Discover Education

Research

Mathematical skills classification through primary education

Giannis Karagiannakis[1] · Marie-Pascale Noël[2] · Anna Baccaglini-Frank[3] · Cristiano Termine[4]

Received: 24 August 2023 / Accepted: 17 September 2024
Published online: 10 October 2024

Abstract
By the end of primary school, children are expected to acquire a range of mathematical skills that progressively develop. This study aimed to gain insight into how a large number of numerical and geometrical measures are grouped and whether the structures shift or remain invariant along child's development based on the data obtained from a sample of 1346 s to sixth grade children. On the basis of correlation analyses and exploratory factor analyses, we came up with an invariant four-factor structure for each grade. The four factors obtained were the following: (a) early and retrieval skills (subitizing, enumeration, number facts) (b) transcoding and ordinality skills, (c) numerical advanced skills and (d) visual-spatial advanced skills. Confirmatory factor analyses showed that the four-factor model fits well to the data (RMSEA < 0.06, CFI > 0.95, GFI > 0.93, SRMR < 0.05) at all grades. The associations between the mathematical measures were captured in each factor and the revealed invariant structure is discussed and compared with previous classification models of mathematical skills. Given that the difficulties in learning mathematics is currently being viewed as a continuum of academic abilities instead of a distinct problem, the underlying invariant four-factor structure can facilitate experts and educators to better understand how a broad area of mathematical skills are related across the primary education, in order to carry our comprehensive assessment of both the mathematical strengths and weaknesses of their students and to apply the appropriate customized teaching strategies.

Keywords Mathematical skills · Arithmetical skills geometrical skills · Primary education · Structural model · Mathematical learning difficulties · Dyscalculia

1 Introduction

By the end of primary school, children are expected to acquire a range of mathematical skills, especially in the domains of arithmetic and geometry (e.g., [1–3]). Many studies have focused on children's numerical development, and especially, on their skills in the domain of natural numbers. Although mathematical learning formally starts at the beginning of the primary school (at 6 years of age in many countries, including Italy where the current study was carried out), it starts from infancy and it is constructed progressively, so that skills acquired earlier, even informally, become foundational for new more complex ones (e.g., [4, 5]). In this study, we aim to gain insight into how a broad range of numerical and visual-spatial measures that assess mathematical skills from early to advanced, are grouped in children from grades 2 to 6, as well as whether these groups change or remain invariant in these years of important mathematical development.

✉ Giannis Karagiannakis, g.karagiannakis@primedu.uoa.gr; Marie-Pascale Noël, marie-pascale.noel@uclouvain.be; Anna Baccaglini-Frank, anna.baccaglinifrank@unipi.it; Cristiano Termine, cristiano.termine@uninsubria.it | [1]Department of Psychology, University of Athens, Athens, Greece. [2]Department of Psychology, Université Catholique de Louvain in Louvain-La-Neuve, Ottignies-Louvain-la-Neuve, Belgium. [3]Department of Mathematics, University of Pisa, Pisa, Italy. [4]Department of Medicine and Technological Innovation, University of Insubria, Varese, Italy.

2 Development of mathematical skills

Infant studies have shown that children are already equipped with two pre-verbal systems for numerical quantification: the approximate number system (ANS or 'analogue magnitude'), and the object tracking system (OTS or 'parallel individuation'). The ANS is a system that supports the estimation of the numerosity of a set of elements without relying on language or symbols [6, 7]. The OTS is a pre-verbal system that enables the parallel individuation and tracking of objects as they move through space (e.g., [8]). The OTS, is considered to be the basis for *subitzing*, the fast and accurate assessment of a small number of dots (range 1 to 4 dots) ([9]. The ANS is typically assessed by determining the larger numerosity between two arrays of images (e.g. dots) that are presented very briefly on a computer screen. There is some positive evidence of a relationship between ANS's precision and math achievement (see the meta-analysis of Schneider et al. [21]) although some studies failed to observe such relations [10, 11].

Similarly, knowledge of geometry is believed to be founded on at least two distinct, core cognitive systems [12]: the first is used to represent the shapes of large-scale navigable surface layouts and the second represents small-scale movable forms and objects. The aforementioned numerical and geometrical early skills is indicative of the view of innate number sense (INS) as an innate sense, which is related to visual and auditory perception [13].

Building on these innate skills or parallel to those, new ones are developed, for example, preschool children learn the number-words sequence, and, through the counting principles [14], how to apply them when counting collections of objects. These skills are often measured by asking the children to determine how many items are in a collection, which refers to the enumeration process. When collections are very small (around 4 items or less) item by item enumeration is usually not necessary as another process takes place: the subitizing, which allows to determine precisely and very rapidly, the cardinal of the collection. This process is based on the OTS system mentioned just above. Some studies have found that subitizing and enumeration skills are significant precursors of later mathematical achievement [15, 16].

Once formally engaged in the primary school children consolidate their knowledge of the Arabic symbols including the complex rules of the positional system that allow them to combine digits to represent larger numerosities; they learn how to transcode numbers from one format to another (analogue, verbal, Arabic, see [17]), and how to compare [18] and order numbers [19]. Tasks like single- or multidigit-numbers comparison, numbers dictation, find the next and the previous number are typically used to assess these skills. The ability to understand, represent and compare Arabic numbers is considered a significant indicator of numerical competence [19–21].

Moreover, they start learning to add and subtract numbers in an increasingly sophisticated way. For example, to find the sum of 3 and 5, first a child utilizes the *count all* strategy: s/he first counts from "one" up to "three" usually on the fingers of one hand, then counts from "one" to "five" on the fingers of the other hand and then counts all of them together (e.g., 1,2,3,4,5,6,7,8) to find the sum. Then s/he is expected to move to the *count on* strategy, counting from "three", and 5 more (e.g., [3–8]) [22]. This shift requires full understanding of ordinality ([23]. Finally, s/he refines the strategy by starting to count on from the greater number, "five" (e.g., [6–8]). The latter strategy presumes that the child has already consolidated symbolic numerical magnitude in order to select the larger number to start with. Alternatively, s/he decomposes the addends, regrouping numbers as s/he feels more comfortable (e.g., $3+5=3+3+2=6+2$ etc.). Being able to compute the sum $3+5$ by counting on from "five", or by decomposing the addends, are skills that depend upon full understanding of the cardinal value and of the early symbolic quantitative knowledge that seems to partially depend upon consolidation of innate skills (e.g., [24, 25]). At this stage, they are about to develop the early number sense (ENS), in contrast to INS, includes learned skills that involve explicit number knowledge [13].

As pupils consolidate these basic math skills, they become progressively automatized with lighter cognitive costs. This, in turn, allows more advanced skills to be built upon them. Eventually, when presented with a sum like $3+5$, the child will retrieve the sum from long term memory, leaving resources to deal with more complex calculations (such as $43+25$). A similar process will also be observed for other arithmetical operations, such as multiplications and divisions. By the end of the primary school, children are also expected to be able to solve numerical problems (for ex. numerical patterns) and arithmetical word problems [26] and are introduced to geometry. For instance, they are required to find the correct position of a number on number lines. The accuracy on number lines tasks is associated with higher mathematical achievement [27, 28]. Moreover, children are requested to reason about shapes on the plane (2D) and in space (3D), classify different kinds of figures and solids based on sets of shared geometrical properties, and calculate perimeters and areas of a variety of 2D figures [29]. The most advanced of these skills are components of the mature number sense (MNS) which continue to develop in later school levels [13, 30].

It should be noted that mathematical learning does not occur independently of general cognitive and executive functions skills. Memory, inhibition, logical reasoning, visual-spatial abilities, flexible thinking and fluid intelligence and so on interact with and support the development of the mathematical skills described above [31–34]. Of course, a variety of non-cognitive factors, such as a student's attitude towards mathematics, mathematics anxiety, pedagogical choices and methodologies, and many other socio-cultural factors [35, 36] also contribute to enhancing or hindering the development of a student's mathematical skills, but we will not discuss these, as they are beyond the scope of this study.

2.1 Associations and classification of mathematical skills

Several studies have used various methodologies to explore associations between sets of skills assumed to be necessary in learning of mathematics. The associations may be outlined with respect to students' general mathematical achievement or to their achievement within a particular mathematical domain, with the intent to identify subgroups of students with different sets of strengths and weaknesses (e.g., [37]). An attempt to capture the different components supporting early stages of mathematical learning is represented by the working model of Aunio and Räsänen [38] which is based on the analysis of various test batteries used to assess children's numerical skills, suggesting a categorization of abilities supporting numerical development between 5 and 8 years of age. In particular, numerical skills were categorized into four main types: (1) symbolic and non-symbolic number sense; (2) understanding mathematical relations (early mathematical-logical principles, arithmetic principles, mathematical operational symbols, place-value and base-ten system); (3) counting skills (knowledge of number-symbols, number word-sequence, enumeration with concrete objects); and (4) basic skills in arithmetic (arithmetic combinations, addition and subtraction skills with number symbols). To our knowledge, no empirical studies were conducted to confirm this model.

In typical populations, to our knowledge, only one study has tapped into the structure of the mathematical skills of primary school children. This study aimed to empirically validate the fourfold model of Karagiannakis, Baccaglini-Frank and Papadatos ([39], that argues that mathematical learning difficulties (MLD) could arise from difficulties from core number, memory, visuo-spatial, or reasoning domains. This fourfold model provides a theoretical basis for exploring both the heterogeneity of MLD and, more in general, students' cognitive strengths and weaknesses with respect to their mathematical learning. A first study attempting to empirically validate this model has been carried out on a sample of 165 typical 5th and 6th grade students using exploratory and confirmatory factor analyses on the performance on a computerized battery involving a wide range of mathematical tasks. The results indicated strong evidence for supporting the solidity of the model and clustered the population into six distinguishable performance groups with the MLD and low achievement (LA) students distributed within five of the clusters (for more see [40]).

There have also been various attempts to understand pattens of skill weakness, which could also give insight into relationships between areas of skills. A number of studies on MLD suggest that the population exhibiting such a condition is quite heterogeneous. A consistent and specific MLD profile wasn't found in a recent meta-analysis which examined the differences in mathematical and cognitive skills between individuals with and without MLD [41]. Indeed, there seems to be growing consensus that MLD depends on the lack of various combinations of skills such as those described above [37, 42, 43]. Most of these studies consider both domain general and domain specific abilities involved in mathematical learning in order to explore the variable profiles associated with low achievers. The purpose of this approach is manifold: to identify "subtypes" of MLD based on tasks that are found to be particularly difficult for groups of students, to make distinctions within the population of students with MLD, and to show the heterogeneity of this population (e.g., [42, 44–48]).

For instance, Geary [49] proposed to distinguish three subtypes of MLD: one characterized by difficulties in the representation or retrieval of arithmetic facts from semantic memory, the second by problems in the execution of arithmetical procedures and the third one by problems in the visuospatial representation of numerical information. Yet, this proposition was not supported by empirical research.

Some other studies tested both number-specific and domain-general skills. For instance, Bartelet, Ansari, Vaessen, and Blomert [42] reported six subtypes of difficulties: (1) weak mental number line, (2) weak ANS, (3) spatial difficulties, (4) access deficit, (5) no numerical cognitive deficit, and (6) garden variety of both numerical and general cognitive deficits. Another recent study of such type was conducted by Chan and Wong [50] on 76 children showing math learning difficulties in both first and second grade testing. They discovered 5 distinctive subtypes among: (1) the number sense deficit subtype (with severe deficit in the ANS), (2) the numerosity coding deficit subtype (weak in enumeration), (3) the symbolic deficit subtype (with deficits in both enumeration and number line and weaknesses in working memory), (4) the working memory subtype (with deficits in working memory), and (5) the mild difficulty group (with average performance in all the cognitive domains assessed). These subtypes showed moderate stability between grade 1 and 2.

Assessing number skills in primary schoolchildren with MD, von Aster [51] identified three subtypes of MDs: the Arabic subtype (with deficits in number transcoding and number comparison), the verbal subtype (with deficits in counting and difficulties in subtraction), and the pervasive subtype (with deficits across most measures in the study).

As can be seen, previous studies have failed to produce clear and consistent results, possibly because of limited measure selection, confusion between different skill levels, and small sample sizes.

As of today, no strong empirical evidence has yet been provided that any of the classification model is in fact suitable to describe associations between key mathematical skills during students' development in their first phase of exposure to formal mathematical instruction. In fact, current gaps in the literature can be looked at from two perspectives. On the one hand, standardized achievement assessments consider a range of mathematical skills, but they do not analyze students' performance based on possible different underlying cognitive components upon which the tasks rely. On the other hand, many psychological studies have focused on a specific skill (or on a small set of them) without studying broader networks through which many other skills are likely to associate.

We have found it very rare to consider simultaneously a broad range of mathematical skills and explore their mutual associations. This gap in the literature is specifically the one we aim to bridge with a line of research that, in the long term, aims at developing theoretical and practical tools to identify students' mathematical strengths and weaknesses in order to propose more effective teaching interventions in case of mathematical learning difficulties, which is extremely important in mathematics education (e.g., [40, 52–56]. In order to accomplish this, many smaller steps need to be taken. In this study, we explore the associations between a broad range of numerical and geometrical measures of students from grade 2 to grade 6, and how these associations shift or remain invariant in these years of important mathematical development.

3 Materials and methods

3.1 Participants

Data were collected from 1489 children in grades 2–6 from 21 primary public schools in Italy. Participants came from a variety of regions and socio-economic statues (SES). In particular, the schools were located in the north (about 65%), center (about 22%) and south (about 13%) of Italy. Overall, about 16% were in urban setting with high SES; 72% in urban or suburban settings with middle SES and 12% in urban or suburban settings with middle-to-low SES. For each grade, outliers were removed by checking whether a child's score on the task Numbers typing (see description below) was < 2 SDs. It is a very simple task that was used to control the individual's reaction time. A low score on accuracy indicates difficulty on using the computer or low motivation to proceed. Children who didn't complete all the tasks were also removed. This led to removing 143 children (9.6%) from the data set resulting in a final sample of N = 1346 (see Table 1 for a breakdown of the participants by grade and gender). Grades did not differ in the number of boys and girls $X^2(4, N = 1346) = 4.82, p = 0.31$.

Table 1 Descriptive statistics of the sample

Grade	Age (years)			Sex		Sum
	M	SD	Range	Boys	Girls	
2	7.83	0.30	7.2–8.3	103	115	218
3	8.85	0.35	8.0–10.6	145	143	288
4	9.83	0.38	8.8–1.8	130	166	296
5	10.79	B 0.33	10.1–12.4	141	137	278
6	11.84	0.35	10.9–13.3	123	143	266
Sum				642	704	N = 1346

3.2 Procedure

The research was approved by the ethics committee of the University of Milano-Bicocca in agreement with the Helsinki ethics declaration (World Health Organization, 2001). Schools and parents received information about the research purpose and procedure. All students provided written informed consent to participate in this study. Students were tested in groups in computer classrooms of their schools where they use to do their computing lessons. They were informed that they could interrupt the test any time they wanted without any penalties. There was no time limit for the session. The data collection took place during the spring period within a period of four weeks. Data were totally anonymous; they were saved in files using an identification code and no personal information regarding the children were associated with it.

The mathematical measures administered were from the MathPro Test [57] which includes one subtest measuring the processing speed and 18 subtests assessing each a very specific mathematical skill. Some of these subtests are related to the core number domain (such as the ability to compare the magnitude of two dots collections, the subitizing skills, or the ability to enumerate dot arrays). Others are tapping into the processing of number symbolic codes (such as the ability to compare the magnitude of two Arabic numbers, to write numbers to dictation, to write the number just preceding or just following a given number). Others are assessing arithmetical skills, using both single-digit additions or multiplications (possibly using arithmetical facts retrieval from memory) and more complex mental calculations. Other subtests measure the ability to understand the properties of the mathematical operations (e.g., commutativity) and the relationships that the operations entertain with one another; the ability to find the appropriate calculations that would be needed to solve a word problem or to find the logic that lies behind a series of numbers. Finally, some other subtests require visual-spatial processing such as positioning numbers on a line (going from 0 to 100 or 0 to 1000) or reasoning on 2D or 3D shapes.

The subtests of the MathPro Test were translated for the purpose of this study in Italian by native speakers experts both in numerical cognition and mathematics education. The Italian version of the MathPro test includes one more control subtest, the Computer mouse speed to control the processing speed of the numbers comparison subtests. For an in-depth description of the procedure and tasks, see [55]. Brief descriptions of the subtests are displayed in Table 2.

3.3 Task scoring

The accuracy (number of correct responses divided by the number of items) was calculated for all tasks except for the number line tasks. At these tasks the pupils are asked to estimate the position of a target number on an empty number line where only the start and end point are marked with the respective numbers (for ex., 0 and 100). For this reason the accuracy of the answers is computed as percentage of absolute errors (PAE) with PAE = (|*estimated position—correct position*| / *numerical range of the line*)*100% [27]. On a group of tasks (Computer mouse speed, Single- and Multi-digit numbers comparison, Screen keyboard use, Numbers dictation, Next number, Previous number, Enumeration, Addition, and Multiplication facts retrieval) response times (correct trials only) were calculated. Since, the accuracy of the children in the Single-digit was at ceiling, we used an RT index penalized for inaccuracy as a measure. Performance scores were calculated as a composite of error rates (ER) and response times (RT), using the formula: $P = RT\,(1 + 2ER)$, where a higher value indicates worse performance. Error rates were multiplied by 2 because the particular tasks were constrained by binary responses ($ER = 0.5$ indicates chance).

The MathPro Test for the younger participants (grades 2 and 3) doesn't include all the subtests and, for some subtests, only the first easier items were presented (see Table A Appendix), following the curriculum in math in Italy. For in-depth description of the psychometric properties of the items in Italian version see Baccaglini-Frank, Karagiannakis, Pini, Termine and Girelli [58].

3.4 Data analysis

We first examined the correlations between all the measures separately for each grade as well as the correlations between the RT measures controlled for the general processing speed. Subsequently, we examined the Kaiser–Meyer–Olkin (KMO) index and Bartlett's sphericity test to perform a factorability check. The communality and skewness of all the MathPro tasks were also examined. Given the limited theoretical background regarding the associations of the mathematical measures that are included in this study (for more see 1.2 section), we performed separately in each grade Principal

Table 2 Brief Descriptions of the MathPro subtests

Subtest	Brief description
1.Dots comparison (DotComp)	Determine which of two dot arrays contains more dots
2.Computer mouse speed	Indicate a cat image from a spot image as quickly and as accurately as possible
3.Single-digit numbers comparison (S-NumComp)	Determine which of two symbolic numbers is larger as quickly and as accurately as possible
4.Multi-digit numbers comparison(M-NumComp)	Determine which of two symbolic numbers is larger as quickly and as accurately as possible
5.Screen keyboard use	Indicate a single-digit number presented on the computer screen on a screen keyboard as quickly and as accurately as possible
6.Numbers dictation (NumDict)	Indicate on the screen keyboard numbers presented verbally as quickly and as accurately as possible
7.Next number (NextNum)	Indicate on the screen keyboard the succeeding number of numbers presented verbally as quickly and as accurately as possible
8.Previous number (PrevNum)	Indicate on the screen keyboard the preceding number of numbers presented verbally as quickly and as accurately as possible
9.Subitizing (Subitizing)	Indicate the numerosity of small arrays of dots (1–6) presented briefly
10.Enumeration (Enumeration)	Count number of dots (1–9) on the screen as quickly and as accurately as possible
11.Addition facts retrieval (AddFacts)	Indicate the correct sum ($\leq$ 10) as quickly and as accurately as possible
12.Multiplication facts retrieval (MultFacts)	Indicate the correct product (both factors never were $\geq$ 5) as quickly and as accurately as possible
13.Mental calculations (MentalCalc)	Indicate the correct result of addition, subtraction, multiplication and division operations with operands up to 3-digits
14.Number Lines 0–100 (NL100)	Indicate on a horizontal line 0–100 where a symbolic number should fall
15.Number Lines 0–1000 (NL1000)	Indicate on a horizontal line 0–100 where a symbolic number should fall
16.Squares (Squares)	Indicate the number of whole squares by joining a combination of squares, half-square triangles and quarter-square triangles
17.Building blocks (BuildBlocks)	Indicate the number of cubes of 3D structures presented in 2D on the screen
18.Word problems (WordsProbl)	Indicate the way several word problems (appear written and read aloud) should be solved (for ex. "37 + 25" and not the final solution "62")
19.Calculation principles (CalcPrinc)	Indicate the solution of a multidigit operation without calculating it but with reference to an already solved associated problem
20.Numerical patterns (NumPatterns)	Indicate the missing number on a series of numbers

axis factoring (PAF) by the maximum-likelihood method with oblique (promax) rotation to both examine the factor loadings of the AC measures and obtain an a posteriori grouping of those with efficient factor loading. Although on the one hand the factors were expected to be independent, on the other the tasks were expected to be correlated since all assess mathematical skills. For this reason, the promax rotation was chosen because it is a nice compromise between orthogonal and oblique rotation. Promax first assumes the factors are orthogonal and then relaxes the rotation to allow them to correlate [59].

Next, we conducted a series of confirmatory factor analysis (CFA) separately for each grade to compare the goodness-of-fit between the one-factor model for a strict unidimensional assessment model and several multidimensional models. The multidimensional models were based on the results of the PAF, the correlation analysis as well as how meaningful they were from theoretical point of view. Starting from the single-factor model, the subsequent model was judged better if its χ^2 value was significantly lower than that for the former model.

Model parameter appraisal used maximum likelihood (ML) estimation. The following criteria were used in evaluating overall goodness of fit for the measurement models: (a) the chi-square/degrees of freedom ratio (χ^2/df) for which a value less than 2.0 indicates a good fit; (b) the robust Comparative Fit Index (CFI); (c) the Goodness of Fit Index (GFI); (e) the Root Mean-Square Error of Approximation (RMSEA) with 90% confidence intervals; and (f) the Standardized Root Mean-Square Residual (SRMR). These indices take sample size into consideration and specify the amount of covariation in the data, which is accounted for by the hypothesized model each time relative to a null model that assumes independence among variables. For the CFI, where 1.0 indicates a perfect fit, a value in the range of 0.90 is generally accepted as indicating a good fit. For the RMSEA, an adequately fitting model will have a value between 0.00 and 0.06, with 90% CIs between 0.00 and 0.10 [60]. Finally, regarding SRMR, a value less than 0.08 is considered a good fit [61]. We also used the Akaike information criterion (AIC; [62]), where a lower AIC values indicative of superior fit.

The magnitude of the skewness in some MathPro subtests indicates that the data was non-normal, which can result in standard error biases [63]. Accordingly, analysis used ML estimation with bootstrapping (500 resamples) to generate accurate estimations of standard errors with accompanying confidence intervals (bias-corrected at the 95% confidence level) and p-values [64]. The Bollen–Stine bootstrap p was calculated. Bollen–Stine gauges fit without normal theory limitations [65] and $p > 0.05$ suggests excellent global fit. Statistical analyses were performed using IBM SPSS 25 and AMOS 21 [66] separately for each grade.

4 Results

In the first result section, performance for each subtest in each grade is provided as well as analyses testing the change of performance across school grades. Section 2 presents the correlations between performances in each subtest, in each grade. Section 3 presents PAF and CFA for each grade.

4.1 Preliminary analyses

Table 3 shows performance per grade in terms of accuracy (AC) for all subtests of the MathPro Test and reaction time (RT) for the subtests where the response time was measured. We analyzed children's reaction times only when task accuracy was high enough. To this end, for all grades, we excluded from the reaction times analyses the items where accuracy was below 0.85 (i.e., the decimal number items of the Multi-digit numbers comparison subtest, the five-digit numbers of the Numbers dictation subtest and the three-digit numbers of the Next and Previous numbers subtests).

ANOVAs were then run separately for each task, measuring the effects of grade (see Table B in Appendix). A significant grade effect (all ps < 0.001) was found in all the tasks and for all the measures (AC and RT) indicating that, as expected, performance improved over the school grades.

Table 3 Descriptive statistics of MathPro subtests per grade

	Grade 2				Grade 3				Grade 4				Grade 5				Grade 6			
	M	(SD)	Skew	Kurt	M	(SD)	Skew	Kurt	M	(SD)	Skew	Kurt	M	(SD)	Skew	Kurt	M	(SD)	Skew	Kurt
Dots comparison (AC)	0.58	(0.11)	–0.15	–0.21	0.61	(0.10)	0.00	0.09	0.63	(0.09)	–0.27	0.03	0.64	(0.09)	–0.13	0.00	0.65	(0.09)	–0.07	–0.25
Single digit numbers comparison (AC)	0.97	(0.05)	–4.02	25.32	0.98	(0.06)	–8.92	108.89	0.99	(0.03)	–2.74	10.35	0.99	(0.03)	–6.19	63.34	0.98	(0.03)	–6.56	72.01
Single digit numbers comparison (RT)	2593	(794)	–1.28	1.74	2319	(832)	–1.24	2.14	2278	(786)	–0.53	7.13	1979	(787)	–1.25	0.71	2137	(691)	–1.62	1.88
Multi–digit numbers comparison (AC)	0.89	(0.16)	–1.73	3.14	0.89	(0.12)	–2.01	7.35	0.78	(0.14)	–0.28	0.18	0.85	(0.13)	–0.68	0.30	0.87	(0.12)	–0.97	0.91
Multi–digit numbers comparison (RT)	3619	(1308)	0.37	3.32	3115	(1166)	–0.34	0.14	3073	(1126)	–0.12	4.39	2617	(1042)	–1.16	0.57	2806	(912)	–1.42	1.65
Numbers dictation (AC)	0.60	(0.18)	–0.59	0.34	0.78	(0.17)	–1.80	4.19	0.86	(0.17)	–2.29	7.64	0.91	(0.12)	–3.40	16.06	0.90	(0.12)	–3.08	14.43
Numbers dictation (RT)	4798	(1749)	0.95	4.79	4621	(2002)	0.68	3.14	4329	(1556)	–0.18	2.59	3784	(1628)	0.69	4.72	3859	(1363)	–0.17	1.77
Next number (AC)	0.70	(0.24)	–1.02	0.17	0.80	(0.25)	–1.54	1.49	0.87	(0.21)	–2.39	5.18	0.88	(0.19)	–2.60	6.32	0.88	(0.21)	–2.58	6.26
Next number (RT)	5087	(1770)	0.44	0.17	4302	(1936)	0.16	1.45	3896	(1400)	0.24	4.20	3319	(1274)	–0.36	1.15	3395	(1035)	–0.77	1.75
Previous number (AC)	0.73	(0.23)	–1.23	0.17	0.85	(0.21)	–2.08	3.56	0.90	(0.18)	–3.01	8.54	0.93	(0.13)	–3.85	16.42	0.93	(0.14)	–3.94	17.15
Previous number (RT)	5268	(1801)	0.60	0.89	4475	(2698)	9.94	136.85	3903	(1330)	0.24	2.89	3250	(1133)	–0.55	1.53	3305	(957)	–0.44	3.57
Subitizing (AC)	0.72	(0.19)	–1.03	0.98	0.79	(0.15)	–1.32	3.36	0.84	(0.14)	–1.28	2.53	0.86	(0.13)	–1.92	6.79	0.85	(0.15)	–2.03	6.38
Enumeration (AC)	0.85	(0.20)	–2.02	3.84	0.87	(0.17)	–2.35	6.18	0.91	(0.13)	–2.89	12.15	0.92	(0.12)	–2.95	13.31	0.91	(0.14)	–2.86	10.93
Enumeration (RT)	8833	(2590)	0.52	2.09	7382	(2141)	1.06	3.41	6979	(1815)	0.87	3.93	5951	(1545)	–0.28	2.19	5967	(1672)	0.09	3.43
Addition facts retrieval (AC)	0.90	(0.17)	–2.85	8.89	0.94	(0.13)	–3.92	16.92	0.97	(0.09)	–6.15	43.06	0.97	(0.10)	–5.69	38.58	0.95	(0.12)	–4.29	25.31
Addition facts retrieval (RT)	7061	(3910)	4.35	31.71	4946	(2435)	2.75	12.97	4345	(2915)	2.80	13.39	3412	(1085)	0.83	3.18	3390	(1268)	1.46	6.96
Multiplication facts retrieval (AC)	[redacted]				0.84	(0.21)	–1.86	2.58	0.89	(0.17)	–2.59	6.94	0.91	(0.16)	–3.28	11.94	0.88	(0.20)	–2.45	5.78
Multiplication facts retrieval (RT)					8914	(4808)	2.32	6.97	7848	(5457)	3.69	18.37	6004	(3146)	3.26	18.89	3603	(2982)	4.799	42.70
Mental calculations (AC)	0.52	(0.22)	–0.24	–0.63	0.57	(0.26)	–0.44	–0.84	0.66	(0.24)	–0.78	–0.11	0.74	(0.22)	–1.33	1.32	0.69	(0.25)	–1.03	0.23
Number Lines 0–100 (PAE)	11.03	(5.99)	1.38	1.99	7.85	(4.74)	2.79	12.63	6.45	(3.88)	3.29	16.09	5.25	(2.32)	2.11	7.16	5.16	(2.55)	3.01	14.65
Number Lines 0–1000 (PAE)	[redacted]				15.72	(8.23)	0.63	–0.02	11.47	(7.70)	1.43	2.48	8.71	(6.05)	1.56	2.29	8.14	(6.16)	2.25	6.49
Squares (AC)	0.43	0.26	0.24	–0.99	0.52	(0.24)	–0.24	–0.86	0.61	(0.23)	–0.76	–0.02	0.69	(0.21)	–1.17	1.74	0.72	(0.20)	–1.31	2.39
Building blocks (AC)	0.58	(0.26)	–0.29	–0.62	0.64	(0.26)	–0.75	–0.03	0.66	(0.25)	–0.78	0.13	0.73	(0.23)	–0.93	0.42	0.74	(0.24)	–1.09	1.09
Word problems (AC)	0.48	(0.29)	0.01	–0.92	0.42	(0.24)	0.07	–1.22	0.52	(0.25)	–0.52	–1.09	0.57	(0.24)	–0.78	–0.67	0.60	(0.22)	–1.02	–0.06
Calculation principles (AC)	0.35	(0.28)	0.45	–0.74	0.36	(0.29)	0.18	–1.55	0.45	(0.31)	–0.13	–1.57	0.55	(0.30)	–0.06	–0.96	0.52	(0.31)	–0.43	–1.32
Numerical patterns (AC)	0.35	(0.20)	0.05	–0.32	0.39	(0.21)	–0.12	–0.41	0.46	(0.21)	–0.14	–0.19	0.50	(0.21)	–0.04	–0.10	0.48	(0.21)	–0.17	0.06

Reaction time (RT) calculated for the items with the accuracy leading to a ceiling (> 0.8)

4.2 Correlations

Table 4 presents the correlations between the measures for grade 2, Table 5 for grade 3 and 4 and Table 6 for grades 5 and 6. Both the PAE and RT scores were multiplied by – 1 as a high score in these cases actually means a low performance. None of the raw correlations between the AC measures was above 0.80 (the maximum value was 0.70), suggesting that multicolinearity was not a problem in the present data at the measurement level [67]. This was unsurprisingly the case for the correlations between the RT measures between the Single-digit and Multi-digit numbers comparison (correlations of 0.81, 0.88, 0.84, 0.96, 0.96 respectively for grades 2, 3, 4, 5, 6). However, when the general processing speed was controlled for by partialing out the RT measure of the Computer mouse speed subtest, the correlation values were all below 0.8 (0.69, 0.68, 0.77, 0.75, 0.76; $p < 0.001$). First-order correlations were also calculated between the other RT measures controlling for the general processing speed by partialling out the RT measure of the Screen keyboard use subtest. All the first-order correlations values were lower compared with the zero-order values (see Tables C, D, E in Appendix). This underlines the effect of the general processing speed on all the RT measures. The correlations between the RT measures and the AC were low or moderate and not significant in most cases. Therefore, the strong correlations between the RT measures but weak between the RT and the AC measures indicate that the associations between the RT measures might be due more to the reaction time itself than to mathematical skills. This is why in the factor analyses we have used only the AC measures.

4.3 Mathematical skills construct structure evaluation

Dots comparison task wasn't included in the factor analyses because of the poor internal reliability and the low correlation coefficients (< 0.30) with the other measures. This was also the case for the accuracy measure of the single-digit numbers comparison task. When the P measure (RT penalized for inaccuracy) of the Single-digit numbers comparison task was included in the factor analyses, its loading was low (< 0.20) and not significant.

Table 4 Pearson's correlations between the tasks for grades 2

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
1.Dots comparison (AC)	–																									
2.Single digit numbers comparison (AC)	0.29**	–																								
3.Single digit numbers comparison (RT)	0.05	0.02	–																							
4.Multi–digit numbers comparison (AC)	0.19**	0.39**	0.08	–																						
5.Multi–digit numbers comparison (RT)	0.02	0.08	0.81**	0.17*	–																					
6.Numbers dictation (AC)	0.22**	0.26**	0.19**	0.26**	0.30**	–																				
7.Numbers dictation (RT)	0.04	0.05	0.60**	0.02	0.52**	0.05	–																			
8.Next number (AC)	0.22**	0.16*	0.03	0.21**	0.04	0.55**	0.16*	–																		
9.Next number (RT)	0.07	0.15*	0.59**	0.12	0.60**	0.25**	0.63**	0.01	–																	
10.Previous number (AC)	0.22**	0.30**	0.02	0.18**	0.09	0.55**	0.14*	0.66**	0.06	–																
11.Previous number (RT)	0.10	0.13	0.46**	0.25**	0.53**	0.32**	0.43**	0.09	0.63**	0.13	–															
12.Subitizing (AC)	0.13	0.22**	0.02	0.14*	0.16*	0.24**	0.03	0.30**	0.07	0.41**	0.23**	–														
13.Enumeration (AC)	0.08	0.23**	0.02	0.18**	0.06	0.21**	0.06	0.26**	0.05	0.40**	0.10	0.60**	–													
14.Enumeration (RT)	0.08	0.10	0.31**	0.09	0.36**	0.12	0.38**	0.08	0.49**	0.09	0.43**	0.08	0.05	–												
15.Addition facts retrieval (AC)	0.14*	0.34**	0.02	0.25**	0.08	0.22**	0.02	0.23**	0.08	0.44**	0.22**	0.48**	0.70**	0.05	–											
16.Addition facts retrieval (RT)	0.19**	0.32**	0.18**	0.24**	0.32**	0.29**	0.32**	0.14*	0.51**	0.21*	0.49**	0.19**	0.04	0.48**	0.14*	–										
17.Multiplication facts retrieval (AC)																	–									
18.Multiplication facts retrieval (RT)																		–								
19.Mental calculations (AC)	0.19**	0.25**	0.03	0.26**	0.07	0.36**	0.19**	0.31**	0.14*	0.42**	0.19**	0.41**	0.55**	0.03	0.54**	0.28**			–							
20.Number Lines 0–100 (PAE)	0.29**	0.32**	0.01	0.30**	0.17*	0.30**	0.13	0.33**	0.17*	0.39**	0.24**	0.43**	0.40**	0.05	0.44**	0.24**			0.51**	–						
21.Number Lines 0–1000 (PAE)																					–					
22.Squares (AC)	0.21**	0.11	0.10	0.17*	0.12	0.32**	0.09	0.28**	0.15*	0.36**	0.26**	0.37**	0.34**	0.08	0.40**	0.23**			0.50**	0.48**		–				
23.Building blocks (AC)	0.18**	0.17*	0.02	0.18**	0.05	0.28**	0.12	0.28**	0.11	0.33**	0.17*	0.40**	0.38**	0.08	0.34**	0.21**			0.50**	0.47**		0.53**	–			
24.Word problems (AC)	0.19**	0.14*	0.01	0.14*	0.01	0.23**	0.10	0.28**	0.12	0.37**	0.23**	0.35**	0.34**	0.02	0.34**	0.23**			0.46**	0.38**		0.46**	0.43**	–		
25.Calculation principles (AC)	0.22**	0.18**	0.04	0.14*	0.11	0.29**	0.14*	0.32**	0.09	0.37**	0.13	0.29**	0.34**	0.01	0.34**	0.22**			0.49**	0.43**		0.38**	0.47**	0.53**	–	
26.Numerical patterns (AC)	0.16*	0.17*	0.10	0.21**	0.14*	0.27**	0.06	0.27**	0.13	0.35**	0.24**	0.33**	0.37**	0.09	0.42**	0.19**			0.46**	0.43**		0.34**	0.43**	0.49**	0.53**	–

The grid format indicates the tasks aren't given; Correlations between RTs are in blue; r ≥0.03 are in bold.

**p < 0.01

*p < 0.05

Table 5 Pearson's correlations between the tasks for grades 3–4

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
1.Dots comparison (AC)	–	0.00	0.05	0.15*	0.04	0.27**	0.03	0.22**	0.03	0.23**	0.02	0.22**	0.01	0.10	0.03	0.17**	0.15**	0.06	0.15**	0.14*	0.23**	0.23**	0.13*	0.19**	0.08	0.14*
2.Single digit numbers comparison (AC)	0.08	–	0.00	0.45**	0.04	0.01	0.03	0.14*	0.08	0.09	0.03	0.00	0.02	0.01	0.09	0.07	0.09	0.04	0.07	0.02	0.02	0.04	0.02	0.06	0.05	0.06
3.Single digit numbers comparison (RT)	0.12*	0.04	–	0.05	0.88**	0.01	0.74**	0.09	0.73**	0.08	0.30**	0.03	0.02	0.35**	0.03	0.27**	0.15*	0.11	0.09	0.07	0.05	0.05	0.08	0.07	0.06	0.03
4.Multi–digit numbers comparison (AC)	0.21**	0.14*	0.10	–	0.04	0.23**	0.09	0.18**	0.09	0.27**	0.05	0.18**	0.12*	0.09	0.24**	0.29**	0.31**	0.13*	0.30**	0.25**	0.20**	0.10	0.18**	0.19**	0.29**	0.24**
5.Multi–digit numbers comparison (RT)	0.07	0.04	0.84**	0.23**	–	0.04	0.67**	0.11	0.64**	0.06	0.29**	0.06	0.03	0.30**	0.04	0.28**	0.11	0.11	0.08	0.06	0.07	0.00	0.05	0.03	0.04	0.05
6.Numbers dictation (AC)	0.26**	0.16**	0.08	0.32**	0.16*	–	0.20**	0.56**	0.18**	0.58**	0.31**	0.33**	0.27**	0.16**	0.30**	0.37**	0.28**	0.23**	0.30**	0.25**	0.28**	0.23**	0.21**	0.22**	0.20**	0.16**
7.Numbers dictation (RT)	0.09	0.02	0.67**	0.05	0.63**	0.00	–	0.10	0.74**	0.10	0.72**	0.02	0.08	0.30**	0.04	0.27**	0.08	0.22**	0.12*	0.02	0.00	0.05	0.07	0.09	0.00	0.07
8.Next number (AC)	0.12*	0.09	0.07	0.20**	0.14*	39**	0.11	–	0.09	0.56**	0.25**	0.31**	0.28**	0.17**	0.23**	0.25**	0.23**	0.16**	0.24**	0.19**	0.22**	0.13*	0.06	0.11	0.13*	0.17**
9.Next number (RT)	0.10	0.02	0.67**	0.11	0.65**	0.12*	0.77**	0.05	–	0.24**	0.26**	0.12	0.19**	0.44**	0.17**	0.39**	0.24**	0.29**	0.22**	0.10	0.11	0.13*	0.12	0.14**	0.08	0.13*
10.Previous number (AC)	0.20**	0.12*	0.02	0.17**	0.09	0.38**	0.02	0.43**	0.10	–	0.28**	0.43**	0.36**	0.24**	0.47**	0.37**	0.47**	0.27**	0.35**	0.33**	0.31**	0.25**	0.18**	0.22**	0.17**	0.22**
11.Previous number (RT)	0.06	0.02	0.61**	0.18**	0.67**	0.16**	0.74**	0.22**	0.76**	0.13*	–	0.10	0.11	0.19**	0.02	0.17**	0.12*	0.21**	0.06	0.09	0.12*	0.10	0.04	0.12	0.05	0.02
12.Subitizing (AC)	0.12*	0.03	0.08	0.11	0.13*	0.20**	0.07	0.15**	0.11	0.31**	0.18**	–	0.41**	0.21**	0.53**	0.29**	0.43**	0.14*	0.38**	0.37**	0.33**	0.33**	0.27**	0.25**	0.25**	0.23**
13.Enumeration (AC)	0.17**	0.07	0.05	0.11	0.11	0.16*	0.03	0.23**	0.05	0.33**	0.13*	0.38**	–	0.02	0.48**	0.27**	0.47**	0.14*	0.41*	0.29**	0.24**	0.30**	0.13*	0.26**	0.23**	0.21**
14.Enumeration (RT)	0.20**	0.03	0.41**	0.17**	0.47**	0.15**	0.48**	0.22**	0.49**	0.20**	0.58**	0.30**	0.18**	–	0.17**	0.51**	0.23**	0.42**	0.27**	0.18**	0.23**	0.22**	0.12*	0.17**	0.12	0.18**
15.Addition facts retrieval (AC)	0.20**	0.11	0.00	0.25**	0.13*	0.31**	0.03	0.30**	0.00	0.48**	0.08	0.36**	0.59**	0.19**	–	0.29**	0.62**	0.19**	0.47**	0.42**	0.22**	0.26**	0.23**	0.28**	0.25**	0.25**
16.Addition facts retrieval (RT)	0.12*	0.05	0.26**	0.36**	0.49**	0.27**	0.21**	0.29**	0.28**	0.33**	0.48**	0.35**	0.39**	0.49**	0.65**	–	0.43**	0.60**	0.53**	0.42**	0.46**	0.31**	0.18**	0.35**	0.29**	0.30**
17.Multiplication facts retrieval (AC)	0.25**	0.26**	0.05	0.25**	17*	0.41**	0.06	0.36**	0.17**	0.41**	0.19**	0.31**	0.42**	0.20**	0.54**	0.45**	–	0.37**	0.71**	0.52**	0.48**	0.41**	0.29**	0.44**	0.41**	0.43**
18.Multiplication facts retrieval (RT)	0.06	0.15**	0.20*	0.21**	0.32**	0.19**	0.30**	0.18**	0.41**	0.17**	0.48**	0.18**	0.04	0.46**	0.33**	0.70**	0.45**	–	0.35**	0.17**	0.27**	0.19**	0.09	0.26**	0.19**	0.17**
19.Mental calculations (AC)	0.25**	0.24**	0.11	0.37**	0.23**	0.40**	0.16**	0.33**	0.24**	0.31**	0.29**	0.31**	0.31**	0.33**	0.32**	0.43**	0.68**	0.48**	–	0.52**	0.55**	0.48**	0.34**	0.56**	0.61**	0.56**
20.Number Lines 0–100 (PAE)	0.29**	0.11	0.20**	0.36**	0.34**	0.36**	0.10	0.26**	0.19**	0.30**	0.28**	0.25**	0.34**	0.25**	0.42**	0.55**	0.41**	0.28**	0.44**	–	0.63**	0.40**	0.31**	0.45**	0.42**	0.42**
21.Number Lines 0–1000 (PAE)	0.28**	0.07	0.17**	0.38**	0.31**	0.32**	0.09	0.23**	0.22**	0.30**	0.20**	0.20**	0.28**	0.25**	0.37**	0.39**	0.40**	0.31**	0.50**	0.71**	–	0.48**	0.34**	0.52**	0.46**	0.47**
22.Squares (AC)	0.25**	0.04	0.05	0.26**	0.15*	0.20**	0.05	0.18**	0.10	0.14*	0.14*	0.22**	0.29**	0.19**	0.32**	0.28**	0.31**	0.12*	0.38**	0.53**	0.55**	–	0.42**	0.50**	0.36**	0.39**
23.Building blocks (AC)	0.21**	0.08	0.06	0.23**	0.13**	0.23**	0.05	0.19**	0.04	0.25**	0.14*	0.19**	0.20**	0.17**	0.22**	0.17**	0.22**	0.11	0.30**	0.40**	0.48**	0.46**	–	0.45**	0.25**	0.34**
24.Word problems (AC)	0.29**	0.20**	0.16**	0.39**	0.21**	0.37**	0.20**	0.28**	0.28**	0.29**	0.31**	0.20**	0.22**	0.33**	0.24**	0.37**	0.46**	0.41**	0.67**	0.46**	0.60**	0.51**	0.39**	–	0.55**	0.52**
25.Calculation principles (AC)	0.29**	0.03	0.10	0.34**	0.17**	0.29**	0.13*	0.25**	0.23**	0.26**	0.22**	0.23**	0.21**	0.29**	0.19**	0.30**	0.42**	0.34**	0.65**	0.37**	0.46**	0.35**	0.31**	65**	–	0.56**
26.Numerical patterns (AC)	0.22**	0.15**0	0.11	0.32**	0.23**	0.29**	0.12*	0.30**	0.18**	0.24**	0.22**	0.23**	0.29**	0.32**	0.28**	0.33**	0.45**	0.33**	0.61**	0.40**	0.52**	0.42**	0.40**	0.62**	0.63**	–

Up right: correlations for grade 3; Bottom left: correlations for grade 4. Correlations between RTs are in blue; r ≥0.03 are in bold

**p < 0.01

*p < 0.05

Table 6 Pearson's correlations between the tasks for grades 5–6

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
1.Dots comparison (AC)	–	0.06	0.01	0.17**	0.03	0.09	0.08	0.08	0.12*	0.07	0.07	0.08	0.05	0.09	0.04	0.10	0.12	0.08	0.20**	0.13*	0.22**	0.16**	0.17**	0.16**	0.16**	0.22**
2.Single digit numbers comparison (AC)	0.15*	–	0.02	0.34**	0.03	0.11	0.03	0.03	0.05	0.05	0.06	0.06	0.08	0.06	0.08	0.08	0.12*	0.04	0.15*	0.17**	0.18**	0.12*	0.13*	0.10	0.03	0.11
3.Single digit numbers comparison (RT)	–0.08	0.02	–	0.02	0.96**	0.04	0.64**	0.05	0.77**	0.04	0.74**	0.11	0.03	0.41**	0.06	0.28**	0.15*	0.16**	0.17**	0.01	0.04	0.04	0.16**	0.21**	0.11	0.110.
4.Multi–digit numbers comparison (AC)	0.17*	0.25**	0.08	–	0.09	0.33**	0.00	0.13*	0.09	0.26**	0.13*	0.19**	0.22**	0.09	0.24**	0.31**	0.25**	0.25**	0.37**	0.49**	0.53**	0.39**	0.32**	0.39**	0.36**	0.43**
5.Multi–digit numbers comparison (RT)	0.10	0.05	0.96**	0.14*	–	0.01	0.60*	0.02	0.76**	0.01	0.74**	0.08	0.06	0.43**	0.00	28**	0.10	0.18*	0.09	0.06	0.03	0.01	0.15*	0.16*	0.07	0.05
6.Numbers dictation (AC)	0.16**	0.16*	0.10	0.31**	0.12	–	0.17**	0.41**	0.06	0.57**	0.08	0.34**	0.37**	0.07	0.33**	0.22**	0.35**	0.16**	0.35**	0.28**	0.27**	0.22**	0.09	0.34**	0.28**	0.28**
7.Numbers dictation (RT)	0.09	0.01	0.79**	0.05	0.76**	0.22**	–	0.02	0.77**	0.01	0.75**	0.05	0.01	0.48**	0.09	0.38**	0.09	0.27**	0.13*	0.08	0.08	0.09	0.09	0.04	0.01	0.04
8.Next number (AC)	0.05	0.03	0.05	0.20**	0.09	0.23**	0.13*	–	0.04	0.57**	0.03	0.29**	0.21**	0.07	0.24**	0.11	0.24**	0.08	0.22**	0.21**	0.16**	0.24**	0.12*	0.24**	0.27**	0.19**
9.Next number (RT)	0.09	0.00	0.20**	0.10	0.70**	0.25**	0.84**	0.06	–	0.09	0.88**	0.04	0.03	0.54**	0.05	0.42**	0.06	0.28**	0.09	0.00	0.02	0.00	0.10	0.12	0.05	0.02
10.Previous number (AC)	0.16*	0.12	0.03	0.19**	0.02	0.42**	0.11	0.46**	0.10	–	0.11	0.39**	0.30**	0.10	0.33**	0.22**	0.30**	0.13*	0.32**	0.28**	0.24*	0.32**	0.12*	0.39**	0.30**	0.24**
11.Previous number (RT)	0.17**	0.00	0.75**	0.16*	0.73**	0.21**	0.79**	0.21**	0.80**	0.23**	–	0.09	0.09	0.56**	0.03	0.47**	0.00	0.34**	0.01	0.09	0.04	0.08	0.04	0.02	0.06	0.05
12.Subitizing (AC)	0.18**	0.27**	0.14*	0.28**	0.17*	0.24**	0.10	0.23**	0.18**	0.37**	0.24**	–	0.32**	0.06	0.36**	0.14*	0.32**	0.08	0.33**	0.18**	0.24**	0.27**	0.11	0.19**	0.29**	0.25**
13.Enumeration (AC)	0.16**	0.35**	0.06	0.23**	0.10	0.19**	0.01	0.18**	0.04	0.32**	0.03	0.37**	–	0.08	0.40**	0.26**	0.36**	0.11	0.47**	0.25**	0.19**	0.23**	0.15*	0.25**	0.38**	0.31**
14.Enumeration (RT)	0.08	0.15*	0.52**	0.18**	0.56**	0.22**	0.56**	0.22**	0.61**	0.09	0.66**	0.25**	0.08	–	0.03	0.67**	0.06	0.46**	0.12*	0.13*	0.08	0.22**	0.09	0.11	0.10	0.09
15.Addition facts retrieval (AC)	0.16*	0.19	0.11	0.22**	0.13*	0.34**	0.05	0.29**	0.13*	0.47**	0.20**	0.43**	0.50**	0.13*	–	0.20**	0.44**	0.02	0.40**	0.33**	0.34**	0.29**	0.11	0.23**	0.27**	0.22**
16.Addition facts retrieval (RT)	0.04	0.06	0.46**	0.28**	0.49**	0.27**	0.55**	0.23**	0.58**	0.19**	0.18**	0.22**	0.03	0.70**	0.15*	–	0.28**	0.59**	0.37**	0.31**	0.30**	0.32**	0.21**	0.29**	0.35**	0.30**
17.Multiplication facts retrieval (AC)	0.10	0.10	0.09	0.33**	0.12	0.30**	0.06	0.23**	0.13*	0.38**	0.21**	0.34**	0.36**	0.13*	0.48*	0.37**	–	0.39**	0.72**	0.29**	0.36**	0.31**	0.25**	0.45**	0.48**	0.44**
18.Multiplication facts retrieval (RT)	0.10	0.01	0.24**	0.16*	0.22**	0.27**	0.33**	0.15*	0.36**	0.32**	0.39**	0.24**	0.11	0.45**	0.16**	0.57**	0.37**	–	0.34**	0.23**	0.17**	0.18**	0.14*	0.29**	0.32**	0.29**
19.Mental calculations (AC)	0.18**	0.26**	0.14*	0.44**	0.19**	0.40*	0.15*	0.21**	20**	0.43**	0.29**	0.37**	0.36**	0.25*	0.38**	0.47**	0.70**	0.50**	–	0.36**	0.48**	0.42**	0.32**	0.58**	0.63**	0.61**
20.Number Lines 0–100 (PAE)	0.19**	0.12	0.20**	0.45**	0.24**	0.38**	0.15*	0.23**	0.19**	0.33**	0.29**	0.42**	0.18**	0.24**	0.32**	0.37**	0.37**	0.29**	0.49**	–	0.58**	0.47**	0.30**	0.40**	0.42**	0.34**
21.Number Lines 0–1000 (PAE)	0.25**	0.14*	0.21**	0.44**	0.23**	0.39**	0.18**	0.18**	0.23**	0.34**	0.30**	0.28**	0.25**	0.28**	0.31**	0.43**	0.38**	0.30**	0.53**	0.58**	–	0.49**	0.38**	0.46**	0.49**	0.49**
22.Squares (AC)	0.30**	0.26**	0.13*	0.30**	0.15*	0.22**	0.09	0.16*	0.10	0.32**	0.22**	0.36**	0.24**	0.22**	0.27**	0.22**	0.35**	0.25**	0.49**	0.38**	0.48**	–	0.42**	0.43**	0.46**	0.41**
23.Building blocks (AC)	0.28**	0.21**	0.17**	0.32**	0.20*	0.19**	0.16**	0.11	0.19**	0.25**	0.22**	0.28**	0.23**	0.24**	0.14*	0.31**	0.26**	0.27**	0.41**	0.37**	0.42**	0.50**	–	0.38**	0.39**	0.38**
24.Word problems (AC)	0.15*	0.08	0.20**	0.45**	0.20*	0.30**	0.16**	0.17**	0.22**	0.32**	0.32**	0.31**	0.14*	0.21**	0.31**	0.38**	0.45**	0.33**	0.60*	0.45**	0.50**	0.44**	0.46**	–	0.62**	0.53**
25.Calculation principles (AC)	0.26**	0.24**	0.07	0.38**	0.10	0.25**	0.12	0.23**	0.18**	0.30**	0.24**	0.27**	0.19**	0.21**	24**	0.30**	0.36**	0.27**	0.58**	0.38**	0.39**	0.37**	0.39**	0.61**	–	0.60**
26.Numerical patterns (AC)	0.23**	0.19**	0.10	0.39**	0.13*	0.31**	0.14*	0.23**	0.17**	0.32**	0.21**	0.28**	27**	0.23**	0.31**	0.36**	0.43**	0.32**	0.62**	0.43**	0.43**	0.44**	44**	0.59**	0.65**	–

Up right: correlations for grade 5; Bottom left: correlations for grade 6.

**p < 0.01.

*p < 0.05; RT measures are in blue; r ≥0.03 are in bold.

4.4 Grade 2

Due to non-normality of some measures (see Table 3), analysis used ML estimation with bootstrapping (500 resamples) to generate accurate estimations of standard errors. The Bollen–Stine bootstrap $p=0.573$ suggests excellent global fit of the data. The Kaiser–Meyer–Olkin measure (KMO = 0.90) verified the sampling adequacy for the analysis, and Bartlett's test of sphericity [$\chi(91)=1250.47$, $p<0.001$] indicated that correlations between items were sufficiently large for PAF [68]. An initial analysis was run to obtain eigenvalues for each component in the data. Four factors emerged and explained in total 59.9% of the variance. The first factor (A) was loaded on Subitizing, Enumeration and Addition facts. The second factor (B) was loaded on Number dictation, Next and Previous numbers. The third factor (C) was loaded on Mental calculations, Word problems, Calculation Principles, Numerical patterns, Numbers comparison. Number lines 0–100, Squares and Building blocks loaded on the last factor (D).

Between the single-factor and the four-factor model, we adopted the four-factor model because the CFA showed that the four-factor model had a significant decrease in χ^2 value, lower (better) AIC (Table 7) and good overall fit $\chi^2/df=1.49$, CFI = 0.971, GFI = 0.935, SRMR = 0.046, RMSEA = 0.048, and AIC = 174.075. The structure of the four-factor model for grade 2 is illustrated in Fig. 1.

4.5 Grade 3–6

Due to non-normality of some measures (see Table 3), analysis used ML estimation with bootstrapping (500 resamples) to generate accurate estimations of standard errors. The PAF run separately for each grade, identified the same four-dimensional structure for grades 3–6. The Kaiser–Meyer–Olkin measure for grade 3,4,5,6 (KMO = 0.90) verified the sampling adequacy for the analysis, and Bartlett's test of sphericity [$\chi(120)=1990.52$, 2098.99, 1838.65, 1761.99 $p<0.001$, respectively]. An initial analysis was run to obtain eigenvalues for each component in the data. Four factors emerged and explained in total 57.97%, 64.25%, 57.07% and 56.11% of the variance for grades 3,4,5

Table 7 Summary of fit indices for factor models per grade

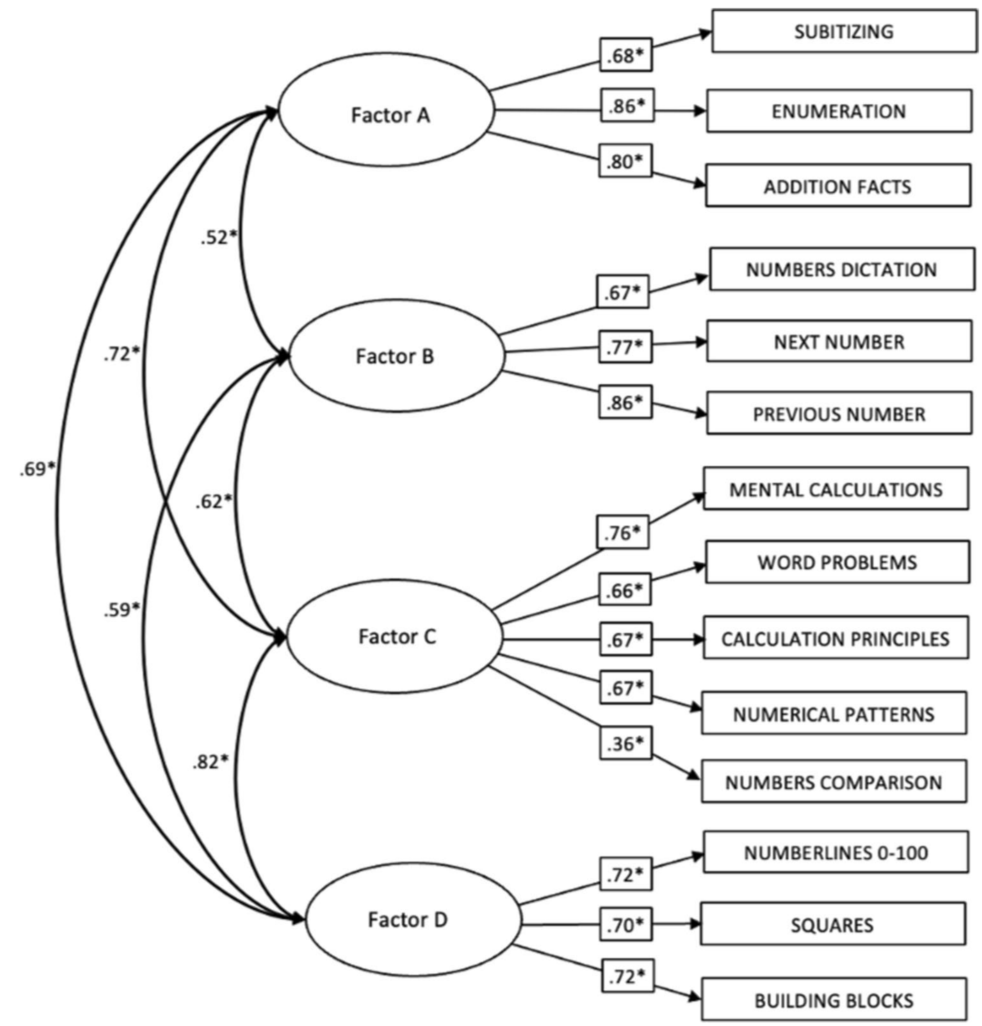
Model	χ^2/df	$\Delta\chi^2$ (df)	CFI	GFI	SRMR	RMSEA (90% CI)	AIC
Grade 2							
One-factor	311.640/77 = 4.05	REF	0.804	0.816	0.078	0.119 (0.105–0.133)	367.640
Four-factor	106.075/71 = **1.49**	**142.242*****(1)	**0.971**	**0.935**	**0.046**	**0.048** (0.027–0.066)	**174.075**
Grade 3							
One-factor	605.058/104 = 5.82	REF	0.738	0.756	0.095	0.130 (0.120–0.140)	701.058
Four-factor	274.699/98 = 2.80	129.304*** (1)	0.908	0.895	0.059	0.079 (0.068–0.090)	350.699
Four-factor CE	185.182/96 = **1.93**	**89.517***** (2)	**0.952**	**0.925**	**0.053**	**0.058** (0.046–0.070)	**264.430**
Grade 4							
One-factor	617.811/104 = 5.94	REF	0.746	0.755	0.088	0.129 (0.120–0.139)	681.811
Four-factor	284.851/98 = 2.91	29.581*** (1)	0.908	0.897	0.060	0.080 (0.070–0.091)	360.851
Four-factor CE	171.281/96 = **1.78**	**113.570***** (2)	**0.962**	**0.934**	**0.046**	**0.052 (**0.039–0.065)	**253.281**
Grade 5							
One-factor	520.493/104 = 5.01	REF	0.763	0.779	0.087	0.120 (0.110–0.131)	584.493
Four-factor	284.197/98 = 2.90	75.757*** (1)	0.894	0.887	0.060	0.083 (0.072–0.094)	360.197
Four-factor CE	177.511/96 = **1.85**	**106.686***** (2)	**0.953**	**0.925**	**0.049**	**0.057** (0.044–0.069)	**261.511**
Grade 6							
One-factor	431.133/104 = 4.15	REF	0.806	0.815	0.076	0.109 (0.098–0.120)	495.133
Four-factor	277.676/98 = 2.83	19.184*** (1)	0.893	0.891	0.061	0.083 (0.072–0.095)	353.676
Four-factor CE	172.729/96 = **1.80**	**104.947***** (2)	**0.953**	**0.926**	**0.054**	**0.056** (0.043–0.069)	**256.729**

The best indices are in bold

CFI = Comparative Fit Index; GFI = Goodness of Fit Index; SRMR = Standardized Root Mean-square Residual; RMSEA = Root Mean-square Error of Approximation; CI = Confidence Intervals; AIC = Akaike Information Criterion; CE = Correlated Errors

***$\Delta\chi^2$ significant at $p<0.001$

Fig. 1 Four-factor model for grade 2. Latent variables are represented by ellipses; measured variables are represented by rectangles



and 6 respectively. The first factor (A) was loaded on the same three subtests (Subitizing, Enumeration, Addition facts) as those of factor A of grade 2 along with the Multiplication facts. The second factor (B) was loaded on exactly the same subtests (Number dictation, Next and Previous numbers) as those of factor B of grade 2. The third factor (C) was also loaded on the same subtests (Mental calculations, Word problems, Calculation principles, Numerical patterns and Numbers comparison) as those of factor C of grade 2. The last factor (D) was loaded on the same three subtests (Number lines 0–100, Squares and Building blocks) as those of factor D of grade 2 along with the Number lines 0–1000 subtest which wasn't given to grade 2 children.

Between the single-factor and the four-factor model for each of the upper grades (3–6), the CFA showed that the four-factor model had a significant decrease in χ^2 value as well as the lowest AIC (Table 7). Moreover, the four-factor solution possessed incidences of high modification indices (MI) between Numbers dictation and Next number in grade 3, Enumeration and Addition facts in grade 4, Number line 0–100 and Number line 0–1000 in grade 5 and 6. The covariances errors indicate similar conceptual content, since it presents a unique variance origin [69]. Allowing errors to covary significantly improved model fit for each grade (Table 7). The Bollen–Stine bootstrap p = 0.008, 0.010, 0.062, 0.036 (for grade 3,4,5,6 respectively) suggests also excellent global fit of the data. Therefore, the four-factor model with correlated errors was adopted as a final model for grades 3–6 (see Fig. 2).

The analysis of the Standardized Residual Covariance Matrix was also considered. The majority of residual covariances for the adopted model in all grades (1–6) were less than two indicating good fit [70]. Moreover, consultation of factor loadings revealed that indicators (MathPro subtests) were positive, possessed a majority of high loadings (i.e., above 0.6) along with few of moderate loadings (between 0.4 and 0.6), and were significant ($p < 0.05$) with lower 95% CI all above 0.5, suggesting all indicators loaded meaningfully [71].

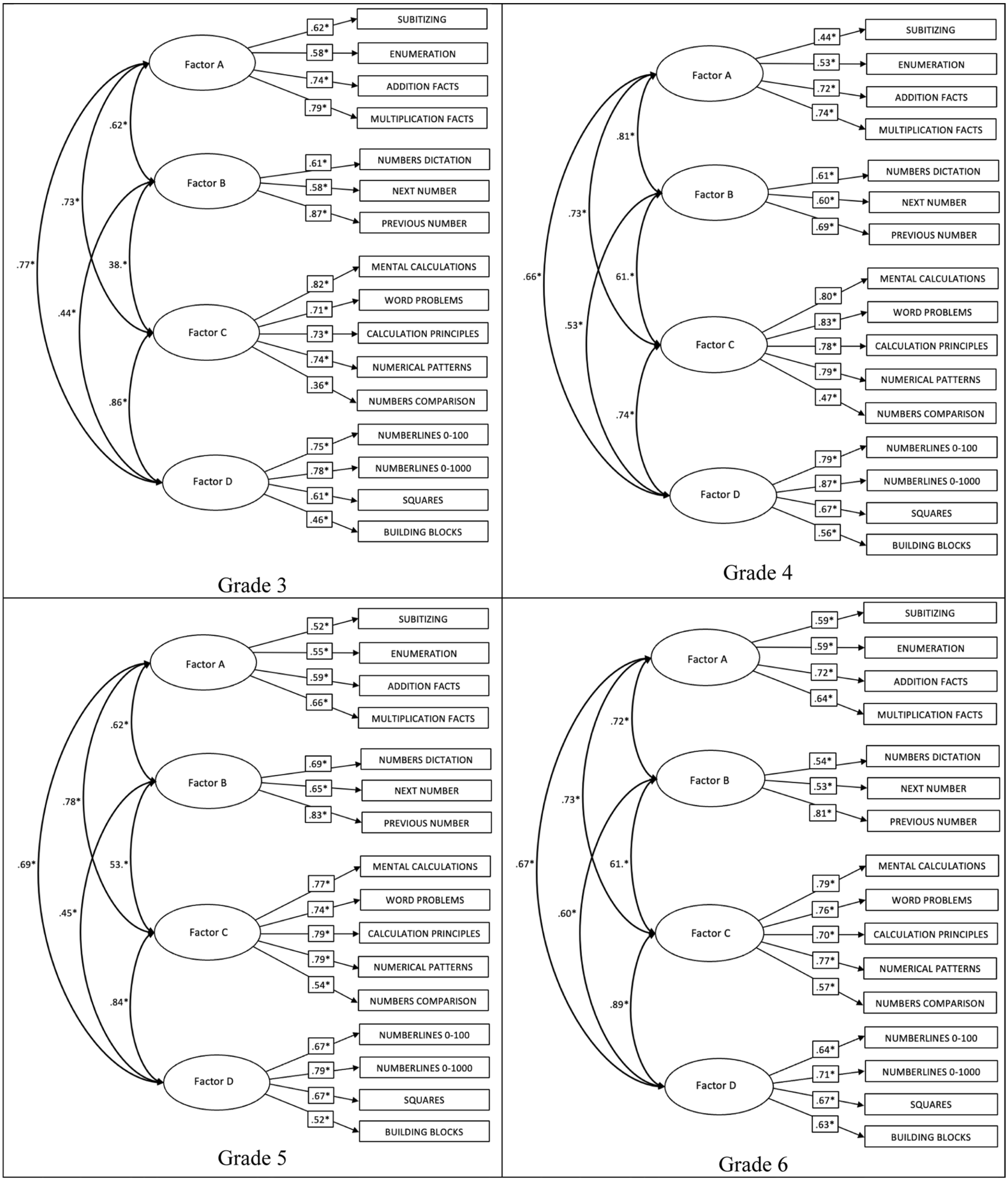


Fig. 2 Four-factor model for grades 3–6. Latent variables are represented by ellipses; measured variables are represented by rectangles. For simplicity error terms associated with the predictor variables and error covariances are not shown on the diagram

5 Discussion

The aims of this study were to gain insight into how a broad range of numerical and geometrical measures, from early ones to more advanced, associate in children from grades 2 to 6 and examine how these associations shift or remain invariant in these years of important mathematical learning. To meet these aims, we tested more than 1300 children from grades 2 to 6 with the MathPro test [57], a large battery of 20 subtests measuring a variety of mathematical skills. A series of both PAF and CFA analyses were run on each grade to come up with the underlying construct structure of these skills. Let us note, however, that for the grade 2 children, not all the subtests, or all the items of the subtests were presented.

The expectations regarding the underlying structures of the math skills were not very clear since to the best of our knowledge there isn't any other study that has recruited that many mathematical measures. Indeed, the current literature underlines the idea that all math skills are quite heterogeneous and that there exist different profiles of typical or atypical development in mathematics. Yet, there is no consensus about the factors underlying this variability, nor any clear theoretical background related to this. An attempt in this direction is the fourfold model proposed by Karagiannakis, Baccaglini-Frank and Papadatos ([39], which classifies mathematical skills into four domains: core number, memory, visual-spatial, or reasoning. Given the lack of previous similar studies and the poverty in theoretical frameworks that could help in drawing clear hypotheses, our study can be considered both as exploratory enabling us to build meaningful structures as well as confirmatory for testing those structures.

The results obtained through the CFA indicated a configural invariant four-factor structure at all grades. In grade 2, all but one (Number lines 0–1000) of the measures were presented to the pupils. Factor A includes the Subitizing, the Enumeration and the Addition facts subtests. Indeed, the ability to enumerate a collection of dots is highly correlated both with Subitizing ($r = 0.60$) and with simple additions ($r = 0.70$). Yet, grade 2 children performed almost at ceiling and rather fast in Addition facts indicating a gradual transition from immature counting addition strategies (such as counting all) to more advanced strategies (such as counting min or decomposition) or even memory retrieval [72]. Starkey and McCandliss [73] have shown that grouping items into subsets in the subitizing range can speed children's enumeration and that this phenomenon strengthens across grades 1–3. Furthermore, this groupityzing is a unique predictor of arithmetic fluency (see [73]). Accordingly, we could consider that children who have stored the arithmetical fact corresponding to 4 plus 3, could more efficiently enumerate a collection of 7 dots if grouped in 4 and 3 dots or, alternatively, that children who have an enriched understanding of numbers as being composed of various combinations of subsets are both able to faster enumerate collection and solve additions. We consider that factor A represents the early addition ability emerging by combining subitizing and counting skills. It seems that factor A include skills which are supported both by the INS and ENS. Factor B included the Numbers dictation, Next and Previous number subtests, which both assess the knowledge of number sequence. In each of these tasks, a spoken number is presented to the child to be transcoded into the Arabic digits (corresponding to the target, or to the next or previous one). This factor appears to assess the child's ability to move (backward or forward) in the verbal sequence and map it to the Arabic code. These ENS skills have received much interest in the last years as their measure is quite strongly correlated with mathematical skills in general (see, [74]). We consider that factor B represents basic transcoding and ordinality skills. Factor C consisted of the Mental calculations, Word problems, Calculation principles, Numerical patterns and Multi-digit numbers comparison and factor D consisted of the Number lines 0–100, Squares and Building blocks subtests, strongly correlated between them. We propose that both factors C and D evaluate advanced skills. One of the main differences between these two sets of measures is that the factor D is more demanding in terms of visual-spatial skills. Accordingly, we propose to consider factor C as representing numerical advanced skills and factor D as representing visual-spatial advanced skills.

Children from grade 3 to grade 6 received the full battery. The factorial structure obtained was identical in these four grades and very similar to the one obtained in grade 2. Factor A included the same subtests as for grade 2, i.e., Subitizing, Enumeration and Addition facts, but now Multiplication facts was also included. In these age groups, simple arithmetic combinations became gradually "facts", thus stored and retrieved from long-term memory (Gear, 2004). Factor B still included Numbers dictation and Next and Previous Number, although children became increasingly accurate and fast across grades, suggesting a progressive consolidation of the mapping between the verbal and the Arabic number systems. Factor C was again the same as factor C of grade 2, corresponding to advanced numerical skills. Similarly, factor D was the same as the one observed in grade 2, but including the Number lines 0–1000 subtest. It represents visual-spatial advanced skills. Not surprisingly, performance on the subtests of factors C and

D improved from grade to grade. Let us note however that the factor loading of the Numbers comparison subtest was smaller in grade 3 than in grades 4–6, probably because decimal numbers were included in this subtest for the grades 4–6 making the task more difficult for those grades.

In summary, this study aimed at discovering the underlying structure of a broad range of mathematical skills, from early to more advanced, and see whether this structure shifted or remained invariant in children from grades 2 to 6. The results obtained indicated that the underlying structure of mathematical measures was invariant at a configural level throughout the school grades. In a previous research on fifth and sixth grade Greek pupils, Karagiannakis et al. (2017) provided empirical evidence to their proposal to classify mathematical skills into four domains [39]: (1) a core number or INS skills factor that included a dots magnitude comparison, a subitizing-enumeration and a number magnitude comparison subtests; (2) a memory factor that included the multiplication and addition facts, (3) a reasoning factor that included the mental calculations equations, the word problems, the number lines 0–1000, the math terms and the calculations principles tasks, and (4) a visual-spatial factor that included the number lines 0–100 task. The structure we obtained in the present study is somewhat different. This is due partly to the fact that the sample here is larger and wider in terms of the age groups considered but also mostly because the set of numerical measures were not fully overlapping. Indeed, here both INS skills dots comparison and single-digit numbers comparison subtests were removed from the analyses due to their very poor internal reliability or/and the low correlation coefficients with the other measures. We used production rather than verification tasks for the addition and multiplication facts tasks, words problems had not to be solved by the child but the operation that would be needed to solve the problem had to be reported. Finally, five new tasks were considered here: the squares, the building blocks, the next number and the previous number and the numbers dictation task. As previously discussed, Subitizing and Enumeration, as core number skills, were grouped now in the same factor as the Addition and Multiplication facts combining both INS and ENS skills. Further, the 0–1000 number-line task fall, along with new measures (Squares and Building blocks), into the visual-spatial advanced skills.

It is important to address here the question of why the dots comparison task had such a poor internal reliability. This was the case in this study, but also in the study of Karagiannakis and Nöel [57] that used another large sample to analyse the psychometric properties of the battery. We argue that this might be the case because the different stimuli not only vary according to the numbers of dots presented and the ratio between the dots in the two collections, but also by different physical dimensions such a the density of the dots, their area, the convex hull of the collections and so on. Is has been shown that all these dimensions actually are processed in parallel with the processing of the number magnitude of the collection and interact with the final judgment [75]. It is possible that on some trials, the children base their answer on the convex hull of the collection, selecting the larger one, on other trials they might select the collection with the bigger dots and on some others, the ones that are more dense. The fact that all these dimensions interact with the numerical judgment could account for the low internal consistency of the task [76].

We can compare our results to the classification of young children's numerical skills (between 5 and 8 years of age) proposed by Aunio and Räsänen [38] into four main types: (1) symbolic and non-symbolic number sense; (2) understanding mathematical relations (early mathematical-logical principles, arithmetic principles, mathematical operational symbols, place-value and base-ten system); (3) counting skills (knowledge of number-symbols, number word-sequence, enumeration with concrete objects); and (4) basic skills in arithmetic (arithmetic combinations, addition and subtraction skills with number symbols). We did not find a factor corresponding to the first category probably because Dot comparison and Digit comparison were not retained in the analyses (due to poor reliability of the tasks or poor correlation with the other measures). Our factor C (Mental calculations, Word problems, Calculation principles, Numerical patterns and multi-digit Numbers comparison) could be seen as relating to their second category, that is, understanding mathematical relations. Factor C involve numerical skills which are considered components of the MNS [13, 30]. Our factor A (Subitizing, Enumeration and Addition facts, Multiplication facts) and B (Numbers dictation and Next and Previous Number) could partly correspond to their categories 4 (basic skills in arithmetic) and 3 (counting skills) respectively except that Enumeration does not belong to the predicted category according to Aunio and Räsänen's proposition. Let us remind that their classification is only based on an analysis of the testing batteries and that they did not bring any empirical evidences to support them.

To sum up, our findings support the idea that the broad area of mathematical skills that were recruited in this study would fall into four main domains, but the domains involved are somewhat different than those proposed by Karagiannakis et al. [39]. Indeed, although we found that some numerical skills fall into the domain of numerical reasoning (our factor C) or visual-spatial reasoning (our factor D), we did not find clear evidence for a distinction between a core number and a memory domain. Rather, we found a distinction between a factor covering transcoding and ordinality skills (factor B) and those involving Subitizing (which is a core number skill) and counting, associated to emerging arithmetic facts (factor A).

We believe that it is highly important to study associations of students' mathematical skills for two main reasons. First, this knowledge can inform about mathematical learning and guide the design of curricula so that they are better aligned with students' cognitive development (thanks to knowledge about how specific skills underlying a certain mathematical task may develop and stabilize over time). Second, cognitive difficulties in mathematical learning may be associated with a particular set of skills, and this information can guide the design of teaching interventions targeted to students exhibiting particular sets of difficulties. This is aligned with a recent meta-analysis on both typical and atypical population which failed to identify a consistent and specific mathematical learning difficulty (MLD) profile [41]. The authors suggest that MLD should not be seen as a distinct disability that could be distinguished from other forms of mathematical difficulties and that mathematics performance should be assessed on a continuum.

Therefore, our findings can facilitate the educators to outline the mathematical profile of their students and to apply the appropriate teaching strategies. For example, students who underperform mainly on the INS and retrieval skills (factor A) will benefit most by remediations stimulating their conceptional understanding and their mathematical flexibility rather than focusing on learning by heart rules through explicit teaching (find more examples in [55]).

A strength of this study is that a very large sample of children from five different grades were tested with a very large set of mathematical tasks. Yet, a limitation of this study is that the same data sample was used for running both the exploratory and the confirmatory factor analyses. This is so because the limited theoretical framework available did not provide us with a well-structured set of hypotheses to test. So, further work will be needed to confirm the structures obtained using other samples and examining also the metric and scalar invariance. Future research could also attempt to validate the obtained structures by assessing associated cognitive skills in children. For instance, it would be interesting to measure whether skills in measures belonging to the visual-spatial factor (D) correlate with performance in visual-spatial reasoning measures while skills belonging to the numerical reasoning factor (C) might correlate with non visual-spatial reasoning tasks. Similarly, it would be interesting to see which mathematical measures more strongly correlate with children's ability to store or retrieve information from memory.

Finally, it will be interesting to study children with MLD and examine whether their mathematical impairment is global or whether it mainly concerns some of the factors more specifically.

Author contributions Data collection: A.B-F., C.T. Statistical analysis: G.K. Main manuscript text: G.K., M-P. N., A.B-F. All authors reviewed the manuscript.

Funding Università di Pisa, PRA-2020-31, PRA-2020-31.

Data availability The data that support the findings of this study are available from the corresponding author, [GK], upon reasonable request.

Declarations

Competing interests The authors declare no competing interests.

Appendix

See Tables 8, 9, 10, 11 and 12 here.

Table 8 Number of items of MathPro subtests administrated per grade

	Grade 2	Grade 3	Grade 4–6
1.Dots comparison	30	30	30
2. Computer mouse speed	10	10	10
3.Single digit numbers comparison	24	24	24
4.Multi-digit numbers comparison	6	9	12
5.Screen keyboard use	10	10	10
6.Numbers dictation*	30	30	30
7.Next number*	18	18	18
8.Previous number*	18	18	18
9.Subitizing	20	20	20
10.Enumeration	14	14	14
11.Addition facts retrieval	12	12	12
12.Mutiplication facts retrieval	0	14	14
13.Mental calculations	12	24	24
14.Number Lines 0–100	22	22	22
15.Number Lines 0–1000	0	22	22
16.Squares	10	10	10
17.Building blocks	8	8	8
18.Word problems*	10	18	18
19.Calculation principles*	15	15	15
20.Numerical patterns*	18	18	18

*Task terminates after 3 consecutive errors

Table 9 Results of the ANOVAs for grade of the MathPro subtests

	Grade
1. Dots comparison (AC)	$F_{4,1346}=32.15^{***}$
3.Single-digit numbers comparison (AC)	$F_{4,1344}=18.29.72^{***}$
3. Single-digit numbers comparison (RT)	$F_{4,1344}=5.34^{***}$
4. Multi-digit numbers comparison[1] (AC)	$F_{2,839}=17.35^{***}$
4. Multi-digit numbers comparison[1] (RT)	$F_{2,838}=31.73^{***}$
6.Numbers dictation (AC)	$F_{5,1346}=318.64^{***}$
6.Numbers dictation[1] (RT)	$F_{2,838}=12.54^{***}$
7.Next number (AC)	$F_{4,1344}=45.23^{***}$
7.Next number[1] (RT)	$F_{2,837}=12.04^{***}$
8.Previous number (AC)	$F_{4,1346}=67.68^{***}$
8.Previous number[1] (RT)	$F_{2,836}=13.25^{***}$
9.Subitizing (AC)	$F_{4,1344}=51.71^{***}$
10.Enumeration (AC)	$F_{4,1346}=21.35^{***}$
10.Enumeration[2] (RT)	$F_{3,1127}=43.18^{***}$
11.Addition facts retrieval (AC)	$F_{4,1346}=41.76^{***}$
11.Addition facts retrieval (RT)	$F_{4,1343}=98.53^{***}$
12.Multiplication facts retrieval[2] (AC)	$F_{3,1125}=9.18^{***}$
12.Multiplication facts retrieval[1] (RT)	$F_{2,835}=23.15^{***}$
13.Mental calculations (AC)	$F_{4,1345}=26.19^{***}$
14.Number line 0–100 (PAE)	$F_{4,1345}=175.79^{***}$
15.Number line tasks 0–100[2] (PAE)	$F_{3,1126}=61.28^{***}$
16.Squares (AC)	$F_{4,1345}=87.32^{***}$
17.Building blocks (AC)	$F_{4,1344}=41.64^{***}$
18.Word problems (AC)	$F_{4,1342}=27.19^{***}$
19.Calculation principles (AC)	$F_{4,1340}=18.35^{***}$
20Numerical patterns (AC)	$F_{4,1342}=59.25^{***}$

AC = Accuracy; RT = Reaction time (in milliseconds); PAE = Percentage of absolute errors

[1]grades 4–6

[2]grades 3–6

$^{***}p<0.001$

Table 10 First-order correlations between the RT measures controlled for the processing speed for grade 2

	7	9	11	14	16
7.Numbers dictation (RT)	–				
9.Next number (RT)	0.39**	–			
11.Previous number (RT)	0.16*	0.49**	–		
14.Enumeration (RT)	0.22*	0.36**	0.34**	–	
16.Addition facts retrieval (RT)	0.21*	0.45**	0.45**	0.43**	–

**p < 0.01

*p < 0.05

Table 11 First-order correlations between the RT measures controlled for the processing speed for grades 3–4

	7	9	11	14	16	18
7.Numbers dictation (RT)	–	0.49**	0.48**	0.12	0.23**	0.26**
9.Next number (RT)	0.66**	–	0.58**	0.29**	0.25**	0.34**
11.Previous number (RT)	0.65**	0.67**	–	0.25**	0.33**	0.38**
14.Enumeration (RT)	0.36**	0.36**	0.43**	–	0.47**	0.42**
16.Addition facts retrieval (RT)	0.37**	0.44**	0.57**	0.52**	–	0.59**
18. Multiplication facts retrieval (RT)	0.22**	0.36**	0.45**	0.43**	0.71**	–

Up right: correlations for grade 3; Bottom left: correlations for grade 4.

**$p < 0.01$

*$p < 0.05$

Table 12 First-order correlations between the RT measures controlled for the processing speed for grades 5–6

	7	9	11	14	16	18
7.Numbers dictation (RT)	–	0.33**	0.31**	0.20*	0.21*	0.17*
9.Next number (RT)	0.53**	–	0.53**	0.24**	0.27**	0.18*
11.Previous number (RT)	0.43**	0.49**	–	0.29**	0.37**	0.29**
14.Enumeration (RT)	0.16*	0.26**	0.37**	–	0.63**	0.42**
16.Addition facts retrieval (RT)	0.16*	0.26**	0.45**	0.52**	–	0.57**
18. Multiplication facts retrieval (RT)	0.12	0.20*	0.29**	0.36**	0.51**	–

Up right: correlations for grade 5; Bottom left: correlations for grade 6.

**$p < 0.01$

*$p < 0.05$

References

1. National Council for Teachers of Mathematics. Principles and standards for school mathematics. Reston, VA: Author. 2000.
2. OECD. PISA 2012 Mathematics Framework. Paris: OECD Publications http://www.oecd.org/dataoecd/8/38/46961598.pdf. 2010.
3. Stacey K, Turner R. Assessing mathematical literacy: the PISA experience. New York: Springer; 2015.
4. Chu FW, Geary DC. Early numerical foundations of young children's mathematical development. J Exp Child Psychol. 2015;132:205–12. https://doi.org/10.1016/j.jecp.2015.01.006.
5. LeFevre JA, Fast L, Skwarchuk SL, Smith-Chant BL, Bisanz J, Kamawar D, Penner-Wilger M. Pathways to mathematics: Longitudinal predictors of performance. Child Dev. 2010;81(6):1753–67. https://doi.org/10.1111/j.1467-8624.2010.01508.x.
6. Halberda J, Feigenson L. Developmental change in the acuity of the "number sense": the approximate number system in 3-, 4-, 5-, and 6-year-olds and adults. Dev Psychol. 2008;44:1457–65. https://doi.org/10.1037/a0012682.
7. Wood JN, Spelke ES. Infants' enumeration of actions: numerical discrimination and its signature limits. Dev Sci. 2005;8:173–81. https://doi.org/10.1111/j.1467-7687.2005.00404.x.
8. Feigenson L, Dehaene S, Spelke E. Core systems of number. Trends in Cogn Sci. 2004;8(7):307–14. https://doi.org/10.1016/j.tics.2004.05.002.
9. Piazza M, Facoetti A, Trussardi AN, Berteletti I, Conte S, Lucangeli D, et al. Developmental trajectory of number acuity reveals a severe impairment in developmental dyscalculia. Cognition. 2010;116(1):33–41. https://doi.org/10.1016/j.cognition.2010.03.012.
10. Holloway ID, Ansari D. Mapping numerical magnitudes onto symbols: the numerical distance effect and individual differences in children's mathematics achievement. J Exp Child Psychol. 2009;103:17–29. https://doi.org/10.1016/j.jecp.2008.04.001.
11. Vanbinst K, Ghesquière P, De Smedt B. Numerical magnitude representations and individual differences in children's arithmetic strategy use. Mind Brain Educ. 2012;6:129–36. https://doi.org/10.1111/j.1751-228X.2012.01148.x.
12. Spelke ES, Lee SA, Izard V. Beyond core knowledge: natural geometry. Cogn Sci. 2010;34:863–84. https://doi.org/10.1111/j.1551-6709.2010.01110.x.
13. Whitacre I, Henning B, Atabaş Ş. Disentangling the research literature on number sense: three constructs, one name. Rev Educ Res. 2020;90(1):95–134. https://doi.org/10.3102/0034654319899706.
14. Gelman R, Gallistel CR. The child's understanding of number. Cambridge, MA: Harvard University Press; 1978.
15. Kroesbergen EH, Van Luit JEH, Van Lieshout ECDM, Van Loosbroek E, Van de Rijt BAM. Individual differences in early numeracy: the role of executive functions and subitizing. J Psychoeduc Assess. 2009;27:226–36. https://doi.org/10.1177/0734282908330586.
16. Aunola K, Leskinen E, Lerkkanen MK, Nurmi JE. Developmental dynamics of math performance from preschool to Grade 2. J Educ Psychol. 2004;96:699–713. https://doi.org/10.1037/0022-0663.96.4.699.
17. Noël M-P, Turconi E. Assessing number transcoding in children. Eur Rev Appl Psychol. 1999;49(4):295–302.

18. Butterworth B. The development of arithmetical abilities. J Child Psychol Psychiatry. 2005;46(1):3–18. https://doi.org/10.1111/j.1469-7610.2004.00374.x.
19. Lyons I, Beilock S. Mathematics anxiety: separating the math from the anxiety. Cereb Cortex. 2011;22:2102–10. https://doi.org/10.1093/cercor/bhr289.
20. De Smedt B, Noël MP, Gilmore C, Ansari D. 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 behaviour. Trends Neurosci, Educ. 2013;2:48–55. https://doi.org/10.1016/j.tine.2013.06.001.
21. Schneider M, Beeres K, Coban L, Merz S, Susan Schmidt S, Stricker J, De Smedt B. Associations of non-symbolic and symbolic numerical magnitude processing with mathematical competence: a meta-analysis. Dev Sci. 2017;20:1–16. https://doi.org/10.1111/desc.12372.
22. Groen GJ, Parkman JM. A chronometric analysis of simple addition. Psychol Rev. 1972;79(4):329–43. https://doi.org/10.1037/h0032950.
23. Picozzi M, de Hevia MD, Girelli L, Cassia VM. Seven-month-olds detect ordinal numerical relationships within temporal sequences. J Exp Child Psychol. 2010;107(3):359–67. https://doi.org/10.1016/j.jecp.2010.05.005.
24. Geary DC. Sexual selection and sex differences in mathematical abilities. Behav Brain Sci. 1996;19(2):229–47. https://doi.org/10.1017/S0140525X00042400.
25. van Marle K, Chu FW, Li Y, Geary DC. Acuity of the approximate number system and preschoolers' quantitative development. Dev Sci. 2014;17(4):492–505. https://doi.org/10.1111/desc.12143.
26. Parmar RS, Cawley JF, Frazita RR. Word problem-solving by students with and without mild disabilities. Except Child. 1996;62(5):415–29. https://doi.org/10.1177/001440299606200503.
27. Geary DC. Cognitive predictors of achievement growth in mathematics: a 5-year longitudinal study. Dev Psychol. 2011;47:1539–52. https://doi.org/10.1037/a0025510.
28. Siegler RS, Booth JL. Development of numerical estimation in young children. Child Dev. 2004;75:428–44. https://doi.org/10.1111/j.1467-8624.2004.00684.x.
29. Mix KS, Levine SC, Cheng Y-L, Young C, Hambrick DZ, Ping R, Konstantopoulos S. Separate but correlated: the latent structure of space and mathematics across development. J Exp Psychol Gen. 2016;145(9):1206–27. https://doi.org/10.1037/xge0000182.
30. Pittalis M, Pitta-Pantazi D, Christou C. A longitudinal study revisiting the notion of early number sense: algebraic arithmetic as a catalyst for number sense development. Math Think Learn. 2018;20(3):222–47. https://doi.org/10.1080/10986065.2018.1474533.
31. Cragg L, Gilmore C. Skills underlying mathematics: the role of executive function in the development of mathematics proficiency. Trends Neurosci, Educ. 2014;3(2):63–8. https://doi.org/10.1016/j.tine.2013.12.001.
32. Peng P, Wang T, Wang C, Lin X. A meta-analysis on the relation between fluid intelligence and reading/mathematics: effects of tasks, age, and social economics status. Psychol Bull. 2019;145(2):189–236. https://doi.org/10.1037/bul0000182.
33. Holmes VM, McGregor J. Rote memory and arithmetic fact processing. Mem Cognit. 2007;35:2041–51. https://doi.org/10.3758/BF03192936.
34. Friso-van den Bos I, van der Ven S, Kroesbergen E, van Luit J. Working memory and mathematics in primary school children: a meta-analysis. Educ Res Rev. 2013;10:29–44. https://doi.org/10.1016/j.edurev.2013.05.003.
35. Evans J, Hannula M, Zan R, Brown L. Affect in mathematics education—exploring theoretical frameworks. Educ Stud Math. 2006; 63(2). 113–121. http://www.jstor.org/stable/25472116
36. Lee YS, Park YS, Ginsburg H. Socio-economic status differences in mathematics accuracy, strategy use, and profiles in the early years of schooling. ZDM Mathematics Education. 2016;48:1065–78. https://doi.org/10.1007/s11858-016-0783-y.
37. De Souza Salvador L, Moura R, Wood G, Geraldi Haase V. Cognitive heterogeneity of math difficulties: a bottom-up classification approach. J Numerical Cogn. 2019;5:55–85. https://doi.org/10.5964/jnc.v5i1.60.
38. Aunio P, Räsänen., P. Core numerical skills for learning mathematics in children aged five to eight years – a working model for educators. Eur Early Child Educ Res. 2016;24(5):684–704. https://doi.org/10.1080/1350293X.2014.996424.
39. Karagiannakis G, Baccaglini-Frank A, Papadatos Y. Mathematical learning difficulties subtypes classification. Front Hum Neurosci. 2014;8(57):10705519909540118.
40. Karagiannakis G, Baccaglini-Frank A, Roussos P. Detecting strengths and weaknesses in learning mathematics through a model classifying mathematical skills. Aust J Learn Difficulties. 2017;21(2):115–41. https://doi.org/10.1080/19404158.2017.1289963.
41. Kroesbergen E, Huijsmans M, Friso-van den Bos I. A meta-analysis on the differences in mathematical and cognitive skills between individuals with and without mathematical learning disabilities. Rev Educ Res. 2022. https://doi.org/10.3102/00346543221132773.
42. Bartelet D, Ansari D, Vaessen A, Blomert L. Cognitive subtypes of mathematics learning difficulties in primary education. Res Dev Disabil. 2014;35(3):657–70. https://doi.org/10.1016/j.ridd.2013.12.010.
43. Ashkenazi S, Black J, Abrams D, Hoeft F, Menon V. Neurobiological underpinnings of math and reading learning disabilities. J Learn Disabil. 2013;46(6):549–69. https://doi.org/10.1177/0022219413483174.
44. Andersson U, Ostergren R. Number magnitude processing and basic cognitive functions in children with mathematical learning disabilities. Learn Individ Differ. 2012;22:701–14. https://doi.org/10.1016/j.lindif.2012.05.004.
45. Fias W, Menon V, Szucs D. Multiple components of developmental dyscalculia. Trends Neurosci, Educ. 2013;2(2):43–7. https://doi.org/10.1016/j.tine.2013.06.006.
46. De Visscher A, Szmalec A, Van der Linden L, Noël M-P. Serial-order learning impairment and hypersensitivity-to-interference. Cognition. 2015;144:38–48. https://doi.org/10.1016/j.cognition.2015.07.007.
47. Skagerlund K, Träff U. Number processing and heterogeneity of developmental dyscalculia: subtypes with different cognitive profiles and deficits. J Learn Disabil. 2016;49(1):36–50. https://doi.org/10.1177/0022219414522707.
48. Träff U, Olsson L, Östergren R, Skagerlund K. Heterogeneity of developmental dyscalculia: cases with different deficit profiles. Front in Psychol. 2017;7:1–15. https://doi.org/10.3389/fpsyg.2016.02000.
49. Geary DC. Mathematical disabilities: cognitive, neuropsycho- logical, and genetic components. Psychol Bull. 1993;114:345–62. https://doi.org/10.1037/0033-2909.114.2.345.

50. Chan WWL, Wong TT-Y. Subtypes of mathematical difficulties and their stability. J Educ Psychol. 2020;112(3):649–66. https://doi.org/10.1037/edu0000383.
51. von Aster M. Developmental cognitive neuropsychology of number processing and calculation: varieties of developmental dyscalculia. Eur Child Adolesc Psychiatry. 2000;9:41–57. https://doi.org/10.1007/s007870070008.
52. Baccaglini-Frank A. To tell a story, you need a protagonist: how dynamic interactive mediators can fulfil this role and foster explorative participation to mathematical discourse. Educ Stud Math. 2021;106:291–312. https://doi.org/10.1007/s10649-020-10009-w.
53. Lewis KE, Fisher MB. Taking stock of 40 years of research on mathematical learning disability: methodological issues and future directions. J Res Math Educ. 2016;47:338–71. https://doi.org/10.5951/jresematheduc.47.4.0338.
54. Lewis KE. Designing a bridging discourse: re-mediation of a mathematical learning disability. J Learn Sci. 2017;26(2):320–65. https://doi.org/10.1080/10508406.2016.1256810.
55. Noël MP, Karagiannakis G. Effective teaching strategies for dyscalculia and learning difficulties in mathematics. Perspectives from cognitive neuroscience. New York: Routledge; 2022. p. 30.
56. Robotti E. Designing innovative learning activities to face up to difficulties in algebra of dyscalculia students: how exploit the functionality of AlNuSet. In: Baccaglini-Frank A, Leung A, editors. Digital technologies in designing mathematics education tasks—potential and pitfalls. Berlin: Springer Publisher; 2017. p. 193–214.
57. Karagiannakis G, Noël M-P. Mathematical profile test: a preliminary evaluation of an online assessment for mathematics skills of children in grades 1–6. Behav Sci. 2020;10:126. https://doi.org/10.3390/bs10080126.
58. Baccaglini-Frank A, Karagiannakis G, Pini C, Termine C, Girelli L. Identifying mathematical learning profiles of children between age 6 and 12: the Italian standardization of the MathPro battery. Six-mounthly J Learn Res Innov Educ RICERCHE. 2020;12:167–99. https://doi.org/10.32076/RA12109.
59. Russell DW. In search of underlying dimensions: the use (and Abuse) of factor analysis in personality and social psychology bulletin. Pers Soc Psychol Bull. 2002;28(12):1629–46. https://doi.org/10.1177/014616702237645.
60. Browne MW, Cudeck R. Alternative ways of assessing model fit. In: Bollen KA, Long JS, editors. Testing structural equation models. Beverly Hills, CA: Sage; 1993. p. 136–62.
61. Hu L, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. Struct Equ Modeling. 1999;6(1):1–55. https://doi.org/10.1080/10705519909540118.
62. Akaike H. Factor analysis and AIC. Psychometrika. 1987;52(3):317–32. https://doi.org/10.1007/BF02294359].
63. Bentler PM, Wu EJC. EQS 6.1 for windows user's guide. Encino, CA: Multivariate Software; 2005.
64. Byrne BM. Structural equation modeling with AMOS: basic concepts, applications, and programming. New York, NY: Routledge; 2010.
65. Bollen KA, Stine RA. Bootstrapping goodness-of-fit measures in structural equation models. Sociol Methods Res. 1992;21:205–29. https://doi.org/10.1177/0049124192021002004.
66. Arbuckle JL. Amos 23.0 user's guide. Chicago, IL: IBM SPSS; 2014.
67. Kline RB. Principles and practice of structural equation modeling. New York, NY: GuilfordPress; 1998.
68. Field A. Discovering statistics using SPSS. 3rd ed. London: Sage Publications Ltd.; 2009.
69. Brown TA. Confirmatory factor analysis for applied research. 2nd ed. New York, NY: Guilford Publications; 2015.
70. Jöreskog KG. Testing structural equation models. In: Bollen KA, Long JS, editors. Testing structural equation models. Thousand Oaks, CA: Sage; 1993. p. 294–316.
71. Arifin W, Yusoff M. Confirmatory factor analysis of the Universiti Sains Malaysia emotional quotient inventory among medical students in Malaysia. SAGE Open. 2016. https://doi.org/10.1177/2158244016650240.
72. Ashcraft MH, Fierman BA. Mental addition in third, fourth, and sixth graders. J Exp Child Psychol. 1982;33(2):216–34. https://doi.org/10.1016/0022-0965(82)90017-0.
73. Starkey GS, McCandliss BD. The emergence of "groupitizing" in children's numerical cognition. J Exp Child Psychol. 2014;126:120–37. https://doi.org/10.1016/j.jecp.2014.03.006.
74. Lyons IM, Price GR, Vaessen A, Blomert L, Ansari D. Numerical predictors of arithmetic success in grades 1–6. Dev Sci. 2014;17(5):714–26. https://doi.org/10.1111/desc.12152.
75. Gebuis T, Cohen-Kadosh R, Gevers W. Sensory-integration system rather than approximate number system underlies numerosity processing: a critical review. Acta Physiol. 2016;171:17–35.
76. Clayton S, Gilmore C, Inglis M. Dot comparison stimuli are not all alike: The effect of different visual controls on ANS measurement. Acta Physiol. 2015;161:177–84.