Please find the published paper at:

https://www.tandfonline.com/doi/full/10.1080/19404158.2017.1289963

https://doi.org/10.1080/19404158.2017.1289963

Detecting strengths and weaknesses in learning mathematics through a model classifying mathematical skills

1. Introduction

Most of the literature on students' difficulties in learning mathematics from the field of cognitive psychology investigates development of basic number processing, introducing terms for describing atypical situations. These terms include "Developmental Dyscalculia" (DD), "Mathematical Learning Disabilities" (MLD), "Mathematical Learning Difficulties" among many others (Passolunghi & Siegel, 2004; Piazza et al., 2010; Rousselle & Noël, 2007). However, these definitions are still a topic of debate (e.g., Kosc, 1974; Mazzocco, 2005; Lewis & Fisher, 2016), and the terminology is inconsistent. For example Mazzocco and Räsänen (2013) note that mathematical learning difficulties "has been used as synonymous with DD [...], but also as distinct from DD when [it] is used to refer to the larger category of mathematics difficulties." (ibid., p. 66). The cause of the terminological confusion is that currently there is no clear generally accepted classification of developmental mathematical weaknesses (Szucs, 2016). Lewis and Fisher (2016), in conducting a systematic review of 164 studies on MLD, note that "it is unknown how much variability exists across the body of research, which raises questions about the reliability and validity of MLD identification particularly related to differentiating cognitive and non-cognitive sources of low achievement" (ibid., p. 341). Fletcher, Lyon, Fuchs and Barnes (2007) also note that there are still "no consistent standards by which to judge the presence or absence of LDs [learning difficulties] in math" (ibid, p. 207). The lack of consensus to identify the central characteristics of an MLD as well as the comorbidity and heterogeneity that characterize the MLD students (Bartelet, Ansari, Vaessen & Blomert, 2014; Watson & Gable, 2013; Szucs & Goswami, 2013) have also led researchers to propose various models in order to explain different MLD subtypes.

In this paper we explain how and why we have developed a literature-driven a priori four-pronged model for detecting difficulties in learning mathematics (Authors & other, 2014), and we bring evidence, through an empirical study, to support the model's solidity. The purpose of the study presented in this manuscript is (a) to empirically examine our proposed model in order to determine whether and how it can differentiate students with and without difficulties in learning mathematics, and (b) to provide educators with a means for sketching students' mathematical profiles that can be used to inform educational choices.

2. Review of ~~Attempts at~~Efforts to Define~~ing~~ Types of MLD

Our proposal of a model for detecting difficulties in learning mathematics has roots in a line of research in which various models have been advanced, constituting interesting attempts at explaining differences in the population of students who under achieve in mathematics (e.g, Fuchs & Fuchs, 2002; Geary, 2004; von Aster & Shalev, 2007). A common feature to many of these models, including our own, is the attempt to uncover underlying cognitive factors to students' underachievement in mathematics (Shalev et al., 2001; Augustyniak, Murphy & Phillips, 2005; Fuchs et al., 2007; Butterworth, 2010; Andersson, & Östergren, 2012; Mazzocco & Räsänen, 2013; Szucs, 2016), either related to a specific cognitive domain (e.g., the "core number system"), or to general cognitive domain. For example, recent studies have investigated the roles of general executive functions, such as working memory and inhibition, in mathematical achievement (Andersson, 2008; Geary, 2004; Cragg & Gilmore, 2014; Passolunghi & Siegel, 2004). Also visual-spatial deficits have been attributed to poor mathematical achievement, including achievement in geometry (Mammarella & Cornoldi, 2005; Mammarella, Lucangeli & Cornoldi 2010; Mammarella, Giofrè, Ferrara & Cornoldi, 2013; Szucs, 2016).

Models stemming from within this line of research are constructed upon the assumption that it is possible to link students' cognitive abilities to their performances on appropriately designed assessment tasks. At the methodological level this assumption leads to another one common to these studies, that is, that high performance on one or a set of tasks corresponds to the presence or absence of a particular cognitive ability "in" the student. We do see this as a limitation, but in this type of studies we have not yet found a way around it. However, there are still significant differences underlying the approaches used for the development of such models and underlying the methodology used to validate each model. These differences ~~currently~~ contribute to make it very difficult to compare results across studies (Lewis & Fisher, 2016).

An important difference between the models is in *how* these links between (internal) cognitive abilities and (externally visible) performance are theorized. For example, Geary (2004) hypothesized a classification based on types of possible underlying deficits (procedural, semantic memory, spatial), and used the notion of "supporting competencies" that are either conceptual or procedural to link a set of underlying "cognitive systems" within which the deficits may reside to a "mathematical domain". Geary's classification of subtypes of MLD includes reference to parts of the brain containing each type of deficit.

von Aster and Shalev's model (2007) arises from a previous classification of subtypes – verbal, Arabic, and pervasive – (von Aster, 2000), detected through clinical case studies and quantitative research using cluster analysis of students' performance on batteries of tasks elaborated to investigate their abilities considered relevant to mathematical performance. The researchers took into account findings pointing to the genetic underpinning of a spatially oriented number line that develops through elementary school, together with working memory and number symbolization (explicit reference is made to Dehaene's research on number sense (1997, 2001)). The four steps of the model include acquiring the core system of magnitude (or cardinality), the verbal number system, the Arabic number system, and finally the mental number line that also involves the spatial-ordinal properties of number. The brain areas involved in each step are listed explicitly, as well as the student's mathematical abilities that develop from infancy through school, ~~thanks to~~in parallel with the increasing working memory. Therefore this model theorizes links between students' performance on numerical tasks and their overall mathematical achievement in school and the development of specific abilities in specific brain areas.

As a third example of how these links have been theorized, we refer to a model proposed by Mulligan and her colleagues, which makes use of a specific theoretical construct: "awareness of mathematical patterns and structure" (AMPS) (Mulligan, 2009; Mulligan, 2011; Mulligan & Mitchelmore, 2013). The model describes different levels of structural development, or of AMPS, and relates these, o~~O~~n one hand~~, the researchers show how AMPS correlates~~ to students' general mathematical achievement~~;~~ , and, on the other hand~~, AMPS is theorized as an underlying construct common~~ to a range of concepts and skills based on a broad range of cognitive factors, including (though they are not limited to) visualization, visual memory and representation, reasoning and inference (Mulligan, Mitchelmore & Stephanou, 2015).

Moreover, methodologically~~, some~~ models seem to be developed in two different ways. The first, stemming from from personal elaborations of theoretical considerations emerging from a review of the literature, previous studies, and clinical analyses (as in the case of Geary's model, 2004), lead~~ing to an~~s to "a priori" models (we shall refer to this approach in developing models as top-down)~~; while other~~. The second, developed through~~ models seem to be developed as~~ attempts of grouping students' performances on batteries of tasks in various ways (e.g, Fuchs & Fuchs, 2002; von Aster, 2000), leads~~ing~~ to~~ an~~ "a posteriori" models (we shall refer to this approach as bottom-up). We do not have evidence to claim that either methodological choice is more sound than the other. However, we note that the choice does make a difference in the role played by the assessment tasks used in the studies. In a bottom-up approach, analyses of students' performance on sets of assessment tasks constitute the emerging models themselves, while in top-down approach, assessment~~In particular, if~~ tasks are designed to bring empirical evidence to potentially support an "a priori" model. In this case they must be aligned with the basic theoretical assumptions upon which the model is grounded, and this alignment must be made explicit in the experimental design. This is the approach we take here and we will make such alignment explicit in section 5.

3. A four-pronged model for detecting strengths and weaknesses in learning mathematics

The models presented in the previous section, together with an analysis of existing literature on hypotheses underlying MLD, and the clinical experiences of the authors, contributed to the development of the four-pronged model we now present (published in Authors, 2014). In particular, our model tries to overcome some limitations of the previous models. For example, in Geary's model (2004) it is not clear how the underlying "cognitive systems", within which the deficits may reside, correspond to the "conceptual and procedural supporting competencies" in mathematical achievement. Moreover, this model does not take into account any "core number" hypotheses.

On the other hand, von Aster and Shalev's model (2007) focuses exclusively on development of the core number domain. Furthermore, in such model cardinality is assumed to be developed before ordinality. However, there is increasing evidence suggesting that development proceeds in the opposite direction: already in the late 1970's some neuro-scientific experiments suggested that ordinality occurred in young children at a much earlier age than cardinality (Brainerd, 1979); and more recent neural evidence shows that accessing ordinal information from numerical symbols relies on a different network of brain regions and that such accessing has qualitatively different behavioral patterns when compared to ordinal processing of perceptual magnitudes (Lyons & Beilock, 2013; Coles, 2014; Lyons & Ansari, 2015).

The model we have been developing hypothesizes associations between sets of mathematical skills. We chose not to introduce new theoretical constructs, but instead to use the existing literature to identify fundamental mathematical cognitive skills, and group them into four previously identified domains[1], revisiting MLD hypotheses introduced in the cognitive psychology literature and intertwining them in a mathematical holisticly meaningful way. Below we summarize the sets of mathematical skills that hypothetically characterize each domain, as presented in Authors (2014). The skills considered in Table 1 are not a comprehensive list of all mathematical skills, but ones that, from the literature, seem to be particularly rooted in each of the domains[2].

Table 1

Domain	*Mathematical skills associated with the domain*
Core number	estimating accurately a small number of objects (e.g., 4-5); estimating approximately quantities; placing numbers on number

[1] The domains are not considered, a priori, to be hierarchical in any way.

[2] In particular, many complex mathematical skills such as counting, recognizing patterns, base ten structure, multiplicative reasoning, etc., are not included, since they critically involve more than one domain. Moreover, we see their connection with a particular domain to be heavily based on how they are assessed.

	lines; managing Arabic symbols; transcoding a number from one representation to another (analogical-Arabic-verbal)
Memory (retrieval and processing)	retrieving numerical facts; decoding terminology (numerator, denominator, isosceles, equilateral); remembering theorems and formulas; performing mental calculations fluently; remembering procedures and keeping track of steps
Reasoning	grasping mathematical concepts, ideas and relations; understanding multiple steps in complex procedures/algorithms; grasping basic logical principles (conditionality – "if…then…" statements – commutativity, inversion); grasping the semantic structure of problems; (strategic) decision making; generalizing
Visual-spatial	interpreting and using spatial organization of representations of mathematical objects (for example, numbers in decimal positional notation, exponents, geometrical figures or rotations); placing numbers on a number line; confusing Arabic numerals and mathematics symbols; performing written calculation when position is important (e.g. borrowing/carrying); interpreting graphs and tables

4. Aims of this study

Our first aim in this study is to test the four-pronged theoretical model described above and developed through a top-down approach, by studying students' performances on a newly designed computer-based experimental battery of mathematical tasks. Such tasks, designed by the first author of this paper, are grouped a priori assuming that certain sets of tasks tap on a particular domain of mathematical cognitive skills. These assumptions are made explicit, for each, set of tasks, in section 5. Students' performances are analyzed a posteriori, leading to a bottom-up grouping that we compare with the a priori grouping to test the solidity of the four-pronged model introduced above. The second aim of this study is to ~~detect the most common~~gain insight into mathematics learning profiles of students with or without difficulties in learning mathematics.

The age we chose to target with this first trial of the experimental battery was 10-12 years, corresponding to 5th and 6th grade in Greek primary schools. This choice was made due to the assumption that the development of the mathematical skills elicited during the battery~~, in cases of typical development,~~ was likely to be complete in children of this age range. The population we report on in this paper is a "typical population", in order to register "typical" performance on the battery, and validate a tool for detecting a 10-12 year old (Greek[3]) student's mathematical learning profile (at least with respect to the skills elicited). To put our findings in relation with other

[3] Or possibly of other nationalities once the tool has been calibrated on other populations.

research on MLD, we used a standardized test, typically used in Greece to select underachieving students (with different levels of severity: Low Achievement (LA), or MLD) in mathematics.

In particular, this paper addresses the following questions:

1) How are tasks of the experimental battery correlated with each other?
2) Is there evidence for supporting our theoretical four-pronged model?
3) How do performances of students in the MLD and LA groups (according to the selection criterion) compare on the mathematical tasks of the experimental battery?
4) What types of mathematical profiles emerge in general, and, in particular, do the under achieving students constitute separate groups?

5. Method

5.1 Participants

The participants were 165 grade 5 and grade 6 Greek children (mean age=11.26, SD=.59 years), 91 of whom were males. They were randomly recruited within four public schools in Athens and the surrounding areas, from different socio-economic backgrounds. All children were fluent speakers of Greek, had normal visual acuity, and no hearing loss. Once schools had agreed to participate in the study, an information letter was sent to each child's parents, together with an opt-out form. Once we received each approval form, a non-verbal IQ test (see section 5.4) was administered and students were excluded if they had a score below 1.5 SD. Approval forms were received for a sample of 169 students. Four students were excluded due to low non-verbal IQ scores.

5.2 Materials and procedure

To the students who met the non-verbal IQ requirements, two other tests were administered: first a standardized test for assessing the student's mathematics achievement (NUCALC battery, see section 5.3), and finally the experimental battery. All tests were administered individually by one of five trained research assistants, within the school context, in a computer lab in the school. The experimental battery was computer-based and was administered under the supervision of the research assistants, who would orally give instructions at the beginning of each new set of tasks. In all tasks there were two practice trials before the testing phase, to ensure that the student understood the task. The order of the tasks of the test was the same for all students. Generally, the individual test administration was completed within one session lasting about 70 min. The administration procedure took place within two weeks.

5.3 Mathematics achievement selection criterion

The selection criterion was based on students' total score on the Greek standardized version of *Neuropsychological Test Battery for Number Processing and Calculation in Children* (NUCALC battery) (Koumoula et al., 2004). It is an untimed paper-and-pencil arithmetic fluency test which consists of six subtests: Dictation of Numbers, Mental Addition, Reading Numbers, Oral Comparison, Problem Solving and Written Comparison. Children with low achievement below the 16th percentile on NUCALC battery were classified as MLD students, children with low average achievement between 17th and 30th percentile were classified as low achievement (LA) students and children with scores above the 40th percentile were classified as typical achievement (TA) students. The above cut off scores were based on the Greek standardized norms of the NUCALC battery.

5.4 Non-verbal IQ

The Colored Raven Progressive Matrices is a normed untimed visual-spatial reasoning test for children in the age range of 5–11 (Raven, Court, & Raven, 1995). Children were assessed on 36 items involving colored patterns and were asked to select the missing piece out of six choices.

5.5 The experimental battery

The experimental battery is a computer-based total of 13 tasks which was developed by the first author especially for the purpose of the present study. All tasks was programmed in the C++ language using the open-source cross-platform application framework QT version 4.7 and the open-source GNU compiler gcc. All the functions were implemented using generic QT/C++ approaches, so that the same code can be compiled for different operating systems (OS) such as Windows, Mac OS X and Linux, with only minor differences in the appearance. The actual tests were executed on Windows machines. Once the battery had been completed on the computer, output was extracted in the form of a bar chart in which the student's Stanine Score (Thorndike, 1982) on each task was shown. Below we describe the tasks of the experimental battery and explain their a priori grouping in the domains of the four-pronged model.

5.5.1 Tasks in the core number domain

Subitizing-Enumeration

Students were instructed to compare a random array of dots shown on the left half of the computer screen to an Arabic digit shown on the right half of the screen. Underneath the array of dots appeared the word "NO", and underneath the Arabic digit appeared the word "YES". Children were asked to respond by pressing a "YES" key (Q, W, E, R, A, S, D, F, Z, X, C) on the left side of the keyboard or a "NO" key (U, I, O, P, J, K, L, N, M, ; , <) on the right side of the keyboard according to whether the two numbers represented the same numerosity or not. Numerosities varied between 1 and 10. Trials including 1–4 dots (10 with the same numerosity and 10 with different numerosities) were combined into a subitizing measure and trials including 5–10 dots (18 with the same numerosity and 18 with different numerosities) were combined into an enumeration measure. The task consisted of 56 stimuli (Cronbach's α=.83) presented in a fixed, pseudo-random order (the same item never appeared in two consecutive trials). Each trial started with the presentation of a pair remaining on the screen until a response was given, followed by an ISI (white screen) of 500 milliseconds (ms). The software recorded both accuracy and reaction times to the order of ms. The inverse efficient score (IES) was used as the measure. IES consists of the mean reaction time of the correct responses divided by the proportion of the correct responses (Townsend & Ashby, 1978). Since the reaction times are expressed~~s~~ in ms and divided by proportions, IES is expressed in ms as well. The larger the IES of a student, the worse his performance.

Number Magnitude Comparison
Two numbers from 1 to 98 in Arabic digits were simultaneously displayed on the computer screen, one on the left half of the computer screen and the other on the right half. Underneath of each number appeared the word “LARGER.” Children were asked to select the larger number by pressing a “LEFT” key (Q, W, E, R, A, S, D, F, Z, X, C) on the left side of the keyboard or a “RIGHT” key (U, I, O, P, J, K, L, N, M, ; , <) on the right side of the keyboard corresponding to the ~~right~~ correct response. The numbers were displayed until the students responded by pressing the button to response. Comparison pairs varied along two variables: size (small: from 1 to 9; large: from 23 to 98) and distance (close: distance of 1; far: distance of 4 or 5). The pairs were presented as follows: eight small pairs with small distance and eight small pairs with large distance, eight large pairs and eight large pairs with large distance and small distance, respectively. Each pair appeared twice, once in ascending and once in descending order. The 64 stimuli (Cronbach’s α=.70) were presented in a fixed, pseudo-random order (the same item never appeared in two consecutive trials). Each trial started with the presentation of a pair, shown until a response was given, followed by an ISI (white screen) of 500 ms. The inverse efficient score (IES) was used as the measure.

Dots Magnitude Comparison
Students were simultaneously presented with two arrays of dots, one on the left half of the computer screen and the other on the right half. Underneath each array of dots appeared the word “MORE.” Children were asked to select the one that contained more dots by pressing a “LEFT” or a “RIGHT” key of the keyboard corresponding to the right response. Stimuli were pairs of black dots, created based on Gebuis and Reynvoet’s work (2011), and the matlab code publicly provided by the authors[4]. Comparison pairs varied along the Weber fraction (1; 0.5; 0.3; 0.25; 0.2; 0.16 and 0.14) in two numerical sizes: seven small pairs (1-8 dots) and seven large pairs (14-28 dots). Each pair appeared twice, once in ascending and once in descending order in a fixed, pseudo-random order (the same item never appeared in two consecutive trials). Each of the 28 trials (Cronbach’s α=.88) started with the presentation of a pair that remained on the screen until a response was given, followed by an ISI (white screen) of 500 ms. The mean response time (of the correct answers only) was used as the measure.

The above three tasks were designed based on main hypotheses advanced in the literature on deficits within the two preverbal (or non-symbolic) systems for

[4] See http://titiagebuis.eu/Materials_files/comp_dots_version180112.m

processing quantities. A first system is~~: (1)~~ the object tracking system (OTS) that is precise, limited by its absolute set size, and that creates an *object file* with concrete information for each objects observed simultaneously (e.g., Piazza, 2010)~~; (2)~~. A second system is the approximate number system (ANS) that is extensible to very large quantities, operates on continuous dimensions, and yields and approximate evaluation in accordance with Weber's law (e.g., Halberda & Feigenson, 2008; Piazza, 2010). The hypotheses on deficits within the OTS or the ANS, as well as hypotheses on deficits in other mechanisms specific to numerical (symbolic and non-symbolic) processing have been reviewed by Andersson and Östergren (2012), and classified into the following categories: defective ANS; defective OTS; defective numerosity-coding; access deficit; multiple deficit.

For example, De Smedt, Noel, Gilmore and Ansari (2013) highlight how results on the specific association between numerical magnitude processing and mathematics achievement differ depending on the number format used: for symbolic numbers (digits), data seem to be consistent and robust across studies and populations, while for non-symbolic formats (dots), many conflicting findings have been reported. These and other hypotheses related to numerical cognition are also being investigated in neuroscience (Dehaene, 1997; Piazza et al., 2004; Pinel, Piazza, Le Bihan & Dehaene, 2004); Nieder, 2005; Butterworth, 2010). Because of the important role these hypotheses have played in the literature, we will refer to them as hypotheses on "core number" deficits. The three tasks designed for our study aim at detecting domain specific deficits in core number processing.

5.5.2 Tasks in the memory domain

Addition Facts Retrieval

Students were simultaneously presented with a single-digit addition (with operands between 2 and 9) that appeared in the center of the screen, and with two possible answers underneath (one on the left half of the screen and the other on the right half). The possible answers were displayed until the child responded by pressing, as quickly as possible, a "RIGHT" or a "LEFT key of the keyboard corresponding to the correct response. Each trial was followed by an ISI (white screen) of 500 ms. The items varied based on the numerical sizes of the sums (equal to or less than 10 or greater than 10), and on the relationship between the two possible answers (they differed by one unit or had the same parity, see Krueger & Hallford, 1984). The two answers always had the same tens place digit. Twelve additions presented unequal operands, with their sum equal to or less than 10, and the possible answers differed by one unit. Fifteen trials consisted of unequal operands with their sum above 10, and for eight trials the possible answers differed by one unit, while the rest had the same parity. Finally, the task included six items with equal operands (Cronbach's α=.77). The inverse efficient score (IES) was used as the measure.

Multiplication Facts Retrieval

Students were simultaneously presented with a single-digit multiplication (with factors between 2 and 9) that appeared in the center of the screen, and with two possible answers underneath (one on the left half of the screen and the other on the right half). They were instructed to choose the right answer as quickly as possible by pressing a "RIGHT" or a "LEFT key of the keyboard. The trials varied based on the numerical size of the factors (equal to or less than 5 or greater than 5) and on the relationship between the two possible answers (the wrong answer could be a multiple of one of the factors or not). The two answers always had the same tens place digit. The wrong answer had the same parity as the correct one, thus preventing the use of a short-cut based on parity checking (Krueger & Hallford, 1984). In 10 trials both factors were equal to or less than 5, in 15 trials one factor was equal to or less than 5 and the other greater than 5, and nine trials contained factors which were both greater than 5. In 28 trials the wrong answer was a multiple of one of the factors (Cronbach's α=.75). The possible answers were displayed until the child responded by pressing the right or left key corresponding to the correct response. The inverse efficient score (IES) was used as the measure.

Math Terms

Students were simultaneously presented each time with a shape or a number on the center of the screen, in red, and with three math terms underneath. They were instructed to choose the term which corresponded to the red stimulus by clicking with the mouse on one of the possibilities. Twelve trials displayed shapes related to geometry (e.g., Is the colored shape **L** called "right", "acute" or "obtuse"?), while the remaining 18 trials pertained to the content area of arithmetic, (e.g., Is the colored digit in 2**3**8 called "unit", "ten" or "tenth"?), presenting numbers as stimuli (Cronbach's α=.68). The stimuli remained on the screen until a response was given and there was no time limit. The percentage of correct responses was used as the measure.

Mental Calculations

This task consisted of 10 trials (Cronbach's α=.78) in which students were asked to type in the result of an operation that appeared horizontally in the center of the computer screen. The operations were between numbers with up to 3 digits and they did not include division (e.g., 245 + 55 = _ ; 52 - 13=_; 3 x 25 =_). The stimuli remained on the screen until a response was given. There was no time limit. After each response the student could move on to the next trial by clicking with the mouse on the label "NEXT" which was on the right corner of the screen. The percentage of correct responses was used as the measure.

The *Addition facts retrieval* and *Multiplication facts retrieval* tasks were elaborated based on literature on retrieval of numerical facts (Geary, 1993; 2004; von Aster, 2000; Woodward & Montague, 2002) and accurate performance of mental calculation (Andersson & Östergren, 2012; Ashcraft, 1992; Campbell, 1987a, 1987b, 1991). Some design aspects were inspired by Krueger and Hallford's work (1984). The *Math terms* task is based on literature on decoding terminology (Geary, 1993; Hecht, Torgesen, Wagner & Rashotte, 2001). Finally, the *Mental calculations* task was based on studies on students' grasping of mathematical relations (e.g., Geary, 1993; Schoenfeld, 1992), classified in our theoretical model as a skill pertaining to the *reasoning* domain. In fact, we were uncertain about the placement of this task: although it was originally designed to elicit skills primarily from the *memory* domain, we assumed it could also be grouped with tasks eliciting skills in the *reasoning* domain.

*5.5.3 Tasks in th*e number lines domain

Number Lines 0-100

A series of twenty-two number lines, in pairs, containing a horizontal line with two endpoints (0 and 100) was presented to the student, together with a target number (e.g., 29) above the center of each line. In this Number to Position task (see Siegler & Opfer, 2003) the student was asked to consider the first number line (the one on top) and use the mouse to click on the position where the target number (above it, in the center) should lie (for a detailed description, see Siegler & Booth, 2004). The number line coordinates for each response were recorded, based on a pixel count along the length of the line. Accuracy was defined here as the absolute difference between the student's placement of a number and its correct position. These measures were taken for the student's placement of numbers on the 11 lines on the top row of each trial (Cronbach's α=.88). The mean of these absolute differences was used as the measure.

Ordinality

This task was presented to the student together with the previous one (Number Lines 0-100) and it was performed on the second line (the one below) of the two presented simultaneously. The student was asked to perform the same task on the second number line (below it and aligned with it) placing the second target number on it. As this task was carried out, the first estimated position remained on the screen. The software checked whether the placement of the number on the 11 lines in the lower row of each trial was coherent with the estimation of the top target number (Cronbach's α=.88). The measure was the percentage of correct responses.

Number Lines 0-1000

This task was designed in the same way as the Number lines 0-100, except that each line was presented alone. The task consisted of 16 trials (Cronbach's α=.74). Accuracy was defined as the absolute difference between the student's placement of a number and its correct position. The mean of these absolute differences was used as the measure.

The three aforementioned tasks focus on rather specific aspects relating visual spatial skills to properties of natural numbers. Based on this choice to focus on number lines and not include other types of visual-spatial skills, in the rest of the paper we will refer to what we called the "visual-spatial domain" in the theoretical model as the *Number lines* domain[5]. The reasons for our focus on natural numbers and number lines are that we were interested in studying relationships between the main core number skills and other skills, while still related to numbers, also might pertain different domains. Indeed, ~~MLD~~ difficulties in mathematics, and in particular in arithmetic, have~~s~~ also been put in relationship with ~~the~~ atypical development of an internal representation of the number line: a number of studies have explored a relationship between space and the processing of numbers (e.g., Pinel et al., 2004; Seron et al., 1991), ever since initial hypotheses on such a relationship advanced by Galton in 1880. These studies suggest that the (mental) number line model corresponds to an intuitive representation and to a natural translation of the sequence of (natural) numbers into a spatial dimension. The number line model is not a static representation, nor is it necessarily innate[6], instead studies suggest that it evolves as the subject develops cognitively, and such evolution depends on cultural influences (see, for example, Zorzi, Priftis & Umiltà, 2002). Typically, the representation in young children seems to be of a logarithmic nature, with "smaller" numbers (e.g., 1, 2, 3) more distant from one another with respect to larger numbers (e.g., 8, 9, 10) which are 'closer'. As the child grows and is exposed to external representations of the number line and to more and more activities that involve numbers, the representation of the number becomes more linear, that is, all numbers assume the same distance from one another, as on the mathematical number line. Moreover, studies have related other factors such as children's perception of structure and mental imagery to their development of the counting sequence 1–100, which is closely related to development of the number line (e.g., Thomas, Mulligan & Goldin, 1992).

Tasks such as *Number Lines 0-100* and *Number lines 0-1000* have been used and described in various studies in the literature (e.g., Siegler & Opfer; 2003; Siegler & Booth, 2004), and they have been put relation with sets of visual-spatial skills (Cooper, 1984; Dehaene & Cohen, 1997; Ward, Sagiv, & Butterworth, 2009). Neuroscience has also shown that numbers and space associate in the parietal cortex for the general population (e.g., Hubbard, Piazza, Pinel & Dehaene, 2005)

The *Ordinality* task was designed to gain deeper insight into students' abilities to spatially relate positions of numbers (given in symbolic format) to one another. Indeed, ordinality refers to the capacity to place numbers in sequence; for example, to know that 6 comes before 7 and after 5 in the sequence of natural numbers. There is neuro-scientific evidence that accessing ordinal information from numerical symbols relies on a different network of brain regions and that such accessing has qualitatively different behavioral patterns when compared to ordinal processing of perceptual magnitudes (Lyons & Beilock, 2013). Other neuro-scientific studies have

---

[5] In this a priori classification we were actually uncertain about the relationship of these 3 tasks with the four proposed domains. In fact, in the classification of skills in the theoretical model we included number lines tasks both in the *visual-spatial* domain and in the *core number* domain.

[6] For a more complete discussion see volume 42(4) of the *Journal of Cross-Cultural Psychology*.

associated activation of the posterior IPS to the ordinal nature of number forms (Tang, Ward & Butterworth, 2008), so a hypothesis that has been discussed is that the same networks may be involved in spatial-form synesthetes (Jonas & Jarick, 2013) and thus in specific visual-spatial abilities. These might also be involved in perceiving pattern and structure, fundamental abilities linked to visual-spatial skills, studied in depth by Mulligan and her colleagues (e.g., Mulligan & Mitchelmore, 2013; Mulligan, Mitchelmore & Stephanou, 2015).

5.5.4 Tasks in the reasoning domain

Calculation Principles

Students were instructed to type a number through the computer's arithmetic keyboards into a gap in an equation that appeared horizontally in the center of the computer screen, above an equality. The number to be typed into the equation could be obtained without computation, using the "relevant principle" introduced in the equality. The principles used were: commutativity of addition (e.g., "what is 253 +147 =_, if 147 + 253 = 400?") and multiplication (e.g., "what is 150 ÷ 12 =_, if 120 x 15 = 1800?"), the property of the equalities representing inverse operations (addition/subtraction and multiplication/division) (e.g., "what is 365 + 135 =_, if 500 - 365 = 135?" or "what is 108 ÷6 =_, if 6 x 18 = 108?"), or "double plus or minus one" (e.g., "what is 4173 +4172 =_, if 4173 + 4173 = 8346?"). The 10 trials (Cronbach's α=.73) were presented in a vertical format on the computer screen. The stimuli remained on the screen until a response was given. There was no time limit. Once the student had responded, s/he could pass to the next questions by clicking on "NEXT" which was on the right corner of the screen. The percentage of correct responses was used as the measure.

Equations

This set consisted of 10 trials (Cronbach's α=.77) in which the students were instructed to fill the gap in an equation containing numbers with 1 to 3 digits, which appeared horizontally in the center of the computer screen. To fill the gap the student by using mouse had to click on the gap and select from a menu of possible answers which appeared under the gap. Only one answer was correct for each equation, and it could be a number or the math symbol of an operation. Four equations were made up of one operation, with the choice having to be made between numbers (e.g., _ ÷ 2 = 400, choosing between 200 or 800); in the other six trials the gaps had to be filled by choosing between the four basic math's symbols (+ - x ÷) and the equations could have more than one operation (e.g., 37_5 = 185 or 10 _ 8 _ 79 = 1). The stimuli remained on the screen until a response was given by choosing the proper number or symbol. There was no time limit. After each response students could move on to the next trial by clicking on the label "NEXT". The percentage of correct responses was used as the measure.

Word Problems

Each student was asked to solve short story problems involving addition, subtraction, multiplication, and division, according to the classification described below. In order to impose as light a linguistic demand as possible, the problems were one to three sentences long and the experimenter read the individual problems while the student followed along on the computer screen. The student was instructed to respond mentally (paper and pencil was not permitted) and type the numerical answer into an empty box, in the right hand corner of the screen by using the computer's arithmetic keyboards. Six trials contained addition-subtraction problems (Carpenter & Moser, 1982): two comparison problems (e.g., "Chris has 35 markers. We know that he has 5 less markers than John. How many markers does John have?"); three change problems (e.g., "Stella has washed 5 pairs of socks. When she went to take them out of the washing machine one sock was missing. How many socks did Stella take out of the washing machine?"); one equalization problem (e.g., "Peter has 40 cards. If Alex lo~~o~~ses 10 cards, he will have as many as cards Peter does. How many cards does Alex have?"). The other seven trials consisted of multiplication-division problems (Vergnaud, 1983) (e.g., "One family has 3 children. Each child of the family drinks 2 glasses of milk every day. How many glasses of milk will the family drink during 10 days?"). One problem contained irrelevant information and four required two calculation steps (Cronbach's α=.73). The stimuli remained on the screen until a response was given. There was no time limit. After each response students could move on to the next trial by clicking on the label "NEXT". The percentage of correct responses was used as the measure.

Calculation principles task was designed based on literature on students' grasping of numerical relations, basic logical principles (e.g., Geary, 1993; Núñez & Lakoff, 2005) and decision making (Desoete, & Roeyers, 2006). Previous studies have used tasks similar this one (e.g., Hanich et al., 2001). The design of *Equations* task was based on the same literature as for *Calculation principles* task. *Word problems* task was designed based on the vast literature on MLD and students' problem solving skills. As has been done in many other studies, we based the addition and subtraction problems on seminal work by Carpenter and Moser (1982), while the multiplication-division problems on ideas of Vergnaud (1983), later developed and studied, for example, by Kouba (1989) and Mullligan and Mitchelmore (1997).

5.6 Statistical analyses

Statistical analyses were performed on IBM SPSS 21 and AMOS 21. Analysis of Variance and Pearson's correlation coefficients were used to test for group differences and bivariate relationships respectively. Principal Components Analysis and Confirmatory Factor Analysis were used to obtain an a posteriori grouping of the tasks in the battery, and elaborate and test the fit of three tested models (one with all tasks grouped into a single factor, the second with the tasks grouped as in the a priori analysis, and the third with the tasks grouped as obtained from the PCA). Confirmatory Factor Analyses (CFA) were conducted using AMOS 21 (Arbuckle, 2012). The following criteria were used in evaluating overall goodness of fit for the measurement models: (a) the chi-square/degrees of freedom ratio, 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); the Adjusted Goodness of Fit Index (AGFI); (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 each time model 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 .95 is generally accepted as indicating a good fit (Hu & Bentler, 1999). For the RMSEA, an adequately fitting model will have a value between .00 and .06, with 90% CIs between .00 and .10 (Browne & Cudeck, 1993). Finally, regarding SRMR, a value less than .08 is considered a good fit (Hu & Bentler, 1999).

Finally, K-means cluster analysis was conducted on the data from the tested population to gain insight into possible types of mathematical profiles, in particular those of underachieving students, as has been done in previous studies (e.g., von Aster, 2000; Bartelet, Ansari, Vaessen & Blomert, 2014). In this method the number of clusters is defined in advance; the criterion used to decide the number of clusters was the maximum number for which the differences between groups remained statistically significant (Tan, Steinbach & Kumar, 2006).

5.7 Descriptive statistics of the three groups

Background information and results on the NUCALC battery and the Nonverbal IQ are displayed separately for the mathematical learning disabilities (MLD), the low achievement (LA) and the typical achievement (TA) groups in Table 21.

TABLE 21 here

Comparisons among groups were made using analysis of variance (ANOVA), and significant group effects were investigated using the Tukey post hoc test, controlling alpha at $p < .05$. The groups differed both in NUCALC battery score, $F(3, 164) = 182.21$, $p < .001$ and in Non-verbal IQ, $F(3, 164) = 4.55$, $p < .01$.

Table 32 presents descriptive statistics (means and standard deviations) for each task of the experimental battery and for each one of the three ability groups. The tasks in Table 32 are presented in the order in which they were administered. In addition, Hedges' g coefficients were calculated on the mean differences of the MLD and TA students only. Hedges' g is a variation of Cohen's d that corrects for biases due to small sample sizes (Hedges & Olkin, 1985). The magnitude of Hedges' g may be interpreted using Cohen's (1988) convention as small (0.2), medium (0.5), and large (0.8). It is apparent that in all cases (with the exception of the Dots Magnitude Comparison task) effect sizes were large.

Table 32 here

6. Results

6.1 Pearson correlation coefficients for the experimental battery

Pearson correlation coefficients were calculated between the values of the tasks of the experimental battery (see Table 4~~3~~). Most of the coefficients were statistically significant. The highest were (all statistically significant at the .001 level of significance): between tasks 4 and 5 (r = .92); between tasks 11 and 12 (r = .66); between tasks 1 and 2 (r = .64); between tasks 11 and 13 (r= .56); between tasks 12 and 13 (r = .55); between tasks 2 and 4 (r = .55); between tasks 9 and 12 (r = .54). Other correlations worthy of attention are between the three tasks (1, 2, 3) designed to elicit *core number* skills. These three tasks also correlate moderately well with tasks 4 and 5 (r = .48, r = .55, r = .30 and r = .47, r = .49, r = .30, respectively). Moreover, task 2 has a moderately high correlation with task 12 (r = .48), and tasks 4 and 5 with tasks 9 (r = .41, r = .42) and 12(r = .43, r = .40). Task 9, was found to correlate moderately well not only with task 12, but also with task 10 (r = .46), task 11 (r = .47), and task 13 (r = .42). Also tasks 11, 12, and 13 were found to correlate moderately well. Task 8, Number Lines 0-1000, was unexpectedly found to correlate moderately with tasks 9, 11, 12 and 13. Finally we remark on one unexpected low correlation between tasks 6, Number Line 0-100 and 8, Number Line 0-1000.

TABLE 4~~3~~ here

6.2 Principal Component Analysis

Principal Component Analysis (PCA) was conducted on the tasks of the experimental battery. An orthogonal rotation (varimax) was chosen since the components were expected to be independent. The Kaiser–Meyer–Olkin measure (KMO = .81) verified the sampling adequacy for the analysis, and Bartlett's test of sphericity [$\chi^2(78)$ = 654.60, p < .001] indicated that correlations between items were sufficiently large for PCA (Field, 2009). All KMO values for individual items were >.51, which is above the acceptable limit of .5 (Field, 2009). An initial analysis was run to obtain eigenvalues for each component in the data. Four components had eigenvalues over Kaiser's criterion of 1 and average communality was .68; in combination these four components explained 67.9% of the variance. Table 5~~4~~ shows the factor loadings after rotation. The items that cluster on the same components suggest that the first component groups tasks eliciting skills from the *reasoning* domain, the second groups tasks eliciting skills from the *memory* domain[7], the third groups tasks eliciting skills from the *core number* domain and four component groups tasks eliciting skills from the *number lines* domain.

TABLE 5~~4~~ here

6.3 Confirmatory Factor Analysis

[7] Since both the tasks that loaded on this component elicit fact retrieval, in the rest of the paper we will refer to this component as *facts retrieval* domain.

The following three models were tested through CFA. In Model I all thirteen tasks were hypothesized to load on a single factor. In Model II the thirteen domains served as indicators for four factors as grouped in a priori four-pronged model: Dots magnitude comparison, Subitizing-Enumeration and Number Magnitude Comparison for factor 1; Multiplication Facts Retrieval, Addition Facts Retrieval, Maths Terms and Mental Calculations for factor 2; Equations, Word Problems and Calculation Principles for factor 3; and, finally, Number Lines 0-100, Ordinality and Number Lines 0-1000 for factor 4 (see also Table 9~~8~~). Finally, the last model (III) tested was the one identified by the PCA reported earlier (see Table 5~~4~~). The fit indices of the three structure models of CFA are shown in Table 6~~5~~. To compare the three models, the Akaike information criterion (AIC) was used with smaller values representing a better fit for the hypothesized model.

TABLE 6~~5~~ here

As shown in Table 6~~5~~, only model III provided acceptable fit to the data and exhibited the lowest AIC value. Based on these results, the a posteriori groupings identified by the PCA was found to be the best in capturing the structure of the battery of tasks.

Based on the results of the PCA and CFA reported above, mean values for the four components (*reasoning, facts retrieval, core number* and *number lines*) were calculated. To rescale students' raw scores on the 13 tasks into a standardized scale, Stanine scores were calculated. Table 7~~6~~ presents the Pearson's correlation coefficients between the four components of Grouping III. Four coefficients were statistically significant, but their values were low suggesting independence between the four components. Only those between *facts retrieval* and *reasoning* (r = .43) and *facts retrieval* and *core number* (r = .49) were large enough, indicating a percentage of shared variance around 20%.

TABLE 7~~6~~ here

6.4 K-means cluster analysis

Finally k-means clustering was used to partition the sample into homogenous subgroups. Table 8~~7~~ shows the results obtained for six distinguishable clusters, the maximum number for which the differences between groups remained statistically significant (Tan, et al., 2006) as well as the distribution of MLD and LA students among the six clusters. The clusters describe two TA groups, the 2nd (no MLD or LA students) and the 3rd (no MLD and 3 LA students) with performances above average on all tasks, and with the following differences: the 2nd group performs less well on the core number tasks, but better than the 3rd group on the *number lines* and *reasoning* domain. The four other groups in which the LA and MLD students are distributed are characterized by the presence of different specific weaknesses we will discuss in section 7.4.

Table 8~~7~~ here

7. Discussion

7.1 Correlations among the mathematical tasks of the experimental battery

Tasks designed to elicit skills from the *core number* domain were correlated with each other, as expected. The same was found for the tasks assumed to elicit skills from the *reasoning* domain. However high correlations were also found between these tasks and two of the tasks designed to elicit skills from the *memory* domain: the Maths Terms task (task 9) and the Mental Calculations task (task 11). This result is not surprising when we consider recent findings correlating memory (working memory) and high achievement in mathematics (e.g., Passolunghi & Siegel, 2004; Andersson, 2008; Mammarella et al., 2010). A similar explanation holds for the correlation found between the Equation task (task 12) and the Facts Retrieval tasks (tasks 4 and 5). On the other hand, the correlation found between the same tasks (4 and 5) and Maths Terms (task 9) can be explained by their common reliance on the use of memory. The correlation between the Equation task (task 12) and the Number Magnitude Comparison (task 2) can possibly be explained by their eliciting the ability of dealing with the symbolic representations of numerosity. Weaknesses in this ability have been studied elsewhere in relation to underachievement in mathematics (e.g., Rousselle & Noël, 2007).

A surprising result was the negative correlation of the Number Lines 0-1000 task (task 8) with the Number Lines 0-100 task (task 6). This finding needs further investigation, but a possible hypothesis could be that mathematical instruction frequently (this is the case in Greece, where the study was carried out) focuses on numbers up to 100 in early grades (pre-K, 1^{st} and 2^{nd} grade), and only later is extended to larger numbers up to 1000 (3^{rd} grade), with less focus. Therefore skills related to placing numbers on 0-100 lines and ordering numbers may be fundamentally different from those developed for 0-1000 lines.

7.2 Evidences for supporting the four-pronged model for detecting strengths and weaknesses in learning mathematics

PCA revealed that the tasks designed do indeed fall into four components, which correspond to our a priori grouping of the tasks, with only a few changes. Both the a priori and posteriori grouping of the tasks are shown in Table 9~~8~~.

TABLE 9~~8~~ HERE

The differences of this grouping compared to the original hypothesized grouping appear clearly in Table 9~~8~~ comparing the a priori and a posteriori classification of the tasks: three of the thirteen types of tasks did not load on the expected components. 1) The Number Lines 0-1000 task loaded with the tasks designed to elicit skills from the *reasoning* component; 2) the Maths Terms task also loaded on the *reasoning* component; 3) the Mental Calculations task also loaded on the *reasoning* component. Although the Maths Terms task was expected to fall on the *memory* component, we could explain its tight relationship to the *reasoning* tasks because of a possible significant semantic component of the tasks. The placement of the Number Lines 0-1000 task differed from that of the Number lines 0-100 tasks. This finding is consistent with the weak correlation found and discussed in section 7.1; it needs further investigation, and we suggested a possible hypothesis for it above.

Moreover, within each component the tasks were also highly correlated. Particularly strong correlations were obtained for the following components: *reasoning* – between Mental Calculation and Equations, Mental Calculation and Word Problems, Equations and Word Problems; *facts retrieval* – between Addition Facts Retrieval and Multiplication Facts Retrieval; *core number* – between Subitizing-Enumeration and Number-Magnitude Comparison.

The only other high correlation obtained that does not correlate tasks grouped in a same component is between Number Magnitude Comparison and Addition Facts Retrieval. This finding suggests that managing Arabic digits is highly correlated with facts retrieval, which is consistent with studies that suggest that students who have trouble overcoming difficulties in arithmetic may have weak symbolic comparison abilities (Rousselle & Noël, 2007) and/or weak fact retrieval mechanisms (Andersson, 2008; Geary, 1993).

Confirmatory factor analysis also revealed that such grouping into the four expected components is also the best fit of three models analyzed. Pearson's correlation coefficients between the four components of model III emerging from the PCA revealed that the components are mutually independent. This suggests independence between the sets of skills elicited by the tasks, which can, in turn, and in a much weaker way, suggest independence of the four dimensions of the model. Therefore, we can expect that if the mathematical skills of a student (including those with an MLD or LA profile emerging from a standardized test like NUCALC battery) are weak within a component, they will not necessarily be weak on other components. This is further supported by the different profiles that emerged from cluster analysis.

7.3 MLD and LA students' performance on the mathematical tasks of the experimental battery

As expected, we found that the TA group outperformed both the LA and the MLD groups on all tasks of the battery; moreover, the LA group outperformed the MLD group on all tasks except for the Dots Magnitude Comparison task; and the MLD group performed significantly less well than the ~~control~~ TA group (even with corrections of biases due to the small sample size obtained through use of Hedges' g) on all tasks except the Dots Magnitude Comparison task.
These findings, a part from the anomaly on the performances on the Dots Magnitude Comparison task that we explain in the section on limitations of this study (section 8), suggest a continuum in students' math abilities that goes from low, to average, up

to exceptional, a result which is in line with other studies (e.g., Dowker, 2005; Raghubar et al., 2009; Reigosa-Crespo et el., 2011).

7.4 Types of emerged mathematical profiles in general and those of MLD and LA specifically

Our results from the k-means cluster analysis support that students, both the normal/high achievers and the underachievers, do not all share the same sets of strong and weaker mathematical skills; nor that under achievement in mathematics is related to weaknesses in a single domain. These results are consistent with other studies attempting to identify defining characteristics of MLD (e.g., Geary, 2004; Andersson, & Östergren, 2012; Lewis & Fisher, 2016; Szucs, 2016). Although the population in the present study contains only 9 MLD and 17 LA students, the distribution of these 26 students within the 1^{st}, 4^{th}, 5^{th} and 6^{th} cluster identified suggests that the mathematical profiles of the weaker students are not of a same type. Instead, our results suggest that for these students, just like for the other students, cognitive strengths or weaknesses may rely in any of the four domains of the four-pronged model. In particular, we found that

- 4^{th} cluster performs very poorly on tasks eliciting skills from the *facts retrieval* and *core number* domain, while their performance on tasks on the *number lines* or eliciting skills from the *reasoning* domain are around average, or slightly above average;
- 5^{th} cluster performs very poorly on the *reasoning* domain, but around or above average on tasks eliciting skills from the other domains;
- 6^{th} cluster performs very poorly on the *core number* domain, but around or above average on tasks eliciting skills from the other domains;
- 1^{st} cluster performs poorly on all tasks.

Both the 2^{nd} and the 3^{rd} clusters perform above average on all tasks, but excel respectively on tasks eliciting skills from the *reasoning* domain or the *core number* domain. ~~Consistently with studies suggesting a continuum in students' math abilities that goes from low, to average, up to exceptional (Dowker, 2005), t~~The 1^{st} cluster we found contains the weakest students and the other clusters identify students with performances characterized by weaknesses on certain types tasks (4^{th}, 5^{th}, 6^{th}) up to the 2^{nd} and 3^{rd} clusters identifying the normal to high achievers. This is consistent with studies suggesting a continuum in students' math abilities that goes from low, to average, up to exceptional (Dowker, 2005). These results are also consistent with Szucs's plea to better understand MLD and its possible subtypes by taking a multidimensional parametric approach, positioning individuals in a multidimensional parametric space, in order to understand the multidimensional structure of cognitive functions and their relationship to mathematical performance (Szucs, 2016).

Finally, although the 1^{st} cluster may stand for a persistent and serious math disability due to very low performance in all domains, because of the distribution of the 26 under achieving students in 5 groups, we prefer to see our results as supporting the hypothesis that developmental mathematical difficulties can have multiple origins (Jordan, Hanich & Kaplan, 2003; Mazzocco & Myers, 2003; Dowker, 2005).

8. Limitations

A first limitation of the study is the still limited selection of mathematical skills assessed in the battery: we mostly focused on tasks within the domain of arithmetic, that are related to main hypotheses advanced on MLD in the literature. Moreover, the detection of skills pertaining to the *visual-spatial* domain through the experimental battery is particularly weak. In fact we spoke of the *number lines* domain, instead of the *visual-spatial* domain because the only tasks in the battery used to assess this domain involved the number line. This issue is being addressed in an ongoing study in which a broader range of tasks eliciting visual-spatial skills are used.

A few limitations have to do with programming defects of the computer software used to administer the battery. In evaluating the students' performance on the battery, we did not control for reaction time, which implies that we cannot know whether the correlation between tasks and the factors identified is due to the content of the tasks or to the measure used. Other studies do control for this factor (e.g. Reigosa-Crespo et al., 2011) and we have already implemented this in the newer version of the battery. Moreover, in the Dots Magnitude Comparison task there is a programming defect: when the sets of dots were presented to the students on the screen, they did not disappear until the students provided an answer. This allowed some students to count the dots, delaying their response time, even though they were asked not to do so and answer as quickly as possible. We also note that in the new version of the battery the arrays of dots appear for a controlled period of time.

A final limitation regards the generalizability of the results of the study based on the age-range and nationality of the participants. The authors are currently adjusting and enriching the battery so that it can be administered to an international population of school students.

9 Conclusion and future directions

In this study we described and tested a literature driven four-pronged model designed to identify stronger and weaker sets of students' mathematical skills. To do this we designed a computer based experimental battery for students aged 10-12 inspired by the model, and we compared an a priori grouping of the tasks to a posteriori groupings obtained through explanatory and confirmatory on the data obtained from students' performances. The analyses confirmed a posteriori grouping of the mathematical skills elicited in the experimental battery that is mostly consistent with the a priori grouping supporting the solidity of the four-pronged model.

We also searched for mathematical profiles through k-means cluster analysis, which showed how it is possible for both MLD students and non-MLD students to belong to clusters with quite different characteristics, and thus apparently have completely different mathematical profiles. Moreover, our results suggest a continuum in students' mathematical abilities and supporting the hypothesis that MLD can have multiple origins, as has been suggested by other research studies (e.g., Jordan, Hanich & Kaplan, 2003; Mazzocco & Myers, 2003; Dowker, 2005; Szucs, 2016).

We believe it is a high priority that research on mathematical learning and teaching, including research on difficulties in learning mathematics, is approached in an interdisciplinary way. However, as educators, we acknowledge the difficulty of implementing in the field of mathematics education some important findings from neighboring fields of research in which different research paradigms are used (e.g., Ansari & Lyons, 2016). Our theoretical four-pronged model was a first attempt at intertwining main MLD hypotheses in a mathematically holistic way, with the aim of constructing a tool giving insight to educators (classroom teachers, one-on-one after school coaches, clinicians who propose remedial interventions) on how to better understand the needs of the students they are working with. This is the direction we have been working in, trying to develop tools for unearthing a student's cognitive weaknesses and strengths *in mathematics*, no longer focusing on specific "syndromes" (frequently labeled as "dyscalculia", "dyslexia", "ADHD", or "autistic spectrum") but instead bringing to the forefront their acquisition of *specific mathematical skills*. This direction of research seems to be in line with the approach to specific learning disorders with impairment in mathematics suggested in DSM V (2013).

One of our more long term aims is to design assessment tools that elicit greater numbers of skills pertaining to the four domains of the four-pronged model. This could provide further insight into relationships between students' stronger and weaker skills and their overall mathematical performance. More in general, within tThis trend of research, we propose to ~~could, more in general,~~ investigate relationships between students' performance on this (and more complete versions of the) battery of tasks and their performance on tasks in the mathematics curricula introduced by their teachers. This line of research should explore the potential of the model for sketching out students' *mathematical learning profiles*,, which could eventually lead to more efficient design of remedial interventions. Indeed, we expect that (but for the time being this is only a working hypothesis) students with different profiles respond differently to a same remedial intervention. In particular, interventions could be better tailored to lead the student they are designed for to repeated success by building on his/her strengths, while avoiding to propose repetitive tasks that cause repetitive failure experiences (Author and other, 2014), maximizing the learning opportunities of *all* students (as proposed in other & Author, 2015).

References

American Psychiatric Association. (2013). *Diagnostic and Statistical Manual of Mental Disorders. Fifth Edition. DSM-5.* American Psychiatric Publishing, Inc.

Andersson, U. (2008). Working memory as a predictor of written arithmetical skills in children: the importance of central executive functions. *British Journal of Educational Psychology, 78*, 181–203.

Andersson, U., & Östergren, R. (2012). Number magnitude processing and basic cognitive functions in children with mathematical learning disabilities. *Learning and Individual Differences, 22*, 701–714.

Ansari, D. & Lyons, I.M. (2016). Cognitive neuroscience and mathematics learning: how far have we come? Where do we need to go? *ZDM Mathematics Education, 48*, 379–383.

Ashcraft, M. H. (1992). Cognitive arithmetic: A review of data and theory. *Cognition*, *44*, 75-106.

Augustyniak, K., Murphy, J., & Phillips, D. K. (2005). Psychological perspectives in assessing mathematics learning needs. *Journal of Instructional Psychology, 32*, 277–286.

Bartelet, D., Ansari, D., Vaessen, A., & Blomert, L. (2014). Research in Developmental Disabilities Cognitive subtypes of mathematics learning difficulties in primary education. *Research in Developmental Disabilities, 35*(3), 657–670.

Brainerd, J. (1979). *The origins of the number concept*. New York: Praeger Publishers.

Butterworth, B.(2010). Foundational numerical capacities and the origins of dyscalculia. *Trends in Cognitive Science, 14*, 534–541.

Campbell, J. I. D. (1987a). Network interference and mental multiplication. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13*, 109-123.

Campbell, J. I. D. (1987b). Production, verification, and priming of multiplication facts. *Memory & Cognition, 15*, 349-364.

Campbell, J. I. D. (1991). Conditions of error priming in number-fact retrieval. *Memory & Cognition, 19*, 197-209.

Carpenter T. P., & Moser J. M. (1982). The development of addition and subtraction problem - solving skills, In T. P. Carpenter, J.M. Moser & T. P. Romberg (Eds.), *Addition and subtraction. A cognitive perspective* (pp. 9-25). Hillsdale, N.J: Lawrence Erlbaum Associates.

Coles, A. (2014). Ordinality, neuro-science and the early learning of number. In C. Nichol, S. Oesterle, P. Liljedahl & D. Allen (Eds), *Proceedings of the joint PME 38 and PME-NA 36 conference* (Vol. 2, pp. 329-336). Vancouver, BC: PME.

Cooper, R.G. (1984). Early number development: Discovering number space with addition and subtraction. In C. Sophian (Ed.), *Origins of cognitive skills* (pp. 157-192). Hillsdale, NJ. Erlbaum.

Cragg, L., & Gilmore, C. (2014). Skills underlying mathematics: the role of executive function in the development of mathematics proficiency. *Trends in Neuroscience and Education*, *3*(2), 63 - 68.

Dehaene S. (1997) *The Number Sense. How the Mind Creates Mathematics.* New York: Oxford University Press.

Dehaene S. (2001) Précis of the number sense. *Mind Lang* 16: 16–36.

Dehaene S., Cohen L. (1997). Cerebral pathways for calculation: double dissociation between rote verbal and quantities knowledge of arithmetic. *Cortex 33*, 219–250.

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

Desoete, A., & Roeyers, H. (2006). Metacognitive macroevaluations in mathematical problem solving. *Learning and Instruction, 16,* 12-25

Dowker, A. D. (2005). *Individual differences in arithmetic: Implications for psychology, neuroscience and education*. Hove, East Sussex: Psychology.

Field, A. (2009). *Discovering Statistics using SPSS*. London: Sage.

Fletcher, J. M., Lyon, G. R., Fuchs, L. S., & Barnes, M. A. (2007). *Learning disabilities: From identification to intervention*. New York, NY: Guilford Press.

Fuchs, L.S. & Fuchs, D. (2002). Mathematical problem-solving profiles of students with mathematics disabilities with and without comorbid reading disabilities. *Journal of Learning Disabilities, 35*(6), 564-574.

Fuchs, L. S., Fuchs, D., Compton, D. L., Bryant, J. D., Hamlett, C. L., & Seethaler, P. M. (2007). Mathematics screening and progress monitoring at first grade: Implications for responsiveness to intervention*. Exceptional Children, 73,* 311–330.

Galton, F. (1880). Visualized numerals. *Nature, 21*, 252-256

Geary, D. C. (1993). Mathematical disabilities: Cognitive, neuropsychological, and genetic components*. Psychological Bulletin, 114*, 345–362.

Geary, D.C. (2004). Mathematics and learning disabilities. *Journal of Learning Disabilities, 37*(1), 4-15.

Gebuis, T., & Reynvoet, B. (2011). Generating nonsymbolic number stimuli. *Behavior Research Methods, 43*(4), 981-986.

Halberda, J., & Feigenson, L. (2008). Developmental Change in the Acuity of the "Number Sense": The Approximate Number System in 3-, 4-, 5-, and 6-Year-Olds and Adults. *Developmental Psychology, 44*(5), 1457–1465.

Hanich, L. B., Jordan, N. C., Kaplan, D., & Dick, J. (2001). Performance across different areas of mathematical cognition in children with learning disabilities. *Journal of Educational Psychology, 93,* 615–626.

Hecht, S. A., Torgesen, J. K., Wagner, R., & Rashotte, C. (2001). The relationship between phonological processing abilities and emerging individual differences in mathematical computation skills: A longitudinal study of second to fifth grades. *Journal of Experimental Child Psychology, 79*, 192–227.

Hedges, L. V. & Olkin, I. (1985). *Statistical methods for meta-analysis.* New York, NY: Academic Press.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*(1), 1-55.

Hubbard, E. M., Piazza, M., Pinel, P., & Deahaene, S. (2005). Interactions between numbers and space in parietal cortex. *Nature Reviews Neuroscience, 6*, 435-448.

Jonas, C., & Jarick, M. (2013). Synesthesia, sequences and space. In J. Simner & E.M. Hubbard, *The Oxford Handbook of Synesthesia* (pp. 123-148). Oxford: Oxford University Press.

Authors, & other (2014).

Authors (2014).

Author & Other (2014).

Kosc, L. (1974). Developmental Dyscalculia. *Journal of Learning Disabilities, 7*(3), 164-177.

Kouba, V.L. (1989). Children's solution strategies for equivalent set multiplication and division word problems. *Journal for Research in Mathematics Education, 20*, 147-158.

Koumoula, A., Tsironi, V., Stamouli, V., Bardani, I., Siapati, S., Annika, G., …,von Aster, M. (2004). An epidemiological study of number processing and mental calculation in Greek schoolchildren. *Journal of learning disabilities, 37*, 377-388.

Krueger, L. E., & Hallford, E. W. (1984). Why 2 + 2 = 5 looks so wrong: On the odd-even rule in sum verification. *Memory & Cognition, 12*, 171-180.

Lewis, K. E., & Fisher, M. B. (2016). Taking stock of 40 years of research on mathematical learning disability: Methodological issues and future directions. *Journal for Research in Mathematics Education*, *47*, 338-371.

Lyons, I., & Beilock, S. (2013). Ordinality and the nature of symbolic numbers. *Journal of Neuroscience, 33*(43), 17052-17061.

Lyons, I.M. & Ansari, D. (2015). Numerical Order Processing in Children: From Reversing the Distance-Effect to Predicting Arithmetic. *Mind, Brain, and Education, 9*(4), 207-221.

Mammarella, I. C., & Cornoldi, C. (2005). Difficulties in the control of irrelevant visuospatial information in children with visuospatial learning disabilities. *Acta Psychologica*, *118*, 211–228.

Mammarella, I. C., Giofrè, D., Ferrara, R., & Cornoldi, C. (2013). Intuitive Geometry and Visuospatial Working Memory in Children showing symptoms of Non-verbal Learning Disabilities. *Child Neuropsychology, 19*, 235-249

Mammarella, I.C., Lucangeli, D., & Cornoldi, C. (2010). Spatial Working Memory and Arithmetic Deficits in Children With Nonverbal Learning Difficulties. *Journal of Learning Disabilities, 43(*5), 455-468.

Mazzocco, M. M. M. (2005). Challenges in identifying target skills for math disability screening and intervention. *Journal of Learning Disabilities, 38*(4), 318–323.

Mazzocco, M. M. M., & Räsänen, P. (2013). Contributions of longitudinal studies to evolving definitions and knowledge of developmental dyscalculia. *Trends in Neuroscience and Education, 2*(2), 65-73.

Mazzocco, M. M. M., & Myers, G. F. (2003). Complexities in identifying and defining mathematics learning disability in the primary school-age years. *Annals of Dyslexia*, 53, 218–253

Mulligan, J. (2009). Awareness of Pattern and Structure in Early Mathematical Development. *Mathematics Education Research Journal*, *21*, 33–49.

Mulligan, J. (2011). Towards understanding the origins of children's difficulties in mathematics learning. *Australian Journal of Learning Difficulties, 16*(1), 19-39.

Mulligan, J. T., & Mitchelmore, M. C. (1997). Young children's intuitive models of multiplication and division. *Journal for Research in Mathematics Education*, 309–330.

Mulligan, J. T. & Mitchelmore, M.C. (2013). Early Awareness of Mathematical Pattern and Structure. In L. English & J. Mulligan (Eds.), *Reconceptualizing Early Mathematics Learning* (pp. 29-46). Dordrecht: Springer Science-Business Media.

Mulligan, J.T., Mitchelmore, M.C., & Stephanou, A. (2015). *Pattern and Structure Assessment (PASA): An assessment program for early mathematics (Years F-2) teacher guide*. Australian Council for Educational Research. Melbourne: ACER Press.

Nieder, A. (2005). Counting on neurons: The neurobiology of numerical competence. *Nature Reviews Neuroscience,* 6, 177–190.

Núñez, R., & Lakoff, G. (2005). The cognitive foundations of mathematics: the role of conceptual metaphor. In J. I. D. Campbell (Ed.), *Handbook of Mathematical Cognition* (pp. 109–125). New York, NY: Psychology Press.

Passolunghi, M. C., & Siegel, L. S. (2004). Working memory and access to numerical information in children with disability in mathematics. *Journal of Experimental Child Psychology, 88*, 348–367.

Piazza, M. (2010). Neurocognitive start-up tools for symbolic number representations. *Trends in Cognitive Sciences 14*, 542-551.

Piazza, M., Facoetti, A., Trussardi, A.N., Berteletti, I., Conte, S., Lucangeli, D., …, & Zorzi, M. (2010). Developmental trajectory of number acuity reveals a severe impairment in developmental dyscalculia. *Cognition, 116*(1), 33–41.

Pinel, P., Piazza, M., Le Bihan, D., & Dehaene, S. (2004). Distributed and overlapping cerebral representations of number, size, and luminance during comparative judgments, *Neuron, 41*(6), 983–993

Raghubar, K., Cirino, P., Barnes, M., Ewing-Cobbs, L., Fletcher, J., & Fuchs, L. (2009). Errors in multi-digit arithmetic and behavioral inattention in children with math difficulties. *Journal of Learning Disabilities, 42*(4), 356–371.

Reigosa-Crespo, V., Valdés-Sosa, M., Butterworth, B., Estévez, N., Rodríguez, M., Santos, E., ... & Lage, A. (2011). Basic numerical capacities and prevalence of developmental dyscalculia: The Havana Survey. *Developmental Psychology*, *48*(1), 123-135.

Raven, J., Court, J. H., & Raven, J. C. (1995). *Coloured progressive matrices*. Oxford, England: Oxford Psychologists Press.

Rousselle, L., & Noël, M. P. (2007). Basic numerical skills in children with mathematics learning disabilities: A comparison of symbolic vs. non-symbolic number magnitude processing. *Cognition, 102,* 361-395.

Other & Author (2015).

Schoenfeld, A. (1992). Learning to think mathematically: Problem solving, metacognition, and sense making in mathematics. In D.A. Grouws & the NCTM (Eds.), *Handbook of research on mathematics teaching and learning*, NY: Maxwell Macmillan International, pp. 334–370.

Seron, X., Deloche, G., Ferrand, I., Cornet, J.-A., Frederix, M., & Hirsbrunner, T. (1991). Dot counting by brain damaged subjects. *Brain and Cognition, 17*, 116–137.

Shalev, R. S., Manor, O., Kerem, B., Ayali, M., Badichi, N., Friedlander, Y., & Gross-Tsur, V. (2001). Developmental dyscalculia is a familial learning disability. *Journal of Learning Disabilities, 34*(1), 59–65.

Siegler, R. S., & Opfer, J. E. (2003). The development of numerical estimation: evidence for multiple representations of numerical quantity. *Psychological Science, 14*, 237-243.

Siegler, R. S., & Booth, J. L. (2004). Development of numerical estimation in young children. *Child Development, 75*, 428-444.

Szűcs, D., (2016). Subtypes and co-morbidity in mathematical learning disabilities: Multi-dimensional study of verbal and visual memory processes is key to understanding. *Progress in Brain Research, 227*(11), 277-304.

Szűcs, D., & Goswami, U. (2013). Developmental dyscalculia: Fresh perspectives. *Trends in Neuroscience and Education, 2,* 33–37.

Tan, P. N., Steibach, M., & Kumar, V. (2006). *Introduction to Data Mining*. Pearson.

Tang, J., Ward, J., Butterworth, B. (2008). Number forms in the brain. *Journal of cognitive neuroscience, 20*, 1-10.

Thomas, N., Mulligan, J.T., & Goldin, G.A. (2002). Children's representations and cognitive structural development of the counting sequence 1–100. *Journal of Mathematical Behavior, 21*, 117–133.

Thorndike, R. L. (1982). *Applied Psychometrics*. Boston, MA: Houghton Mifflin.

Townsend, J.T., & Ashby, F.G. (1978). Methods of modeling capacity in simple processing systems. In J. Castellan & F. Restle (Eds.), *Cognitive theory*. Vol. 3. (pp. 200-239). Hillsdale, N.J.: Erlbaum.

Vergnaud, G. (1983). Multiplicative structures. In R. Lesh & M. Landau (Eds.), *Acquisition of mathematical concepts and processes* (pp. 127-174). New York: Academic Press.

von Aster, M. (2000). Developmental cognitive neuropsychology of number processing and calculation: varieties of developmental dyscalculia. *European Child & Adolescent Psychiatry, 9*, 41–58.

von Aster, M.G. & Shalev, R.S. (2007). Number development and developmental dyscalculia. *Developmental Medicine & Child Neurology, 49*, 868-873.

Ward, J., Sagiv, N. & Butterworth, B. (2009). The impact of visuo-spatial number forms on simple arithmetic. *Cortex, 45*, 1261-1265.

Watson, S. M. R., & Gable, R. A. (2013). Unraveling the Complex Nature of Mathematics Learning Disability: Implications for Research and Practice. *Learning Disability Quarterly, 36*(3), 178-187.

Woodward, J. & Montague, M. (2002). Meeting the Challenge of Mathematics Reform for Students with LD. *The Journal of Special education, 36*(2), 89-101.

Zorzi, M., Priftis, K., & Umiltà, C. (2002). Neglect disrupts the mental number line. *Nature, 417*, 138–139.

Table 1.

Domains of the four-pronged model and sets of mathematical skills associated with each domain.

Domain	Mathematical skills associated with the domain
Core number	estimating accurately a small number of objects (e.g., 4-5); estimating approximately quantities; placing numbers on number lines; managing Arabic symbols; transcoding a number from one representation to another (analogical-Arabic-verbal)
Memory (retrieval and processing)	retrieving numerical facts; decoding terminology (numerator, denominator, isosceles, equilateral); remembering theorems and formulas; performing mental calculations fluently; remembering procedures and keeping track of steps
Reasoning	grasping mathematical concepts, ideas and relations; understanding multiple steps in complex procedures/algorithms; grasping basic logical principles (conditionality – "if…then…" statements – commutativity, inversion); grasping the semantic structure of problems; (strategic) decision making; generalizing
Visual-spatial	interpreting and using spatial organization of representations of mathematical objects (for example, numbers in decimal positional notation, exponents, geometrical figures or rotations); placing numbers on a number line; confusing Arabic numerals and mathematics symbols; performing written calculation when position is important (e.g. borrowing/carrying); interpreting graphs and tables

Table 2.

Descriptive statistics for the NUCALC and nonverbal IQ tests for students in Grades 5 and 6.

	MLD		LA		TA	
N (Number of boys)	9 (6)		17 (6)		121 (74)	
N per Grade 5/6	5/4		8/9		66/55	
	M	SD	M	SD	M	SD
Age	10.92	0.29	11.27	0.64	11.33	0.59
NUCALC battery	51.33	7.05	58.97	1.67	66.27	1.49
Nonverbal IQ	27.67	3.67	27.59	3.73	30.44	3.60

Table 3.

Descriptive statistics for each task of the experimental battery per group and Hedges' g coefficients

	MLD (N=9)				LA (N=17)				TA (N=121)				
	M	**SD**	z_{skew}	z_{kurt}	**M**	**SD**	z_{skew}	z_{kurt}	**M**	**SD**	z_{skew}	z_{kurt}	**Hedge's *g*** *
1. Subitizing-Enumeration	3557.80	1491.74	2.61	2.87	3107.62	818.96	.33	-.03	2558.80	960.12	11.08	23.69	1.00
2. Number Magnitude Comparison	1500.21	415.93	1.68	1.57	1292.52	323.35	1.43	.69	1115.90	228.84	3.10	1.19	1.57
3. Dots Magnitude Comparison	1815.52	697.90	-.21	-.45	2167.16	1310.70	2.14	.58	1801.94	1009.08	8.48	9.20	.01
4. Addition Facts Retrieval	4261.41	1167.97	.67	-.19	3361.68	1000.61	1.49	1.06	2913.72	1150.77	6.17	7.92	1.17
5. Multiplication Facts 6. Retrieval	4789.81	1751.48	.72	.48	3346.77	1209.82	1.81	1.65	3012.14	1313.90	6.74	7.13	1.32
7. Number Lines 0-100	7.50	4.19	1.89	1.20	4.89	1.33	.10	-.68	4.47	1.68	7.29	8.57	1.57
8. Ordinality	.86	.14	-.59	-.97	.94	.06	-.56	-.45	.96	.07	-8.04	6.59	1.31
9. Number Lines 0-1000	136.64	57.09	.53	-.44	131.68	53.81	.66	.47	76.02	33.15	5.19	2.70	1.73
10. Maths Terms	.39	.12	.58	-.77	.49	.15	-.63	-.54	.65	.20	-.59	-1.76	1.32
11. Calculation Principles	.43	.20	.02	-.99	.70	.18	-.56	-.69	.81	.15	-6.79	7.03	2.47
12. Mental Calculations	.26	.22	-.29	-1.15	.51	.30	.25	-.84	.76	.22	-3.53	-.23	2.27
13. Equations	.31	.22	-.50	-.94	.54	.16	.36	1.14	.74	.17	-1.60	-1.30	2.48
14. Word Problems	.21	.12	-.37	.39	.48	.21	-.49	.71	.68	.24	-4.02	.66	2.01

* Hedges' *g* was calculated on the mean differences between the control and the MLD students only

Table 4.

Pearson's correlation coefficients between the tasks in the experimental battery.

	2	3	4	5	6	7	8	9	10	11	12	13
1. Subitizing-Enumeration	.64***	.49**	.48**	.47**	.20*	.14	.09	.29**	.23**	.17*	.33**	.29**
2. Number Magnitude Comparison	-	.41***	.55***	.49***	.21*	.21**	.18*	.34***	.30***	.19*	.48***	.30***
3. Dots Magnitude Comparison		-	.30***	.30***	-.00	.00	.16	.03	-.01	-.12	.11	.04
4. Addition Facts Retrieval			-	.92***	.10	.17*	.18*	.41***	.25**	.27**	.43***	.25**
5. Multiplication Facts Retrieval				-	.17*	.18*	.21*	.42***	.25**	.27**	.40***	.18*
6. Number Lines 0-100					-	.33***	.22**	.28***	.33***	.20*	.25**	.28***
7. Ordinality						-	.03	.19*	.28***	.17*	.22**	.24**
8. Number Lines 0-1000							-	.41***	.21**	.45***	.45***	.40***
9. Maths Terms								-	.46***	.47***	.54***	.42***
10. Calculation Principles									-	.36***	.46***	.48***
11. Mental Calculations										-	.66***	.56***
12. Equations											-	.55***
13. Word Problems												-

*p<.05; **p<.01; ***p<.001

Table 5.

Principle Component Analysis (varimax) of the tasks of the experimental battery.

	Components			
	Reasoning	Facts retrieval	Core number	Number lines
Mental Calculations	.77			
Equations	.74			
Word problems	.70			
Number Lines 0-1000	.70			
Maths Terms	.59			
Calculations Principles	.59			.47
Multiplication Facts Retrieval		.88		
Addition Facts Retrieval		.86		
Dots Magnitude Comparison			.80	
Subitizing-Enumeration			.79	
Number Magnitude Comparison			.66	
Ordinality				.84
Number Lines 0-100				.64
Eigenvalues	4.57	2.14	1.09	1.03
% of variance	35.12	16.47	8.41	7.93

Table 6.

Summary of Fit Statistics for the three models tested (N=148).

	χ^2/df	CFI	GFI	AGFI	SRMR	RMSEA	RMSEA 90% CI	AIC
Model I	5.72	.59	.68	.55	.12	.18	.16 - .20	423.87
Model II	3.26	.82	.83	.74	.13	.12	.11 - .14	256.13
Model III	1.19	.99	.93	.89	.07	.04	.00 - .07	134.42

Note. CFI = Comparative Fit Index; GFI = Goodness of Fit Index; AGFI = Adjusted Goodness of Fit Index; SRMR = Standardized Root Mean-square Residual; RMSEA = Root Mean-square Error of Approximation; CI = Confidence Intervals; AIC = Akaike Information Criterion.

Table 7.

Pearson's correlation coefficients between the four components as detected by the experimental battery.

	1	2	3	4
1. Reasoning	-	.43***	.23**	.34***
2. Facts retrieval		-	.49***	.11
3. Core number			-	.09
4. Number lines				-

**p<.01; ***p<.001

Table 8.

Results of K-means cluster analysis (number of clusters = 6).

			Clusters						
	Mean	(SD)	1 (n=6)	2 (n=29)	3 (n=37)	4 (n=31)	5 (n=23)	6 (n=39)	
Core number	4.75	(1.77)	1.58	5.73	6.68	3.29	5.26	3.54	F=82.03, p<.001
Number lines	4.94	(1.11)	2.95	5.77	4.95	5.18	4.24	4.82	F=8.53, p<.001
Facts retrieval	4.77	(2.51)	3.14	6.65	6.52	1.46	4.13	4.99	F=132.25, p<.001
Reasoning	4.91	(1.48)	1.97	6.80	5.02	4.29	3.50	5.16	F=41.93, p<.001
MLD (n=9)			3	0	0	4	2	0	
LA (n=17)			3	0	3	4	2	5	

Table 9.
A priori and posteriori grouping of the tasks of the experimental battery

A priori				Posteriori			
Core number	**Memory**	**Number lines**	**Reasoning**	**Core number**	**Facts retrieval**	**Number lines**	**Reasoning**
Dots Magnitude Comparison	Addition Facts Retrieval	Number Lines 0-100	Equations	Dots Magnitude Comparison	Multiplication Facts Retrieval	Number Lines 0-100	Equations
Subitizing-Enumeration	Multiplication Facts Retrieval	Ordinality	Word Problems	Subitizing-Enumeration	Addition Facts Retrieval	Ordinality	Word Problems
Number Magnitude Comparison	Mental Calculations	Number Lines 0-1000	Calculations Principles	Number Magnitude Comparison			Calculations Principles
	Maths Terms						Mental Calculations
							Maths Terms
							Number Lines 0-1000