% of the variance in the dataset. In what follows, all variables from the previous sections showing statistically significant prediction power are collected and dichotomized to make a simpler model in LRA. The variables in the modelling included the *score below 372 in the test of listening comprehension, L2 status, decision on intensified or special support, attitude toward mathematics, mother's education in three categories,* and *concentration problems of parents*.

During the modeling, mother's education did not have a major effect and it was dropped in the statistical process. Also, using learning disabilities reduces the dataset to almost half which reduced the explaining power of the models. Hence, in the final model only four variables found from pupils' dataset were kept. The outcome is summarized in Table 5.

Table 5. Four factors explaining the low level of preconditions in mathematics

Variables in the model[1]	**B**	**Significance**	**Risk to belong to the group < A1.1 EXP(B)**
Constant	-3.877		
Score in the test of Listening comprehension (1 = score ≤ 372.19. 0 = score > 372,19	3.351	< 0.001	28.53
3-stage support (1 = decision on intensive or special support, 0 = general support meant for all pupils)	0.777	< 0.001	2.176
L2 status (1 = registered L2 status, 0 = no L2 status)	0.461	0.003	1.586
Attitude toward mathematics (1 = mean score < 3.333, 0 = mean score > 3.333 in the scale of 1–5)	0.457	< 0.001	1.579
Explaining power R^2_{Adj}	0.40		

1) Variables are ordered by the risk

After knowing the score of the listening comprehension, the 3-stage support still gives 2.2 times risk, while L2 status and a low score in a simple test of attitudes toward mathematics have 1.6 times risk to belong to the group "below A1.1" in both mathematical concepts and procedures as well as in abstractions and thinking. The explaining power of the model is reasonably high ($R^2_{Adj} = 0.40$). Notably, low score in the test of listening comprehension, alone, had even higher explaining power ($R^2_{Adj} = 0.41$) with a higher risk value of 44.2. The reason for the non-intuitional higher explaining power by a smaller model is that other variables include missing values causing reduction in pupils included in the analysis.

6 Discussion

An obvious conclusion of the analysis is that not all variables explaining the low level of general preconditions in mathematics and language (Ukkola et al., 2020; see Table 1) are valid in explaining the *absolute* low performance in mathematics. A somewhat surprising although understandable finding is that low proficiency in listening comprehension appears to be a key factor in explaining the low proficiency in mathematics too. It is understandable that if the language skills are underdeveloped in general, the pupil is not expected to master the rarer, specific mathematical vocabulary either, hence the low score in mathematics too. These two phenomena, proficiency in mathematics and proficiency in listening comprehension are not totally independent in the dataset though; part of the items in the mathematics test were used also as part of listening comprehension; after all, nearly all mathematics items also included a component of understanding concepts of mathematics and all instructions were given orally. Hence, further studies of independent tests of mathematics and language would be beneficial. Anyhow, we may predict that all activities increasing the language skills, specifically, of the wider vocabulary in the early childhood may also increase mathematical comprehension. The specific vocabulary related to mathematics may need some conscious concentration from guardians and preprimary teachers.

An obvious limitation of the study is that relevant pieces of information concerning the child was collected from guardians and this information was given only for around half of the pupils and of these, more likely, for better performing children. Hence, with relevant variables explaining pupils belonging to the group “below A1.1”, the number of pupils was reduced from n = 608 to n = 246. In future phases of the longitudinal setting, it is aimed to collect more information from those families that did not answer the questionnaire in the first phase. Hence, the results reported in this article may get more power, specifically, when it comes to parents’ and guardians’ role in the early development of the child.

Teachers in the primary education are facing an interesting challenge at the beginning of the school: how to raise the standard of those who are at the lowest level in mathematics and, at the same time, to keep the lessons interesting also for those advanced pupils who may not learn anything new during the two first years. Earlier studies (Metsämuuronen, 2013a; 2017a; Metsämuuronen & Tuohilampi, 2014) indicate that the schooling and supporting system in Finland can turn the wide distribution of performances at grade 1 into a normal (at grade 3) and even a kurtic normal (at grade 6) after which it, again, widens at grade 9 and even more at the end of secondary education when the national distribution is even wider than what is at the beginning of grade 1.

One part of the challenge is, how to prevent pupils with very low (or even average) level of proficiency in mathematics to fall into an abyss of mathematic anxiety, low self-esteem in mathematics, and underachieving in the studies during the basic education. It may be valuable to try to detect those pupils who have real challenges related to dyscalculia or parallel learning disability related to mathematics. Maybe, at some point, some kind of numeracy screening tests such as functional numeracy assessment (Funa) test (see Funa consortium, 2019; Räsänen et al., 2021) could be used in an early phase, and relevant scaffolding techniques and teaching methods could be developed to help these children during the first stages of development of mathematical concepts, procedures, as well as abstraction and thinking—maybe even before the grade 1.

References

Aunola, K., Leskinen, E., Nurmi, J.-E. (2006). Developmental dynamics between mathematical performance, task motivation, and teachers' goals during the transition to primary school. *British Journal of Educational Psychology, 76*(1), 21–40. https://doi.org/10.1348/000709905X51608

Bandura, A. (2012). *On the functional properties of perceived self-efficacy revisited. Journal of Management, 38(*1), 9–44. https://doi.org/10.1177/0149206311410606

Baye, A., & Monseur, C. (2016). Gender differences in variability and extreme scores in an international context. *Large-Scale Assessments in Education, 4*(4), 1–16. https://doi.org/10.1186%2Fs40536-015-0015-x

Cohen, J. (1988). *Statistical power analysis for the behavioural sciences* (2nd ed). Lawrence Erlbaum.

Dhuey, E., Figlio, D., Karbownik, K., & Roth, J. (2019). School starting age and cognitive development. *Journal of Policy Analysis and Management*, *38*(3), 538–578. https://doi.org/10.1002/pam.22135

Dilnot, J., Hamilton, L., Maughan, B., & Snowling, M. J. (2016). Child and environmental risk factors predicting readiness for learning in children at high risk of dyslexia. *Development and Psychopathology*, *29*(1), 235–244. http://dx.doi.org/10.1017/S0954579416000134

EDUFI (2003). *National core curriculum for upper secondary schools 2003*. Finnish National Agency for Education. [In Finnish]

EDUFI (2004). *National core curriculum for basic education 2004*. Finnish National Agency for Education. [In Finnish]

EDUFI (2014). *National core curriculum 2014*. Määräykset ja ohjeet 2014:96. Finnish National Agency for Education. [In Finnish]

EDUFI (2015). *National core curriculum for upper secondary Schools 2015*. Määräykset ja ohjeet 2015:48. Finnish National Agency for Education. [In Finnish]

EDUFI (2016). *National core curriculum for preprimary education 2014*. Määräykset ja ohjeet 2016:1. Finnish National Agency for Education. [In Finnish]

Eklund, K. (2017). *School-aged reading skills of children with family history of dyslexia: predictors, development and outcome*. Jyväskylä Studies in Education, Psychology and Social Research 574. University of Jyväskylä.

Erola, J., Jalonen, S., & Lehti, H. (2016). Parental education, class and income over early life course and children's achievement. *Research in Social Stratification and Mobility, 44*, 33–43. https://doi.org/10.1016/j.rssm.2016.01.003

Funa consortium (2019). FUNA - An online test battery for functional numeracy assessment. http://oppimisanalytiikka.fi/funa

Johnson, W., Carothers, A., & Deary, I. J. (2008). Sex differences in variability in general intelligence: A new look at the old question. *Perspectives on Psychological Science, 3*(6), 518-531. http://dx.doi.org/10.1111/j.1745-6924.2008.00096.x

Kass, G. (1980). An exploratory technique for investigating large quantities of categorical data. *Applied Statistics, 29*(2), 119–127. https://doi.org/10.2307/2986296

Kivinen, A. (2018). *The effect of relative school starting age on having an individualized curriculum in Finland*. VATT Working Papers 104. Valtion taloudellinen tutkimuskeskus. https://www.doria.fi/handle/10024/149387

Kivinen, O., Hedman, J., & Kaipainen, P. (2012). Koulutusmahdollisuuksien yhdenvertaisuus Suomessa. [Equality of the educational opportunities in Finland.] *Yhteiskuntapolitiikka, 77*(5), 559–566. https://www.julkari.fi/bitstream/handle/10024/103027/kivinen.pdf

Kuukka, K., & Metsämuuronen, J. (2016). *Perusopetuksen päättövaiheen suomi toisena kielenä* (L2) -oppimäärän oppimistulosten arviointi 2015. [Assessment of the learning outcomes of the L2 Finnish 2015.] Publications 2016:13. Finnish Education Evaluation Centre (FINEEC). http://karvi.fi/app/uploads/2016/05/KARVI_1316.pdf

Lerkkanen, M.-K., Poikkeus, A.-M., Ahonen, T., Siekkinen, M., Niemi, P., & Nurmi, J.-E. (2010). Luku- ja kirjoitustaidon kehitys sekä motivaatio esi- ja alkuopetusvuosina. [Development of reading- and writing skills, and motivation during preprimary and primary education] *Kasvatus, 41*(2), 116–128. https://www.researchgate.net/publication/285029301_Luku-_ja_kirjoitustaidon_seka_motivaation_kehitys_esi-_ja_alkuopetusvuosina

Lerkkanen, M.-K., Kiuru, N., Pakarinen, E., Viljaranta, J., Poikkeus, A.-M., Rasku-Puttonen, H., Siekkinen, M., & Nurmi, J.-E. (2012). The role of teaching practices in the development of children's interest in reading and mathematics in kindergarten. *Contemporary Educational Psychology*, *37*(4), 266–279. https://doi.org/10.1016/j.cedpsych.2011.03.004

Machin, S., & Pekkarinen, T. (2008). Global sex differences in test score variability. *Science*, *322*(5906), 1331–1332. http://dx.doi.org/10.1126/science.1162573

Malanchini, M., Rimfeld, K., Wang, Z., Petrill, S. A., Tucker-Drob, E. M., Plomin, R., & Kovas, Y. (2020). Genetic factors underlie the association between anxiety, attitudes and performance in mathematics. *Translational Psychiatry*, *10*(12). https://doi.org/10.1038/s41398-020-0711-3

Metsämuuronen, J. (2009). *Metodit arvioinnin apuna*. [*Methods assisting the assessment*.] Oppimistulosten arviointi 1/2009. Finnish National Agency for Education. [In Finnish].

Metsämuuronen, J. (2010). Osaamisen ja asenteiden muuttuminen perusopetuksen ensimmäisten vuosien aikana. [Development of achievement and attitudes during the first years in basic education] In E. K. Niemi & J. Metsämuuronen (Eds.), *Miten matematiikan taidot kehittyvät? Matematiikan oppimistulokset peruskoulun viidennen vuosiluokan jälkeen vuonna 2008*. [How are the learning outcomes developing? Learning outcomes in mathematics after the grade 5 in basic education in year 2008?] Koulutuksen seurantaraportit 2010:2. Finnish National Agency for Education. https://karvi.fi/app/uploads/2014/09/OPH_0410.pdf [In Finnish]

Metsämuuronen J (2013a). Matemaattisen osaamisen muutos perusopetuksen luokilla 3-9. [Change in the mathematical achievement in grades 3–9 in the compulsory education] In J. Metsämuuronen (Ed.), *Perusopetuksen matematiikan oppimistulosten pitkittäisarviointi vuosina 2005–2012*. [Longitudinal assessment of learning outcomes in mathematics during years 2005–2012] Koulutuksen seurantaraportit 2013:4. Finnish National Agency for Education. http://karvi.fi/app/uploads/2014/09/OPH-0113.pdf [In Finnish]

Metsämuuronen, J. (2013b). A new method to setting standard for the wide range of language proficiency levels. *International Education Research, 1*(1). https://www.researchgate.net/publication/266908493_A_New_Method_to_Setting_Standard_for_the_Wide_Range_of_Language_Proficiency_Levels

Metsämuuronen, J. (2017a). *Oppia ikä kaikki. Matemaattinen osaaminen toisen asteen koulutuksen lopussa 2015*. [Learning all my days. Mathematical achievement at the end of secondary education 2015.] Publications 1:2017. Finnish Education Evaluation Centre (FINEEC). https://karvi.fi/app/uploads/2017/03/KARVI_0117.pdf [In Finnish]

Metsämuuronen, J. (2017b). *Essentials of research methods in human sciences. Vol 2: Multivariate analysis*. SAGE Publications, Inc.

Metsämuuronen, J. (2017c). *Essentials of research methods in human sciences. Vol 3: Advanced analysis*. SAGE Publications, Inc.

Metsämuuronen J (2018). Common framework for mathematics. Discussions of possibilities to develop a set of general standards for assessing proficiency in mathematics. *International Electronic Journal of Mathematics Education, 13*(2), 13–39. https://doi.org/10.12973/iejme/2693

Metsämuuronen, J. (2020). Somers' *D* as an alternative for the item–test and item–rest correlation coefficients in the educational measurement settings. *International Journal of Educational Measurement, 6*(1), 207–221. https://doi.org/10.12973/ijem.6.1.207

Metsämuuronen, J. (2021). Directional nature of Goodman–Kruskal gamma and some consequences. Identity of Goodman–Kruskal gamma and Somers delta, and their connection to Jonckheere–Terpstra test statistic. *Behaviormetrika, 48*(2), 283–307. http://dx.doi.org/10.1007/s41237-021-00138-8

Metsämuuronen, J. (2022a). Deflation-corrected estimators of reliability. *Frontiers in Psychology, 12*:748672. http://dx.doi.org/10.3389/fpsyg.2021.748672

Metsämuuronen, J. (2022b). Effect of various simultaneous sources of mechanical error in the estimators of correlation causing deflation in reliability. Seeking the best options of correlation for deflation-corrected reliability. *Behaviormetrika, 49*(1), 91–130. https://doi.org/10.1007/s41237-022-00158-y

Metsämuuronen, J. & Tuohilampi, L. (2014). Changes in Achievement in and Attitude toward Mathematics of the Finnish Children from Grade 0 to 9. A Longitudinal Study. *Journal of Educational and Developmental Psychology, 4*(2), 145–169. http://www.ccsenet.org/journal/index.php/jedp/article/view/36185

Metsämuuronen, J. & Ukkola, A. (2019). *Alkumittauksen menetelmällisiä ratkaisuja*. [Methodological solutions in baseline measurement.] Publications 18:2019. Finnish Education Evaluation Centre (FINEEC). https://karvi.fi/app/uploads/2019/08/KARVI_1819.pdf [In Finnish]

Nagelkerke, N. J. D. (1991). A note on a general definition of the coefficient of determination. *Biometrika, 78*(3), 691–692. http://dx.doi.org/10.1093/biomet/78.3.691

Niemi, L., Metsämuuronen, J., Hannula, M., & Laine, A. (2020). Matematiikan parhaaksi osaajaksi kehittyminen perusopetuksen aikana. [To develop to the best achiever in mathematics during the basic education.] *Ainedidaktiikka*, *4*(1), 2–33. https://doi.org/10.23988/ad.83384

Niemi, L., Metsämuuronen, J., Hannula, M., & Laine, A. (2021). Matematiikan parhaat osaajat toisen asteen lopussa ja heidän matematiikka-asenteissaan tapahtuneet muutokset. [The best achievers in mathematics at the end of secondary education and changes in attitudes toward mathematics] *LUMAT: International Journal on Math, Science and Technology Education, 9*(1), 457–494. https://doi.org/10.31129/LUMAT.9.1.1511

O'Dea, R. E., Lagisz, M., Jennions, M. D., & Nakagawa, S. (2018). Gender differences in individual variation in academic grades fail to fit expected patterns for STEM. *Nature communications*, *9*(1), 3777. https://doi.org/10.1038/s41467-018-06292-0

OECD. (2015). *Education at a Glance 2015*: *OECD indicators*. OECD Publishing. http://www.oecd.org/education/education-at-a-glance-2015.htm

OECD. (2016). *Technical Report of the Survey of Adult Skills (PIAAC)*. Second Edition. OECD. http://www.oecd.org/skills/piaac/PIAAC_Technical_Report_2nd_Edition_Full_Report.pdf

Paju, P. (2020). *Best childhood money can buy. Child poverty, inequality and living with scarcity, the Finnish case*. Lastensuojelun Keskusliitto.

Palomäki, S., Laherto, L., Kukkonen, T., Hakonen, H., & Tammelin, T. (2016). Vanhempien hyvä koulutus- ja tulotaso on yhteydessä nuorten liikkumiseen etenkin urheiluseuroissa. [Parents' high level of education and salary are connected to the sports activity of adolecents specifically through sports clubs.] *Liikunta & Tiede, 53*(4), 92–98. https://www.lts.fi/media/lts_vertaisarvioidut_tutkimusartikkelit/2016/lt416_tutkimusartikkelit_pa-lomaki_lowres_nimeton-liite-00004.pdf

PISA (2019). *PISA 2018 Results (Volume I): What students know and can do*. PISA, OECD Publishing. https://doi.org/10.1787/5f07c754-en

Räsänen, P., Aunio, P., Laine, A., Hakkarainen, A., Väisänen, E., Finell, J., Rajala, T., Laakso, M.-J., & Korhonen, J. (2021). Effects of gender on basic numerical and arithmetic skills: Pilot data from 3rd to 9th grade for a large-scale online dyscalculia screener. *Frontiers in Education, 6*. https://doi.org/10.3389/feduc.2021.683672

Šidák, Z. (1967). Rectangular confidence region for the means of multivariate normal distributions. *Journal of the American Statistical Association, 62*(318), 626–633. http://dx.doi.org/10.1080/01621459.1967.10482935

Sirniö, O. (2016). *Constrained life chances: Intergenerational transmission of income in Finland*. Publications of the Faculty of Social Sciences 29. Sociology. University of Helsinki. http://urn.fi/URN:ISBN:978-951-51-1098-5

TIMSS (2020). Mullis, I. V. S., Martin, M. O., Foy, P., Kelly, D. L., & Fishbein, B. (2020). *TIMSS 2019 International results in mathematics and science*. https://timssandpirls.bc.edu/timsL2019/international-results/wp-content/themes/timssandpirls/download-center/TIMSS-2019-International-Results-in-Mathematics-and-Science.pdf

Tuohilampi, L. & Hannula, M. S. (2013). Matematiikkaan liittyvien asenteiden kehitys sekä asenteiden ja osaamisen välinen vuorovaikutus 3., 6. ja 9. luokalla. [Changes of attitudes related to matematics and their interaction at grades 3, 6, and 9.] In J. Metsämuuronen (Ed.), *Perusopetuksen matematiikan oppimistulosten pitkittäisarviointi vuosina 2005–2012* [Longitudinal assessment of learning outcomes in mathematics during years 2005–2012.] (pp. 231–252). Koulutuksen seurantaraportit 2013:4. Finnish National Agency for Education. http://karvi.fi/app/uploads/2014/09/OPH-0113.pdf [In Finnish]

Ukkola, A. & Metsämuuronen, J. (2019). *Alkumittaus – Matematiikan ja äidinkielen ja kirjallisuuden osaaminen ensimmäisen luokan alussa*. [Zero-level measurement. Proficiency of mathematics and mother language and literature at the beginning of grade 1.] Publications

17:2019. Finnish Education Evaluation Centre (FINEEC). https://karvi.fi/app/uploads/2019/07/KARVI_1719.pdf [In Finnish]

Ukkola, A. & Metsämuuronen, J. (2021). Matematiikan ja äidinkielen ja kirjallisuuden osaaminen kolmannen luokan alussa. [Achievement in mathematics and mother language and literature at grade 3.] Publications 20:2021. Finnish Education Evaluation Centre (FINEEC). https://karvi.fi/wp-content/uploads/2021/08/KARVI_2021.pdf

Ukkola, A., Metsämuuronen, J. & Paananen, M. (2020). *Alkumittauksen syventäviä kysymyksiä*. [Deepening questions of zero-level measurement] Publications 10:2020. Finnish Education Evaluation Centre (FINEEC). https://karvi.fi/app/uploads/2020/08/KARVI_Alkumittaus.pdf [In Finnish]

Van der Schoot, F. (2009). Cito variation of the Bookmark method. Reference supplement, Section I. In S. Takala (Ed.), *Relating language examinations to the Common European Framework of Reference for Languages: Learning, teaching, assessment (CEFR). A Manual*. Language Policy Division, Strasbourg. https://rm.coe.int/1680667a24